Environmental
Studies
Research
Fund

219

Assessing the Quality of Marine Mammal Detections Using Three Complementary Methods

Évaluation de la qualité de la détection de mammifères marins au moyen de trois méthodes complémentaires

Canada

August 2019

ASSESSING THE QUALITY OF MARINE MAMMAL DETECTIONS USING THREE COMPLEMENTARY METHODS

Prepared by

LGL LIMITED
environmental research associates

ALFRED-WEGENER-INSTITUT
HELMHOLTZ-ZENTRUM FÜR POLAR-
UND MEERESFORSCHUNG

Prepared for

Environmental Studies Research Fund

LGL Report FA0049-04
30 April 2019

Assessing the Quality of Marine Mammal Detections Using Three Complementary Methods

Prepared by

Heather R. Smith[1], Daniel P. Zitterbart[2,a], Olaf Boebel[2], and Valerie D. Moulton[3]

[1] **LGL Limited, environmental research associates**
22 Fisher Street, P.O. Box 280, King City, Ontario L7B 1A6

[2] **Alfred-Wegener-Institute Helmholtz-Centre for Polar and Marine Research**
Am Handelshafen 12, 27570 Bremerhaven, Germany

[3] **LGL Limited, environmental research associates**
388 Kenmount Road, PO Box 13248, Stn. A
St. John's, Newfoundland and Labrador A1B 4A5

[a] Present address:
Woods Hole Oceanographic Institution, Ocean Acoustics Lab,
Applied Ocean Physics and Engineering
Woods Hole, MA 02543, USA

Prepared for

Environmental Studies Research Fund
444 7th Avenue S.W.
Calgary, Alberta
T2P 0X8

LGL Report FA0049-04
30 April 2019

Suggested citation:

Smith, H.R., D.P. Zitterbart, O. Boebel and V.D. Moulton. 2019. Assessing the quality of marine mammal detections using three complementary methods. Environmental Studies Research Funds Report No. 219. LGL Report FA0049-04. Report by LGL Limited, King City, ON & St. John's, NL and AWI, Bremerhaven, Germany for Environmental Studies Research Fund, Calgary, AB. 79 p. + appendices. ESRF Report No. 219

Disclaimer

This report is one of the final deliverables for the Environmental Studies Research Fund Project 2014-03S. This report summarizes all project activities related to data collection and analysis and presents an overview of the results. The two scientific manuscripts listed below were prepared using the data collected during this project. We direct readers interested in citing this work to these manuscripts.

Smith, H.R., D.P. Zitterbart, T.F. Norris, M. Flau, E.L. Ferguson, C.G. Jones, O. Boebel, and V.D. Moulton. 2020. A field comparison of marine mammal detections via visual, acoustic, and infrared (IR) imaging methods offshore Atlantic Canada. Marine Pollution Bulletin 154. doi.org/10.1016/j.marpolbul.2020.111026

Zitterbart, D.P., H.R. Smith, M. Flau, S. Richter, E. Burkhardt, J. Beland, L. Bennet, A. Cammareri, A. Davis, M. Holst, C. Lanfredi, H. Michel, M. Noad, K. Owen, A. Pacini, and O. Boebel. 2020. Scaling the laws of thermal imaging-based whale detection. Journal of Atmospheric and Oceanic Technology. doi.org/10.1175/JTECH-D-19-0054.1

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

List of Symbols, Abbreviations and Acronyms

~	approximately
§	section
α	alpha, significance level
p	*p*-value, probability value
3-D	three-dimensional
ad lib	*ad libitum*, spontaneously, without planning
AIMMMS	Automatic Infrared-based Marine Mammal Mitigation System
aMMO	assisted marine mammal observer (i.e., received alerts from IR system)
APS	acoustic processing system
ASL	above sea level
AWI	Alfred-Wegener-Institute, Helmholtz-Centre for Polar and Marine Research
BF	Beaufort wind force
DAQ	data acquisition (device)
dB	decibel
DC	direct current
DCE	Danish Centre for Environment and Energy
DOC	Department of Conservation, New Zealand
DFO	Fisheries and Oceans Canada
DIFAR	directional frequency analysis and recording (sonobuoy)
EA	environmental assessment
eMMO	experienced marine mammal observer
ePAM	experienced passive acoustic monitoring (operator)
ESRF	Environmental Studies Research Fund
FIRST	fast infrared search and track (reconnaissance sensor)
FOV	field of view
FV	fishing vessel
GPS	global positioning system
h	hour
h:mm:ss	time, hours:minutes:seconds
Hz	hertz
ID	identification
iMMO	inexperienced marine mammal observer
iPAM	inexperienced passive acoustic monitoring (operator)
IR	infrared
JNCC	Joint Nature Conservation Committee (UK)
K	kelvin
kHz	kilohertz
km	kilometre
LGL	LGL Limited, environmental research associates
LLC	limited liability company
LT	local time
LWIR	long wavelength infrared

m	metre
MMO	marine mammal observer
n	sample size
N	north
No.	number
NL	Newfoundland and Labrador
NMFS	National Marine Fisheries Service (US)
NOAA	National Oceanic and Atmospheric Administration
MPA	marine protected area
P(a | b)	probability of 'a' occurring given that 'b' has occurred
PAM	passive acoustic monitoring
RDE	Rheinmetall Defence Electronics GmbH
RFP	request for proposals
RME	RME Intelligent Audio Solutions
ROCCA	real-time odontocete call classification algorithm
RPS	RPS Energy Canada Ltd.
RV	research vessel
s	second
SAIL	St. Andrews Instrumentation
SARA	*Species at Risk Act*
SD	standard deviation
SMRU	Sea Mammal Research Unit
SST	Sea Surface Temperature
Statement	*Statement of Canadian Practice with respect to the Mitigation of Seismic Sound in the Marine Environment*
SZ	safety zone
THA	towed hydrophone array
TMA	target motion analysis
UHF	ultra high frequency
UK	United Kingdom
UKCS	United Kingdom Continental Shelf
UTC	Coordinated Universal Time
V	volt
W	west
x	times (multiples)
VHF	very high frequency

EXECUTIVE SUMMARY

Background

Sound in the ocean is generated by a variety of natural and anthropogenic sources, of which marine seismic surveys are recognized as a major contributor. Impulsive sounds generated during seismic surveys have been documented to cause behavioural responses in marine mammals and could result in hearing impairment or injury. The *Statement of Canadian Practice with respect to the Mitigation of Seismic Sound in the Marine Environment* (the "Statement") specifies the minimum mitigation requirements for the planning and conduct of seismic surveys in ice-free marine waters in Canada in order to reduce the potential effects of seismic surveys on marine life. Mitigation actions in the Statement are designed to minimize the probability of marine mammals being exposed to sounds that are loud enough to cause hearing impairment and are often triggered when marine mammals are observed entering, or about to enter, a safety zone around the seismic sound source. Essential elements of effective mitigation include the ability to reliably detect and localize marine mammals in a timely manner, and in some instances, to identify marine mammals to species or some other taxonomic level.

Traditional methods used to detect marine mammals at sea include visual watches conducted by marine mammal observers (MMOs) using binoculars or the naked eye, and passive acoustic monitoring (PAM), typically using a towed hydrophone array. Both methods have recognized weaknesses. MMOs are unable to make detections during darkness and periods of poor visibility (e.g., fog) and are likely to miss animals when they are fatigued or looking in the wrong direction. PAM is only effective when marine mammals vocalize frequently, and when vocalizations are not masked by vessel or other background noise. With the goal of improving marine mammal detections beyond what is currently possible using these traditional methods, studies in recent years have evaluated the use of thermal imaging camera systems to automatically detect marine mammals at the ocean's surface. The Automatic Infrared Marine Mammal Mitigation System (AIMMMS) is a commercially available thermal (IR) system that utilizes a rotating sensor to detect marine mammals at sea.

In addition to the detection method employed, observer experience level has also been shown to affect the ability to detect and identify marine mammals at sea.

This research project was proposed in response to a request for proposals (RFP) issued by the Environmental Studies Research Fund (ESRF). Our project was designed 1) to improve our understanding of how available real-time marine mammal detection methods employed at sea compare and complement each other, and enable effective mitigation action, and 2) to better understand the relationship between MMO and PAM operator experience level and the ability to effectively detect, classify and locate, marine mammals and employ subsequent mitigation measures. Our research was undertaken in conditions typically experienced during seismic surveys conducted offshore Atlantic Canada, however, aspects of our work are applicable to marine mammal monitoring during seismic surveys in general.

Our ESRF project had three objectives

1. Optimize and test the use of the AIMMMS for use in Atlantic Canada;
2. Compare marine mammal detections made using three different methods (i.e., MMOs making visual observations, MMOs utilizing AIMMMS to make detections, and PAM);
3. Evaluate the quality and accuracy of marine mammal detections made by MMOs and PAM operators with different levels of training and experience.

Study Methodology

The first objective required that data be collected over the course of two shore-based field programs: data collected in 2015 were used to optimize the AIMMMS software algorithm for use in Atlantic Canada; which was then tested and further optimized using data collected in 2016. The second and third objectives were addressed using data collected during a vessel-based field program offshore Nova Scotia and Newfoundland in 2017.

The first field program was conducted at Cape Race, NL. Data were collected during 18 July to 23 August 2015. Thermal (IR) data were collected concurrently with visual sightings of marine mammals made using a theodolite. Automatic IR detections were reviewed by observers and matched to sightings made using the theodolite. This dataset was used to optimize the software algorithm used to detect and classify marine mammals for use in the thermal regime offshore Atlantic Canada in the summer.

The second field program was also conducted at Cape Race, NL. Data were collected during 1–31 July 2016. Thermal (IR) data were collected concurrently with visual sightings made by MMOs using reticle binoculars and following a monitoring protocol that would typically be employed during a seismic survey. Performance of the thermal (IR) system was assessed by investigating the number and cause of false positive detections, generating detection functions for different sighting cues and environmental conditions, and comparing the overall performance of the IR system with an experienced MMO.

The final field program was conducted offshore Nova Scotia and Newfoundland during 31 July and 23 August 2017. Marine mammal detections were made concurrently and independently by MMOs, PAM operators and by the automatic thermal (IR) system. MMOs and PAM operators were either experienced or inexperienced, and some MMOs received assistance from the IR system in the form of automatic alerts relayed to them using a tablet computer. PAM detections were primarily made using a towed hydrophone array, though sonobuoys were sometimes deployed. Detection performance was compared in terms of number of detections made, timing/distance to detections, taxonomic level to which marine mammals were classified, number of false positive detections (IR system); and conditional probabilities were used to make pairwise comparisons of individual detection methods.

Results and Manuscripts Produced

The AIMMMS software algorithm was optimized for use in Atlantic Canada. The software detection module *TashDetectC* is openly available at https://gitlab.com/mamaps/TashDetectC.

The vessel-mounted IR camera system was capable of detecting marine mammals in the thermal regime of Atlantic Canada during summer. The IR camera system resulted in 84.5 % of the automatic detections being false positives. This equates to ~5.6 false positives per hour averaged over the entire survey. Sometimes, multiple alerts, often caused by birds or bow-riding dolphins, resulted in several false positive detections within a single minute. It is during these periods of high false positive rates that further work in false alert suppression is necessary.

Detection of marine mammals via PAM, for mitigation purposes, appears to be primarily influenced by hardware and software (as opposed to level of PAM operator training and experience). Thirteen detections of baleen whales were made over the course of ~14 hours of sonobuoy effort compared to a single humpback whale detected using the towed hydrophone array over the course of the entire survey.

Our results support the idea that employing more than one marine mammal detection method concurrently will result in improved marine mammal detection performance overall. PAM and thermal (IR) methods were effective during darkness; PAM and visual methods complemented each other during periods of high sea state and low visibility due to precipitation (including fog); and that thermal (IR) methods can be used to enhance visual methods during periods of good visibility. Employing PAM and visual methods during periods of good visibility also resulted in more detections than if only visual methods were used.

The types of marine mammals detected, and the extent to which they were classified to the species level, varied depending on which detection methods were used. MMOs and the thermal (IR) system effectively detected both baleen and toothed whales at the water's surface, though species were only reliably identified by MMOs. The species detected using PAM methods differed depending whether a towed array (which detected toothed whales almost exclusively) or sonobuoys (which detected baleen whales) were used. Most PAM detections were not classified to species level (e.g., pilot whale spp., unidentified dolphin, Balaenopterid).

Experience level seemingly influenced MMO detection performance in a number of ways. The inexperienced MMO effectively monitored less of the viewable area around the vessel, detected fewer marine mammals, and was less likely to identify these marine mammals to species level, relative to experienced MMOs. The inexperienced MMO was also generally slower to detect marine mammals and made initial sightings of marine mammals when they were closer to the vessel, relative to the experienced MMOs. These findings suggest that the implementation of mitigation measures related to minimizing the amount of sound exposure (i.e., from an airgun array or other sound source) may be less effective when monitoring is conducted by MMOs that are inexperienced.

Data from the 2016 Cape Race field program were combined with data collected during non-ESRF projects conducted at tropical and sub-tropical field sites and analyzed to investigate the performance of the automatic whale detection system in a wide range of environmental conditions. The results of the combined investigations were used to prepare a manuscript that has been submitted for publication (Zitterbart et al. submitted). The thermal perceptibility analysis based on data collected during the 2015 Cape Race field program was also included in the Zitterbart et al. (submitted) manuscript. Data from the 2017 vessel-based field program were used to prepare a manuscript on the at-sea comparison of three methods that currently are, or have been proposed to be, used to detect marine mammals during monitoring programs during seismic surveys. This manuscript has also been submitted for publication (Smith et al. submitted).

The results and conclusions of this study were also used to recommend changes to some components of the Statement of Canadian Practice with respect to the Mitigation of Seismic Sound in the Marine Environment. The recommendation report was authored by Smith and Moulton, and is included here as Appendix C.

Recommendations for Future Study

Our primary recommendation is that future comparisons of detection methods, when the intended application is for monitoring purposes during seismic surveys, be made from aboard a seismic survey vessel(s). This will ensure that a stable and appropriately high platform is available for the IR camera, and that PAM detections will be made in the appropriate acoustic setting, i.e., with airgun pulses.

One of the weaknesses in our 2017 dataset is the lack of distance estimates for the IR detections, which is attributable to frequency with which the capacity of the gimbal to stabilize the IR camera was exceeded. This prevented us from calculating detection functions for the IR system, which would have allowed for a more meaningful comparison of methods. Ensuring that future studies provide a stable platform for the IR system will avoid this issue.

Large numbers of false positive automatic IR detections were encountered during both the shore- and vessel-based field programs. As such, the thermal (IR) system requires further study and refinement so that the delivery of IR alerts to MMOs in real-time can be effectively used as a monitoring tool.

Our results also suggest that PAM methods other than the use of conventional towed hydrophone arrays (THA) coupled with target motion analysis be explored as a means to detect marine mammals acoustically during seismic surveys. Future studies that explore beamforming methods, cross-bearing methods, or hydrophone arrays capable of localizing marine mammals in 3-D space in the context of mitigation would be useful.

RÉSUMÉ

Contexte

Le bruit dans l'océan est produit par diverses sources naturelles et anthropiques et on reconnaît que les relevés sismiques en milieu marin y contribuent grandement. Il a été établi que les impulsions sonores répétitives produites pendant les relevés sismiques entraînaient des réactions comportementales chez les mammifères marins et étaient susceptibles de causer une perte auditive ou un traumatisme auditif. L'*Énoncé des pratiques canadiennes d'atténuation des ondes sismiques en milieu marin* (ci-après « l'Énoncé ») précise les exigences minimales en matière d'atténuation pour la planification et la réalisation des levés sismiques dans les eaux marines au Canada afin de réduire les effets possibles des relevés sismiques sur la faune marine. Les mesures d'atténuation indiquées dans l'Énoncé sont conçues de manière à réduire la probabilité que les mammifères marins soient exposés aux bruits qui sont suffisamment puissants pour causer une perte auditive et sont souvent déclenchées lorsqu'on observe l'entrée ou l'entrée prochaine de mammifères marins dans une zone de sécurité entourant la source de bruits sismiques. Les éléments essentiels d'une atténuation efficace incluent la capacité à détecter et à localiser de manière fiable et rapide les mammifères marins et, dans certains cas, à déterminer leur espèce ou à les identifier à un certain niveau taxinomique.

Les méthodes conventionnelles utilisées pour détecter les mammifères marins en mer consistent en la surveillance visuelle effectuée par les observateurs des mammifères marins (OMM) à l'aide de jumelles ou à l'œil nu, et la surveillance acoustique passive (SAP), effectuée habituellement au moyen d'un réseau d'hydrophones remorqués. Il est établi que les deux méthodes présentent des points faibles. Les OMM ne sont pas à même d'effectuer des détections pendant les périodes d'obscurité ou de visibilité réduite (en raison du brouillard, par exemple) et, s'ils sont fatigués ou regardent dans la mauvaise direction, ils sont susceptibles de ne pas remarquer les animaux. La SAP est seulement efficace lorsque les mammifères marins émettent des vocalisations fréquentes et lorsque ces dernières ne sont pas masquées par le bruit de navires ou d'autres bruits de fond. En ayant comme objectif d'améliorer la détection des mammifères marins par rapport à ce que ces méthodes conventionnelles permettent en ce moment, certaines études au cours des dernières années ont permis d'évaluer l'utilisation de systèmes de caméras à imagerie thermique afin de détecter automatiquement les mammifères marins à la surface de l'océan. L'Automatic Infrared-based Marine Mammal Mitigation System (système automatique d'atténuation à technologie infrarouge pour la détection des mammifères marins, AIMMMS) est un système à technologie infrarouge (IR) offert sur le marché qui utilise un capteur rotatif pour détecter les mammifères marins en mer.

Outre la méthode de détection employée, le niveau d'expérience des observateurs s'est révélé avoir une incidence sur la capacité à détecter et à identifier les mammifères marins en mer.

Ce projet de recherche a été proposé en réponse à une demande de proposition publiée par le Fonds pour l'étude de l'environnement (FEE). Nous avons conçu notre projet de façon 1) à mieux comprendre ce en quoi les méthodes accessibles de détection de mammifères marins en temps réel employées en mer sont comparables ou complémentaires les unes par rapport aux autres et permettent la mise en place de mesures efficaces d'atténuation et 2) à mieux comprendre la relation

entre les OMM et les techniciens de la SAP en ce qui a trait au niveau d'expérience et à la capacité de détecter, de classifier et de localiser efficacement les mammifères marins et d'employer des mesures d'atténuation en conséquence. Nous avons entrepris notre recherche dans les conditions habituelles des relevés sismiques effectués au large du Canada atlantique, mais certains aspects de nos travaux s'appliquent à la surveillance des mammifères marins dans le contexte de relevés sismiques en général.

Notre projet financé par le FEE visait trois objectifs :

1. Optimiser et mettre à l'essai l'AIMMMS en vue de l'utiliser dans le Canada atlantique;
2. Comparer les détections de mammifères marins relevées au moyen de trois méthodes distinctes (soit l'observation visuelle par les OMM, l'utilisation de l'AIMMMS par les OMM aux fins de détection et la SAP);
3. Évaluer la qualité et l'exactitude des détections des mammifères marins relevées par les OMM et les techniciens de la SAP possédant différents niveaux de formation et d'expérience.

MÉTHODOLOGIE DE L'ÉTUDE

En ce qui concerne le premier objectif, il était nécessaire de recueillir des données dans le cadre de deux programmes côtiers : les données recueillies en 2015 ont été utilisées pour optimiser l'algorithme du logiciel d'AIMMMS qui serait utilisé au Canada atlantique, algorithme qui a ensuite été mis à l'essai et optimisé à nouveau au moyen des données recueillies en 2016. Les deuxième et troisième objectifs ont été abordés au moyen des données recueilles dans le cadre d'un programme réalisé à bord d'un navire au large de la Nouvelle-Écosse et de Terre-Neuve en 2017.

Le premier programme a été réalisé à Cape Race à Terre-Neuve. Les données ont été recueillies du 18 juillet au 23 août 2015. Les données thermiques (IR) ont été recueillies en même temps que les observations de mammifères marins réalisées à l'aide d'un théodolite. Les observateurs ont examiné les détections automatiques IR et les ont jumelées aux observations effectuées au moyen du théodolite. On a employé l'ensemble de données pour optimiser l'algorithme du logiciel utilisé pour détecter et classifier les mammifères marins dans le but de s'en servir pendant le régime thermique au large du Canada atlantique en été.

Le second programme sur le terrain a également été réalisé à Cape Race à Terre-Neuve. Les données ont été recueillies du 1er au 31 juillet 2016. Les données thermiques (IR) ont été recueillies en même temps que les observations de mammifères marins réalisées par les OMM à l'aide de jumelles à réticules en suivant un protocole de surveillance habituellement employé en cours de relevé sismique. On a évalué le rendement du système thermique à IR en examinant le nombre et les causes de la détection de faux positifs, ce qui a permis de produire des fonctions de détection pour divers indices de détections et conditions environnementales et de comparer le rendement global du système IR jumelé à un OMM expérimenté.

Le dernier programme sur le terrain a été réalisé au large de la Nouvelle-Écosse et de Terre-Neuve du 31 juillet au 23 août 2017. Les détections de mammifères marins ont été effectuées de façons concomitantes et indépendantes par les OMM, les techniciens de la SAP et par le système thermique à IR automatique. Les OMM et les techniciens de la SAP étaient soit expérimentés, soit non expérimentés, et certains OMM ont reçu l'aide du système IR sous la forme d'alertes automatiques leur étant communiquées au moyen d'une tablette électronique. Les détections de la SAP ont principalement été effectuées au moyen d'un réseau d'hydrophones remorqués, bien que des bouées acoustiques aient parfois été déployées. On a comparé le rendement de détection en fonction du nombre de détections relevées, le moment et la distance par rapport aux détections, le niveau taxinomique dans lequel les animaux ont été classifiés, le nombre de détections de faux positifs (système IR), et les probabilités conditionnelles ont été utilisées pour la comparaison par paire de chaque méthode de détection.

RÉSULTATS ET MANUSCRITS PRODUITS

L'algorithme du logiciel d'AIMMMS a été optimisé pour être utilisé au Canada atlantique. Le module de détection du logiciel intitulé « *TashDetectC* » est accessible de façon ouverte au site suivant : https://gitlab.com/mamaps/TashDetectC (en anglais seulement).

Le système de caméra infrarouge installé sur le navire a été à même de détecter des mammifères marins dans le régime thermique du Canada atlantique pendant l'été. Les détections automatiques relevées par le système de caméra infrarouge étaient des faux positifs dans 84,5 % des cas. Cela équivaut à environ 5,6 faux positifs par heure, en moyenne, pour la totalité de l'enquête. Parfois, un certain nombre d'alertes, souvent occasionnées par des oiseaux ou des dauphins se livrant à la chevauchée en arc, a donné lieu à la détection de plusieurs faux positifs en une seule minute. Il est nécessaire de poursuivre les travaux en ce qui concerne la suppression des fausses alertes pendant ces périodes où l'on relève des taux élevés de faux positifs.

La détection des mammifères marins au moyen de la SAP, aux fins d'atténuation, semble être principalement influencée par l'équipement informatique et les logiciels (contrairement au niveau de formation et d'expérience des techniciens de la SAP). Il y a eu 13 détections de cétacés à fanons au cours d'un exercice d'environ 14 heures mené à l'aide de bouées acoustiques, par rapport à une seule détection de baleine à bosse, laquelle a été détectée au moyen d'un réseau d'hydrophones remorqués, au cours de la totalité de l'enquête.

Nos résultats donnent à penser que le fait d'employer plus d'une méthode de détection des mammifères marins à la fois permettra d'améliorer le rendement global de la détection des mammifères marins. La méthode de SAP et la méthode thermique (IR) se sont avérées efficaces dans l'obscurité; la SAP et les méthodes visuelles se complétaient pendant les périodes de mer de force élevée et de faible visibilité en raison de précipitations (dont le brouillard); et les méthodes thermiques (IR) peuvent être utilisées pour améliorer les méthodes visuelles pendant les périodes de bonne visibilité. Le recours à la SAP et aux méthodes visuelles dans les périodes de bonne visibilité a également permis un plus grand nombre de détections que si l'on avait seulement utilisé les méthodes visuelles.

Les types de mammifères marins détectés et le degré de précision avec lequel ils ont été classifiés en matière d'espèces variaient en fonction des méthodes de détection utilisées. Les OMM et le système thermique (IR) permettaient de détecter efficacement les cétacés à fanons et à dents à la surface de l'eau, mais la détermination des espèces n'était fiable que lorsqu'elle était effectuée par les OMM. Les espèces détectées au moyen des méthodes de SAP différaient selon que l'on utilisait un réseau d'hydrophones remorqués (qui détectait presque exclusivement les cétacés à dents) ou des bouées acoustiques (lesquelles détectaient des cétacés à fanons). La plupart des détections au moyen de la SAP n'étaient pas classifiées au niveau de l'espèce (p. ex. spp globicéphale, dauphin non identifié, Balénoptéridés).

Le niveau d'expérience semble avoir influencé le rendement des détections des OMM, et ce, de différentes façons. L'OMM non expérimenté a efficacement surveillé une étendue plus petite que la zone de visibilité entourant le navire, a détecté moins de mammifères marins et était moins susceptible d'identifier ces mammifères marins au niveau de l'espèce, par rapport aux OMM expérimentés. L'OMM non expérimenté mettait généralement plus de temps à détecter les mammifères marins, et détectait d'abord ces derniers lorsqu'ils se trouvaient plus près du navire, comparativement aux OMM expérimentés. Ces conclusions donnent à penser que la mise en œuvre de mesures d'atténuation liées à la réduction au minimum de l'exposition au bruit (provenant de canons à air et d'autres sources de bruit) est moins efficace lorsque la surveillance est effectuée par des OMM non expérimentés.

Les données issues du programme à Cape Race en 2016 ont été combinées aux données recueillies pendant des projets non financés par le FEE, exécutés dans les zones de recherche tropicales et subtropicales, et analysées dans le but d'étudier le rendement du système de détection automatique des baleines en fonction d'un large éventail de conditions environnementales. Les résultats des enquêtes combinées ont été utilisés pour préparer un manuscrit en attente de publication (Zitterbart et coll. en attente de publication). L'analyse de perceptibilité thermique fondée sur les données recueillies pendant l'exécution du programme à Cape Race en 2015 a également été incluse dans le manuscrit de Zitterbart et coll. (en attente de publication). Les données issues de l'exécution du programme à bord d'un navire en 2017 ont été utilisées pour préparer un manuscrit sur la comparaison, en mer, des trois méthodes qui sont actuellement utilisées ou dont l'utilisation a été proposée, pour détecter les mammifères marins pendant les programmes de surveillance réalisés en cours de levés sismiques. Ce manuscrit est également en attente de publication (Smith et coll. en attente de publication).

Les résultats et les conclusions de cette étude ont également été utilisés pour recommander des modifications à certains éléments de l'Énoncé des pratiques canadienne d'atténuation des ondes sismiques en milieu marin. Le rapport de recommandation a été rédigé par Smith et Moulton et est inclus ici sous l'Annexe C.

RECOMMANDATIONS AUX FINS D'ÉTUDES FUTURES

Nous recommandons principalement que les comparaisons futures des méthodes de détection, lorsque l'application prévue concerne la surveillance pendant les relevés sismiques, soient effectuées à bord de navires de relevés sismiques. Cette pratique permettra de s'assurer qu'une plateforme stable et adéquatement élevée est accessible pour l'installation d'une caméra infrarouge et que les détections de la SAP sont effectuées selon le réglage acoustique approprié, c'est-à-dire correspondant aux pulsations des canons à air.

L'une des faiblesses de notre ensemble de données de 2017 est l'absence d'estimations de la distance quant aux détections IR, laquelle absence est attribuable à la fréquence à laquelle il y a eu dépassement de la capacité de la suspension à cardan à stabiliser la caméra infrarouge. Cette situation nous a empêché de calculer les fonctions de détection pour le système IR, ce qui nous aurait permis de comparer plus concrètement les méthodes. Faire en sorte que, dans le cadre d'études futures, une plateforme stable soit fournie pour le système IR permettra d'éviter ce problème.

Un grand nombre de détections automatiques IR de faux positifs a été relevé pendant le programme réalisé sur la côte et sur le navire. Par conséquent, il est nécessaire d'approfondir l'étude du système thermique (IR) et d'améliorer ce dernier de telle sorte que la communication des alertes diffusées aux OMM en temps réel soit utilisée efficacement en tant qu'outil de surveillance.

Nos résultats donnent également à penser que les méthodes de SAP autres que l'utilisation d'un réseau d'hydrophones remorqués, jumelées à l'analyse des mouvements de l'objectif, devraient être étudiées comme moyen de détection acoustique des mammifères marins pendant les relevés sismiques. Il serait utile à l'avenir d'effectuer des études sur des méthodes de formation de faisceaux, des méthodes de relèvements croisés ou des réseaux d'hydrophones capables de localiser les mammifères marins dans un espace tridimensionnel dans le contexte de l'atténuation.

1.0 Introduction

This final report summarizes our research activities for project #2014-03S, funded by the Environmental Studies Research Fund (ESRF). The overall goal of our project was to compare three methods that are, or could be, used to detect marine mammals as part of a monitoring program during seismic surveys conducted in waters offshore Atlantic Canada.

1.1 Background

1.1.1 Seismic Sound and Marine Mammals

Sound in the ocean is generated by a variety of natural and anthropogenic sources, of which marine seismic surveys are recognized as a major contributor (Klinck et al. 2012a; Miksis-Olds and Nichols 2016; Haver et al. 2017). During marine seismic surveys, acoustic energy is typically generated by a high-energy sound source positioned near the ocean surface and primarily projected downward through the water column. Reflected acoustic energy is then recorded by hydrophones and used to map the underlying geology of the seafloor. The sound source, usually an airgun array, is towed behind a survey vessel following predetermined survey lines (Gisiner 2016). Impulsive sounds generated during seismic surveys have been documented to cause behavioural responses in marine mammals and could result in hearing impairment or injury (e.g., Richardson et al. 1995; Gordon et al. 2003; Romano et al. 2004; Southall et al. 2007 & 2019; Blackwell et al. 2015; Nowacek et al. 2015; Stone 2015a; Erbe et al. 2018).

1.1.2 Mitigation of Seismic Sound in Canada

The *Statement of Canadian Practice with respect to the Mitigation of Seismic Sound in the Marine Environment* (the "Statement"; DFO 2008) specifies the minimum mitigation requirements for the planning and conduct of seismic surveys in ice-free marine waters in Canada in order to reduce the potential effects of seismic surveys on marine life. Mitigation actions in the Statement are designed to minimize the probability of marine mammals being exposed to sounds that are loud enough to cause hearing impairment and are often triggered when marine mammals are observed entering, or about to enter, a safety zone (SZ) based on horizontal distance from the sound source (DFO 2008). Specific mitigation actions are employed only for particular species or species groups. For example, shut-downs of the airgun array are typically required only for species listed as threatened or endangered on Schedule 1 of the *Species at Risk Act* (SARA). Essential elements of effective mitigation therefore include the ability to reliably detect and localize marine mammals in a timely manner, and in some instances, to identify marine mammals to species or some other taxonomic level.

1.1.3 Methods to Detect Marine Mammals During Seismic Surveys

A "traditional" visual watch conducted by marine mammal observers (MMOs) is the primary marine mammal monitoring method required by the Statement. However, MMOs are restricted to daylight hours and detection of marine mammals is often negatively affected by fog and sea state, particularly in Atlantic Canada. Multiple MMOs are needed when seismic programs are long, involve multiple vessels (e.g., wide azimuth seismic surveys), and occur in areas with long daylight periods. With the objective of overcoming constraints due to the absence of watches at night, observer fatigue, and berth limitations aboard seismic vessels, the Alfred-Wegener-Institute (AWI) developed an automatic whale detection software which continuously and in real-time screens a ship's entire perimeter for whale blows on the basis of thermal (infrared, IR) video data provided by a thermographic scanner. This automatic whale detection software is a component of the Automatic Infrared-based Marine Mammal Mitigation System (AIMMMS) available from Rheinmetall Defence Electronics GmbH (RDE). While thermal imaging techniques have previously been used to detect large marine mammals including minke (*Balaenoptera acutorostrata*), fin (*B. physalus*), blue (*B. musculus*), humpback (*Megaptera novaeangliae*), gray (*Eschrichtius robustus*) and sperm (*Physeter macrocephalus*) whales (Cuyler et al. 1992; Perryman et al. 1999; Zitterbart et al. 2013; Horton et al. 2017); and dolphins and pilot whales (*Globicephala* sp.; Baldacci et al. 2005), the AIMMMS is the first thermal imaging system to simultaneously provide detection ranges, circumferential vision, and the sea-worthiness (Zitterbart et al. 2013) required during typical seismic monitoring programs. The automatic whale detection software developed by AWI and used in the AIMMMS was developed and tested in polar and subpolar waters (Zitterbart et al. 2013).

Passive acoustic monitoring (PAM) is another "traditional" technique used to detect marine mammals during seismic surveys (Weir and Dolman 2007; Compton et al. 2008; Verfuss et al. 2018). The Statement requires the use of cetacean detection technology (such as PAM) during the pre-ramp-up watch when the SZ is not fully visible and when the survey area occurs in critical habitat for a vocalizing cetacean listed on Schedule 1 of the SARA. The Statement also requires the use of PAM or other cetacean detection technology in areas where a vocalizing cetacean is expected to be encountered if that species, as predicted during the EA process, may incur significant adverse effects (DFO 2008). The Statement does not specify minimum requirements for the hardware components of PAM systems. However, towed hydrophone arrays (THA) have usually been employed (e.g., RPS Energy Canada 2014). Depending on equipment configuration, and sound produced by the survey vessel and airguns, THA may not be able to detect the low-frequency vocalizations of some baleen whales, a limitation that is explicitly recognized in the mitigation guidelines of some other countries (e.g., JNCC 2017; NMFS 2018). For PAM to be effective, marine mammals must vocalize relatively frequently during monitoring periods, and vocalizations must be detectable. The effectiveness of PAM in detecting marine mammal vocalizations is not negatively affected by fog or darkness, as with visual methods, but may be somewhat lessened during rainfall or high sea states due to an associated increase in background noise.

1.1.4 Effect of Observer Experience Level on Marine Mammal Detections

The ability to detect (and identify) marine mammals has been shown to be influenced by observer experience level. Inexperienced observers are known to detect fewer animals than experienced observers (Barlow et al. 2006; Wright et al. 2016). Inexperienced observers are also less likely to classify a detected marine mammal to species level (Barlow et al. 2006), thereby impacting the effectiveness of mitigation when actions are prescribed only for certain species. Stone (2015b) compared MMOs with and without prior relevant experience, and found that experienced MMOs had higher sighting rates, could detect animals at greater distances, and recorded a wider range of behaviours than MMOs without prior marine mammal experience. The Statement does acknowledge that observer experience level is likely to influence the effectiveness of mitigation actions by making reference to the use of a "qualified MMO", though it provides no guidance on what this means, nor does it specify requirements for PAM operator qualifications (DFO 2008).

1.1.5 Motivation for This Study

Our research is motivated by the current and likely future extent of marine seismic survey activity, in combination with the potential to negatively affect marine mammals. Given that the ability to detect marine mammals is influenced by the detection method used and observer experience level, we designed our research project; 1) to improve our understanding of how available real-time marine mammal detection methods employed at sea compare and complement each other, and enable effective mitigation action, and 2) to better understand the relationship between MMO and PAM operator experience level and the ability to effectively detect, classify, and locate marine mammals and employ subsequent mitigation measures. Our research was undertaken in conditions typically experienced during seismic surveys conducted offshore Atlantic Canada, however, aspects of our work are applicable to marine mammal monitoring during seismic surveys in general.

1.2 Study Objectives

Our ESRF project had three objectives

1. Optimize and test the use of the AIMMMS for use in Atlantic Canada;
2. Compare marine mammal detections made using three different methods (i.e., MMOs making visual observations, MMOs utilizing AIMMMS to make detections, and PAM);
3. Evaluate the quality and accuracy of marine mammal detections made by MMOs and PAM operators with different levels of training and experience.

The first objective required that data be collected over the course of two shore-based field programs: data collected in 2015 were used to optimize the AIMMMS software algorithm for use in Atlantic Canada; which was then tested and further optimized using data collected in 2016. The second and third objectives were addressed using data collected during a vessel-based field program offshore Nova Scotia and Newfoundland in 2017.

2.0 IR Camera Optimization, Cape Race 2015

Our first field program was designed to collect data needed to optimize the automatic whale detection software developed by AWI for use in the thermal regime offshore Atlantic Canada, where seismic surveys typically occur during May to October, and sea surface temperatures range from ~ 0 to 17°C (DFO 2018). The AIMMMS had previously been deployed only in relatively cooler polar and subpolar oceans (i.e., water temperatures between -2 to +10°C; Zitterbart et al. 2013), and it was unknown if, and how well, it would detect marine mammals during seismic surveys conducted in relatively warmer ocean surface waters. Following the example of Boebel and Zitterbart (2014), a shore-based field program was determined to be the most economically and logistically feasible means to collect an adequate amount of data (i.e., thermal imagery and concurrent visual observations of marine mammals) for algorithm optimization.

The thermal perceptibility analysis based on data collected during the 2015 Cape Race field program is included in a recently submitted manuscript that investigates the performance of the automatic whale detection system in a wide range of environmental conditions (Zitterbart et al. submitted).

2.1 Study Area and Logistics

Cape Race, Newfoundland was determined to be an appropriate location for our first field program primarily because it is a place where the likelihood of spotting passing marine mammals in summer months is high. Cape Race is ~140 km south of St. John's, and is located near (~21 km east) the town of Portugal Cove South. (Figure 1). The Cape Race Lightstation lands (DFRP #00004; List of Lights #1; Figure 2) offer a vantage point ~26 m above sea level (ASL; determined using handheld GPS), and have suitable areas to safely and securely set up the equipment needed for data collection (i.e., thermal imaging camera, visibility sensor, and theodolite[1]), and accommodate an office trailer that housed the computers and equipment needed for collection and screening of the thermal imagery data at the field site (Figure 3). Fisheries and Oceans Canada (DFO) gave permission to LGL to the use Cape Race Lightstation Lands use for the purpose of data collection for this project as outlined in the Licence to Occupy issued for July and August 2015.

The office trailer was delivered to Cape Race on 9 July 2015 and approximately one week of preparations at the field site were required prior to arrival of the field team (e.g., electric hook up, gear delivery and set up). A start-up meeting for the field team was held on 16 July and personnel were mobilized to the field site that afternoon. Scientific equipment setup and training began on 17 July. Data were collected during 18 July to 23 August, and demobilization occurred on 24 August 2015.

2.2 Data Collection

Three primary types of data were collected at Cape Race in 2015: environmental conditions, visual marine mammal sightings and automatic IR detections.

[1] A theodolite is an optical instrument used to measure horizontal and vertical angles to a target, which is viewed through a moveable telescope. The measure angles can be used to determine the target's geographic location.

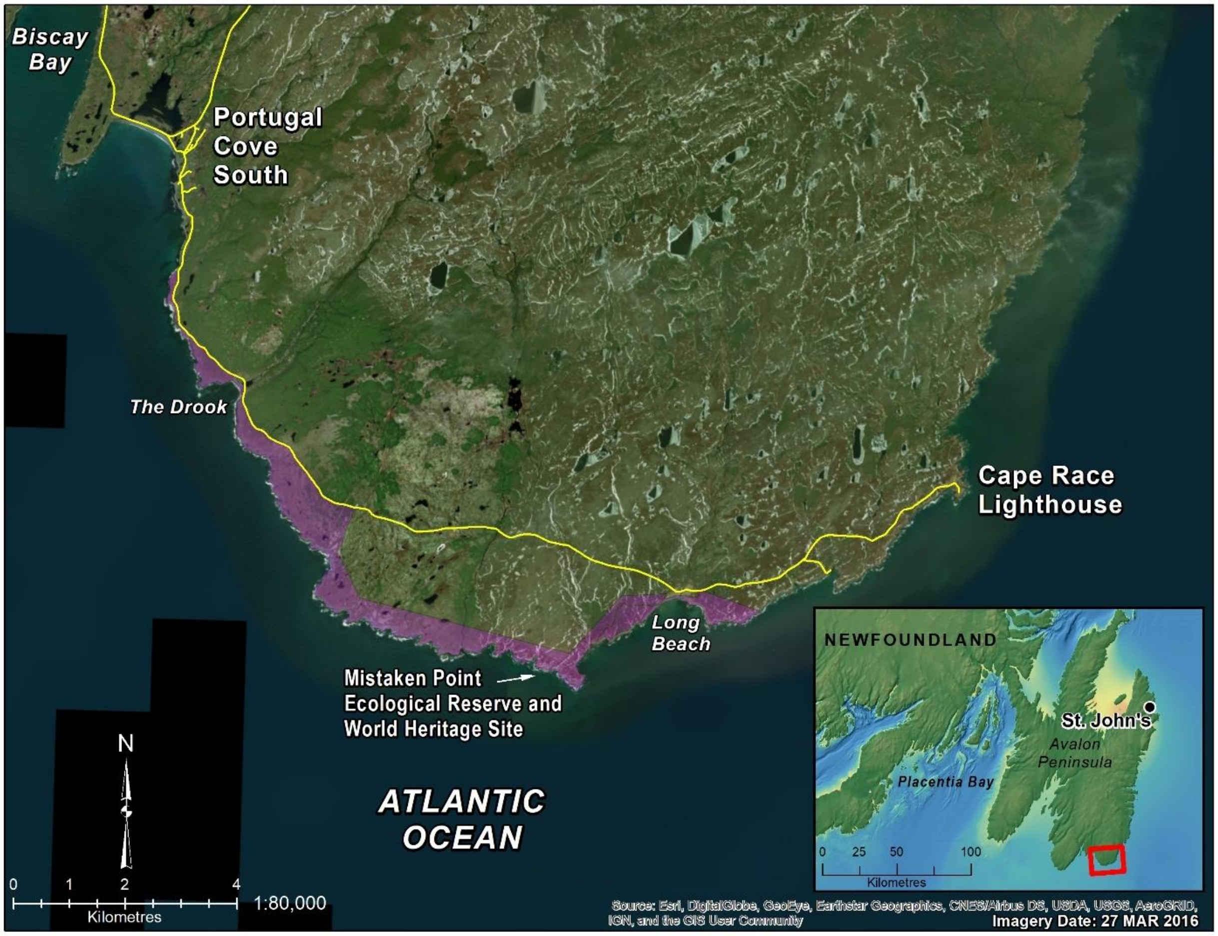

FIGURE 1. Approximate location of the Cape Race Lightstation lands. Inset map shows location of Cape Race relative to St. John's and the Avalon Peninsula, Newfoundland.

FIGURE 2. Cape Race Lightstation lands (DFRP #00004; List of Lights #1), Newfoundland. Data was collected from a vantage point behind the red building at far right (i.e., the old foghorn building).

FIGURE 3. (A) An office trailer was delivered to Cape Race for use by the field team because the old lightstation land buildings were not fit for human occupation; the trailer was anchored to cement jersey blocks in the event of high winds. (B) The trailer was used to store field equipment and computers used to log thermal images; thermal imagery was also reviewed in the trailer.

2.2.1 Environmental Data

Environmental data were recorded by the field team during each observation shift. Environmental data were recorded at the start and end of every observation shift, every half hour, and whenever conditions changed. The following data were collected: number of icebergs present, precipitation type (none, rain, drizzle, fog, snow), cloud cover (%), glare (none, low, moderate, severe; and approximate portion of study area affected), Beaufort wind force (BF; based on appearance of water surface; Appendix A), sightability (subjective judgement of overall viewing conditions classified as being poor, moderate, good, or excellent), and visibility (estimate of how far could be seen; to a maximum of 10 km).

Visibility was also measured using a Vaisala FS11 visibility sensor. The sensor was mounted to the side of the old foghorn building (Figure 4) and logged visibility at 15 second intervals.

Air temperature and wind (speed and direction) data were downloaded from the nearby Environment Canada weather station at 46.660 N 53.076 W (http://climate.weather.gc.ca/index_e.html). Data were logged at 30-minute intervals. Water temperature data were downloaded from the Nickerson Bank buoy at 46.443 N 53.392 W (http://www.ndbc.noaa.gov/station_realtime.php?station=44251). Data were logged at 60-minute intervals.

FIGURE 4. The Vaisala FS11 visibility sensor was mounted to the side of the old foghorn building at Cape Race, 2015.

2.2.2 Marine Mammal Observations

Observations of marine mammals were made from a vantage point overlooking the water that was a safe distance from the cliff edge (Figure 5). Observers worked in pairs to make fixes (i.e., recorded horizontal and angular bearings to sighting location) on marine mammals using a Sokkia CX-105 Total Station (electronic theodolite with telescope magnification of 30 x) connected to a data logger (Mesa Rugged Notepad from Juniper Systems, MAGNET Field data collection software). One observer sighted marine mammals using the total station /theodolite while the second observer made fixes using the data logger and recorded details for each sighting; pairs of observers took alternating observation shifts throughout the day. Observers did not attempt to make fixes on all marine mammals sighted, but instead focused on making accurate locations of as many marine mammals as possible. Ensuring the accuracy of the fixes so that visual sightings could be compared with thermal images was prioritized over documenting the presence of all marine mammals in the viewable area.

Visual observations were made from beside the FIRSTnavy platform (§ 2.2.3) to avoid creating vibrations that would interfere with sensor function.

FIGURE 5. Data were collected from a vantage point behind the old foghorn building at Cape Race, NL. Observers worked in pairs and made fixes on marine mammals using a theodolite.

The following data were recorded for each fix/sighting:

1. Local time: each fix was automatically time stamped by the theodolite. All times in this report are local Newfoundland Standard Time (i.e., UTC – 3.5 h).
2. Location: horizontal (azimuth) and vertical (declination) angles were measured by the theodolite and were used to triangulate geographic locations of fixes.
3. Species: minke whale, humpback whale, unidentified baleen whale, unidentified whale, harbour porpoise (*Phocoena phocoena*) and harbour seal (*Phoca vitulina*).
4. Group size, including group size 1. The distance between individuals within a group varied according to the size of the species being observed such that larger species had more space between individuals within a group. The following group definitions were used:
   - Dolphins and killer whales – individuals within 10 m of each other.
   - Baleen whales – individuals within approximately two body lengths of each other. Whales observed lingering in the same areas (e.g., feeding) were categorized as being a single group even when individuals were more than 2 body lengths apart.
   - Seals – individuals within approximately 5 body lengths of each other.
5. Group ID: sequential numbers were assigned to new groups (where group size may = 1) such that repeat sightings of the same group retained the same number.
6. Cue that alerted observer to the presence of the marine mammal: blow, breach, body (e.g., back at surface), splash, flukes. Blows were categorized according to

strength/opacity (e.g., 0 = barely visible, translucent; 1 = very weak and transparent; 2 = easily visible; and 3 = very dense and opaque).

7. Behaviour: travelling (directed movement), milling (non-directed movement), resting with back exposed (also known as "logging"), resting submerged, porpoising, fluking, diving, flipper slapping, breaching, presumed or observed feeding, unknown. For group size >1, the behaviour that described what the majority of animals in the group was doing was recorded. Note: behaviour data were collected to provide general information about marine mammal behaviour in the area and are not suitable for detailed behavioural analysis.

During periods of reduced visibility (i.e., <500 m) and wet conditions, observers used reticle binoculars (Fujinon 7 x 50 FMTRC-SX) and the naked eye to estimate the location and distance of marine mammals sighted. Observations were made during a wide variety of environmental conditions (e.g., daylight hours with good to poor visibility, variable sea states with and without glare, and in a variety of fog "thicknesses") to collect additional thermal imagery that could be used to explore the application of this technology when visibility is poor.

2.2.3 Thermographic Data

Thermographic data were collected using a rotating IR camera (FIRSTnavy sensor, a component of the AIMMMS, RDE). The FIRSTnavy sensor is cooled to 84 K using a Sterling cooler; scans 360° horizontal x 18° vertical at 5 revolutions per second, providing a 5-Hz video stream of the radiance field of view at horizontal and vertical resolutions of 0.05°/pixel and 0.03°/pixel, respectively.

The FIRSTnavy sensor was situated on a purpose-built platform near the cliff edge at Cape Race; the platform was level and allowed for the sensor tripod to be anchored securely to the ground (Figure 6). Active stabilization (i.e., gimbal[2]) of the thermal imager was not required because data were collected from land. The field of view (FOV) for the FIRSTnavy sensor was ~198° facing southeast (Figure 7).

The FIRSTnavy sensor was operated throughout the field program during the hours when the field team was present at the study site. Thermographic data were recorded when observers were on effort, as well as during some periods of darkness and heavy fog when no observers were on effort.

Thermographic data acquisition and processing were performed with custom developed software (*Tashtego*). *Tashtego* utilizes a multi-step detection and classification approach that was developed with data collected during previous expeditions (Zitterbart et al. submitted). *Tashtego* made automatic detections by tracking contrast in radiance in the IR video stream and applying a set of heuristic rules designed to reduce the number of non-marine mammals detected. Automatic detections are packaged as 6-second video "snippets" (420 pixels x 101 pixels x 30 frames; Figure 8) of the thermal anomaly identified as potentially being created by a marine mammal.

[2] A gimbal is a pivoting support that allows an object mounted on it to remain upright relative to the horizon.

FIGURE 6. The FIRSTnavy sensor was situated on a level platform near the cliff edge at Cape Race, NL.

FIGURE 7. The field of view for FIRSTnavy sensor at Cape Race, NL was 198° facing approximately southeast.

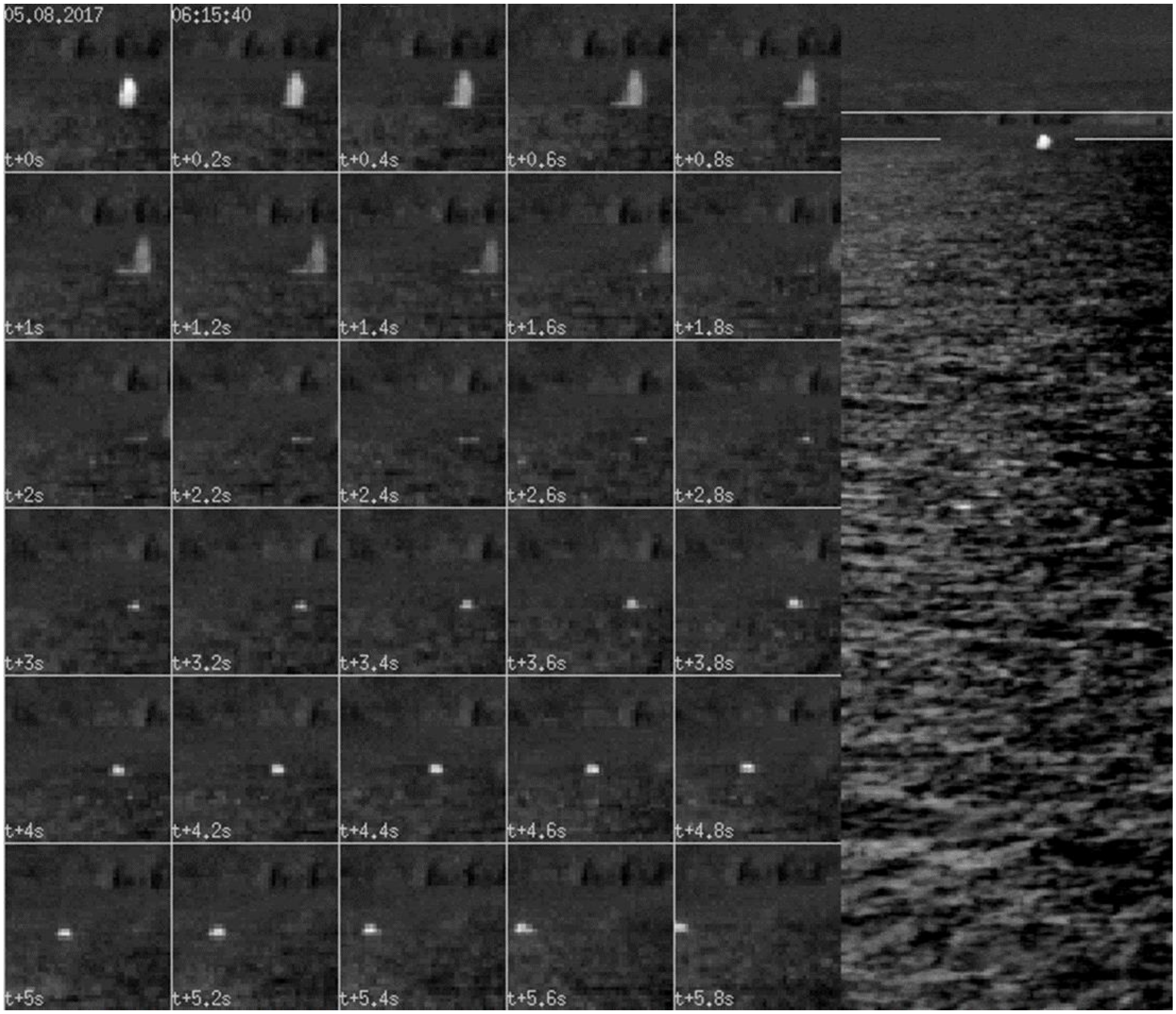

FIGURE 8. Thermal (IR) video "snippet" of an automatic marine mammal detection. Snapshots at 0.2 second intervals are displayed at left; a six-second video snippet of the detection plays on repeat in the vertical panel at right. A blow (tall white image) is visible in the vertical panel and in snapshots 0–1.6 s, followed by the back of the whale (smaller white image) in snapshots 2.8–5.8 s.

2.3 Data Processing and Analysis

2.3.1 Informed Manual Review of Automatic IR Detections

In order to identify which automatic IR detections were true detections of marine mammals and which were false positives, all automatic detections were human-verified. Automatic IR detections were reviewed and classified by observers after each observation shift. Observers were "informed" during their review of the automatic IR detections, i.e., observers compared concurrent visual sightings data (including time and location of sighting) with the thermal (IR) video snippets. Snippet review was performed using *Fedallah* software, specially designed by AWI to allow for *ad lib* navigation (forward, pause, backward, zoom, rotate) within the thermal (IR) data stream.

Automatic detections classified as being created by a whale were then further classified as being either an aerial display (breach, half-breach, pectoral slap, tail slap, back) or whale blow. Classification was subjective and based on the overall appearance of the thermal anomaly in the six-second video snippet. The *Fedallah* software was used to annotate the data snippets accordingly, which were then saved as .svmdb database files. The resulting data were used to provide training data for the machine learner[3] to optimize the automatic detection algorithm. Thermal (IR) was conducted immediately (i.e., no later than 24 h) after the observation shift.

Because the field team received hands-on training with the AIMMMS while in the field, data collected prior to 24 July were considered "practice data" and were not included in the dataset used to optimize the AIMMMS software. Data used to optimize the AIMMMS software were collected from 24 July to 23 August 2015.

2.3.2 Thermal Perceptibility

We define thermal perceptibility as how well a whale cue (e.g., blow, splash, back, breach) in the thermal (IR) data stream can be perceived by an informed human observer. We use this metric as a means to quantify the detection performance of the thermal (IR) system. Performance was quantified as the conditional probability of perception (i.e., P(IR | VIS) = the probability that a cue was perceived in the IR data stream given that it was sighted using the theodolite), and the influence of cue type, Beaufort wind force, and humidity was examined.

2.4 Results

In 2015, thermal (IR) data were collected for the purpose of optimizing the automatic detection algorithm for future applications. Analysis of these data was therefore limited, and results are presented here as summaries of marine mammal sightings made using the theodolite, and the results of the manual video snippet review.

2.4.1 Visual Sightings of Marine Mammals

A total of 1,156 theodolite fixes were made on marine mammals over the course of the field season (Table 1). Fixes were made on humpback whales (n = 981), minke whales (n = 138), harbour porpoise (n = 10), harbour seals (n = 2), unidentified baleen whales (n = 16), and unidentified whales (n = 9). Of these, 1,114 fixes comprised the dataset used to optimize the AIMMMS software.

[3] Machine learning is a subfield of computer science that evolved from pattern recognition. It gives computers the ability to learn without being explicitly programmed (Wikipedia 2019).

TABLE 1. Observation effort and marine mammal sightings made with the theodolite at Cape Race, 2015.

Date	**Effort (h)**	**No. of Theodolite Fixes**			**Notes**
		Humpback	**Minke**	**Other**	
"Practice Data"					
18-Jul-15	2.5		2	1	
19-Jul-15	2.5	1	24	1	
20-Jul-15	2.5				humpbacks sighted, no fixes due to rain
21-Jul-15	0				AIMMMS modifications; rain
22-Jul-15	3.5	13			
23-Jul-15	5.7				too windy for theodolite – no fixes
Totals	16.7	14	26	2	
Data for Software Optimization					
24-Jul-15	7	126	5	9	
25-Jul-15	7	113	7		
26-Jul-15	5.7	162	2		
27-Jul-15	0				IR data review for MMOs
28-Jul-15	5.2	84			
29-Jul-15	7	116			
30-Jul-15	0				rain and heavy fog
31-Jul-15	7	125	5	12	
01-Aug-15	5	30			
02-Aug-15	6	76	2		
03-Aug-15	0				IR data collected at night in fog
04-Aug-15	4.5	40	4		
05-Aug-15	5.5				too foggy for theodolite
06-Aug-15	2.2				too foggy/rainy for theodolite
07-Aug-15	7				
08-Aug-15	6.5	93	4	3	
09-Aug-15	6.25				humpbacks sighted, no fixes, rough seas
10-Aug-15	1				AIMMMS malfunction
11-Aug-15	0				AIMMMS cable on order
12-Aug-15	1.3				fog; AIMMMS cable on order
13-Aug-15	0				fog & wind; AIMMMS cable on order
14-Aug-15	0				fog & rain; AIMMMS cable on order
15-Aug-15	8		2	2	
16-Aug-15	7.5		11	3	
17-Aug-15	7.5		3	1	
18-Aug-15	7.5		1	4	
19-Aug-15	7.5	2	8		
20-Aug-15	7.5		44		
21-Aug-15	6		9		
22-Aug-15	7.5		5	1	
23-Aug-15	3.5				no marine mammal sightings
Totals	146.7	967	112	35	Grand total = 1114

2.4.2 Automatic IR Detections

Video snippets of automatic IR detections were reviewed by observers, and thermal anomalies that matched 700 of the 1,114 theodolite fixes were perceptible in the IR data (Table 2). Of the 414 fixes not matched, 366 were not found in the images and 48 were undetermined (i.e., uncertain). Observers perceived thermal anomalies that matched ≥50 % of the minke and humpback whale sightings made using the theodolite. No thermal anomalies were found to match the harbour porpoise sightings (Table 2).

TABLE 2. Comparison of automatic IR detections and visual sightings made at Cape Race, 2015.

Species	No. of Theodolite Fixes	No. (and %) of Fixes Perceptible in IR Data		
		Yes	No	Uncertain
Minke Whale	112	56 (50.0)	43 (38.4)	13 (11.6)
Humpback Whale	967	640 (66.2)	296 (30.6)	31 (3.2)
Unidentified Baleen Whale	16	1 (6.2)	14 (87.5)	1 (6.2)
Unidentified Whale	9	3 (33.3)	4 (44.4)	2 (22.2)
Harbour Porpoise	10	0	9 (90.0)	1 (10.0)
Totals	1114	700 (62.8)	366 (32.8)	48 (4.3)

2.4.3 Thermal Perceptibility

The maximum distance at which humpback whale cues were perceived in significant numbers was 5 km. Perceptibility decreases substantially at 2–3 km (Figure 9). At 3–4 km, strong blows were ~1.3 times as likely to be detected as weak blows (Figure 9A). Perceptibility is minimally affected by wind force up to 3 km, and relative humidity shows no obvious correlation with perceptibility (Figures 9B & 9C, respectively). Minke whales could be perceived at up to 800 m in distance, but with a much-reduced probability of perception ranging between 0.2 and 0.5.

2.5 Discussion

2.5.1 Thermal (IR) System Performance in Atlantic Canada

Even before optimization of the detection algorithm, whales were successfully perceived by the thermal (IR) system, indicating that this technology will be useful for detecting whales in the thermal regime offshore Atlantic Canada in summer. The probability of perception was >80 % for strong and weak blows at distances ≤3 km. This suggests that the IR system could be a useful tool for MMOs on seismic vessels as it could allow for an increased awareness of, and potentially the ability to track, whales in the vicinity of the safety zone.

2.5.2 Marine Mammal Species Perceived by the Thermal (IR) System

Sighting cues produced by both minke and humpback whales were perceptible in the thermal (IR) data stream, though cues produced by harbour porpoises were not. This is in agreement with previous studies where thermal (IR) cameras have primarily been used to detect large marine mammals including minke, fin, blue, humpback, gray, and sperm whales (Cuyler et al. 1992; Perryman et al. 1999; Zitterbart et al. 2013; Horton et al. 2017). However, dolphins and pilot whales have also been detected with some systems (e.g., Baldacci et al. 2005).

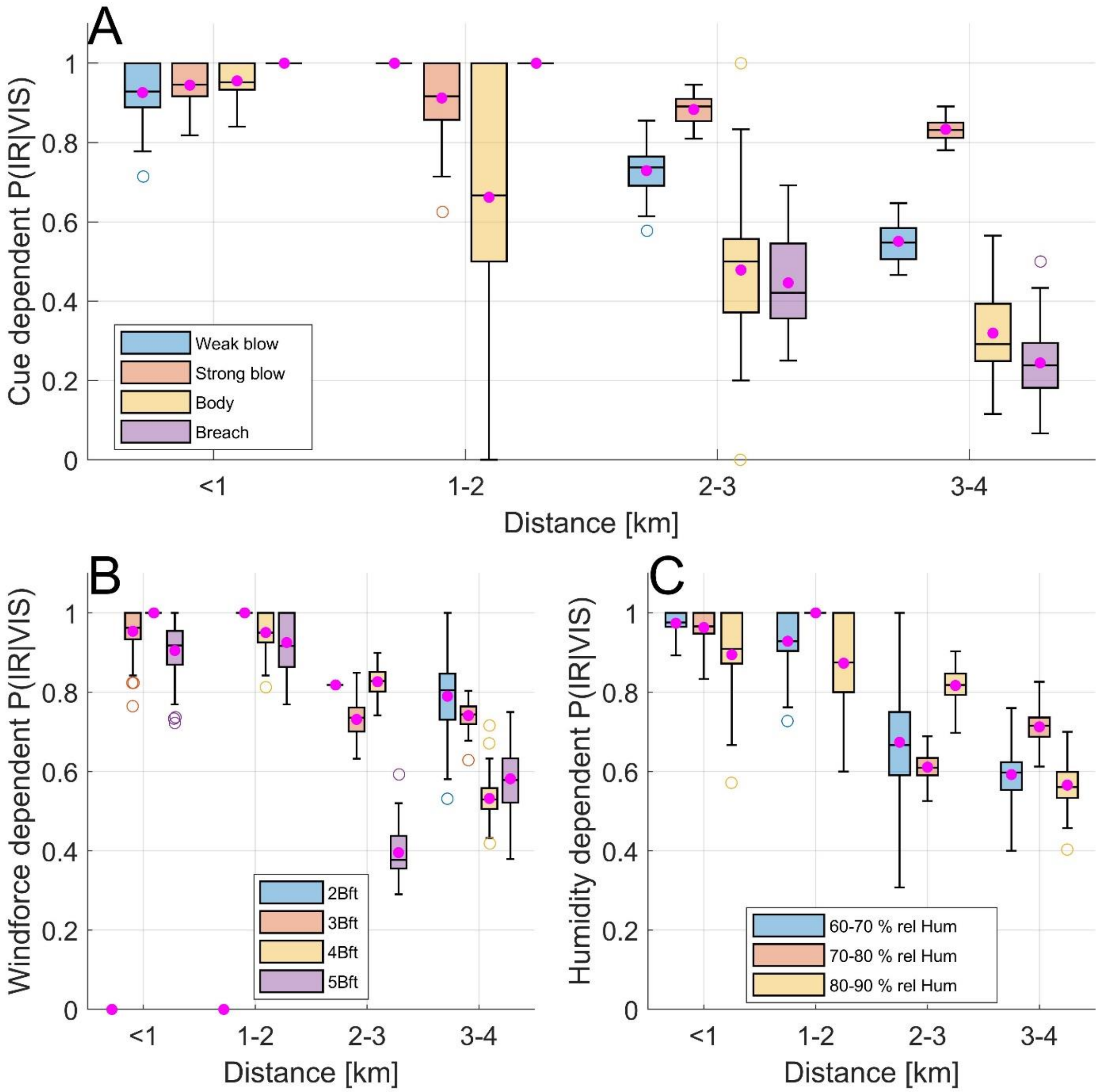

FIGURE 9. Probability of perception of different humpback whale cues (A), and the influence of Beaufort wind force (B) and relative humidity (C). Error bars denote standard deviations obtained by bootstrapping.

2.5.3 Considerations for Use of the Thermal (IR) System in a Mitigation Context

Blows were subjectively classified as strong or weak by the observer in the field. Though we did not attempt to follow individual whales for the purpose of making detailed behavioural observations at Cape Race, it is likely that the strong blows were produced by whales surfacing immediately following longer dives, while weak blows were produced during shallower dives and surface activity. Because whales would produce both weak and strong blows over time, the decrease in perceptibility of weak blows (though still >50 %, see Figure 9A) at distances greater than 3 km, is not overly concerning in a mitigation context especially because mitigation and monitoring zones generally have radii <3 km (DFO 2008, Verfuss et al. 2016).

That perceptibility was shown not to be influenced by relative humidity (see Figure 9C), and was only reduced in the presence of wind force >BF4 (relative to perceptibility at wind force ≤BF3) at distances >3 km (see Figure 9B), suggests that this technology is suitable for mitigation applications in a variety of environmental conditions.

The detection performance on the thermal (IR) system needs to be evaluated in order to assess its utility in assisting with marine mammal monitoring during seismic surveys in Atlantic Canada. The next step towards reaching this goal is to optimize the detection algorithm. An adequate number of marine mammal visual sightings matched to thermal anomalies perceptible in the IR data stream were collected during the 2015 field program at Cape Race to allow for this optimization. The software was optimized prior to the 2016 field program at Cape Race, and detection functions were generated for the thermal (IR) system using the 2016 dataset (§ 3.5.7), which was collected using the optimized software. The optimized software detection module *TashDetectC* has been published in the software repository https://gitlab.com/mamaps/TashDetectC.

3.0 IR CAMERA TESTING, CAPE RACE 2016

The goal of the 2016 field program was to test the newly optimized automatic whale detection software developed by AWI for use in the thermal regime offshore Atlantic Canada. Once again, a shore-based field program was determined to be the most economically and logistically feasible means to collect an adequate amount of data to meet this goal.

Data from the 2016 Cape Race field program were combined with data collected during non-ESRF projects conducted at tropical and sub-tropical field sites and analyzed to investigate the performance of the automatic whale detection system in a wide range of environmental conditions. The results of the combined investigation have been submitted for publication (Zitterbart et al. submitted).

3.1 Study Area and Logistics

The field site at the Cape Race Lightstation Lands was used again for the field program in 2016 (§ 2.1). A Licence to Occupy the Lightstation Lands for this project during June and July 2016 was granted to LGL by DFO.

An office trailer used to house equipment and the field team while on site was delivered to Cape Race on 13 June 2016. Preparations at the field site required prior to arrival of the field team (e.g., electric hook up, gear delivery and setup, MMO booth construction) were completed over the next week and a half. A start-up meeting for the field team was held on 27 June and personnel were mobilized to the field site that evening.

3.2 Data Collection

Three primary types of data were collected during the 2016 field program: environmental conditions, visual marine mammal sightings and automatic IR detections. Data were also collected during a short calibration study (§ 3.3) conducted to allow us to investigate the accuracy of the three methods used to estimate sighting distances to marine mammals during the 2016 program.

3.2.1 Environmental Data

Environmental data were recorded during 2016 as was done in 2015 (§ 2.2.1).

3.2.2 Marine Mammal Observations

Marine mammal observations were made during daylight hours from purpose-built observation booths located on either side of the old foghorn building at Cape Race, NL. The FIRSTnavy sensor was located on a platform between the MMO booths (Figure 10A). MMO booths were placed on either side of the foghorn building so that MMOs made observations independently of one another.

MMOs followed a monitoring protocol that mimicked what would be done during a seismic survey. MMOs scanned the water's surface for marine mammals using binoculars and the naked eye (Figure 10B). The distance of marine mammals from the MMO booth was predominantly determined using reticle binoculars (Fujinon 7 x 50 FMTRC-SX.). However, when marine mammals were too close to shore to use the reticle binoculars, distance was estimated by eye or by using a clinometer (Suunto PM-5/360 PC). MMOs attempted to determine the distance and bearing of all marine mammals sighted. Sightings were recorded using paper datasheets.

All MMOs were experienced in marine mammal monitoring, i.e., all had previously worked as MMOs during seismic surveys offshore Atlantic Canada. During some shifts, MMOs received assistance from the thermal (IR) system. Alerts (audible and visual) were relayed in near real-time to assisted MMOs (aMMOs) via an iPad application developed specifically for this purpose (Figure 11A). An audible "beep" was used to alert the aMMO to the arrival of a new automatic IR detection on the iPad. Visual alerts consisted of 6-second video snippets of thermal (IR) data, along with estimated distance and bearing values to the location of the automatic detection. Alerts sent to aMMOs were "unverified" automatic IR detections (i.e., included both true and false positive detections), and aMMOs could choose to ignore an alert (e.g., if already engaged in making a visual sighting). The 2016 field season was considered a good opportunity to "trial" and troubleshoot this new software application prior to using it on the vessel in 2017.

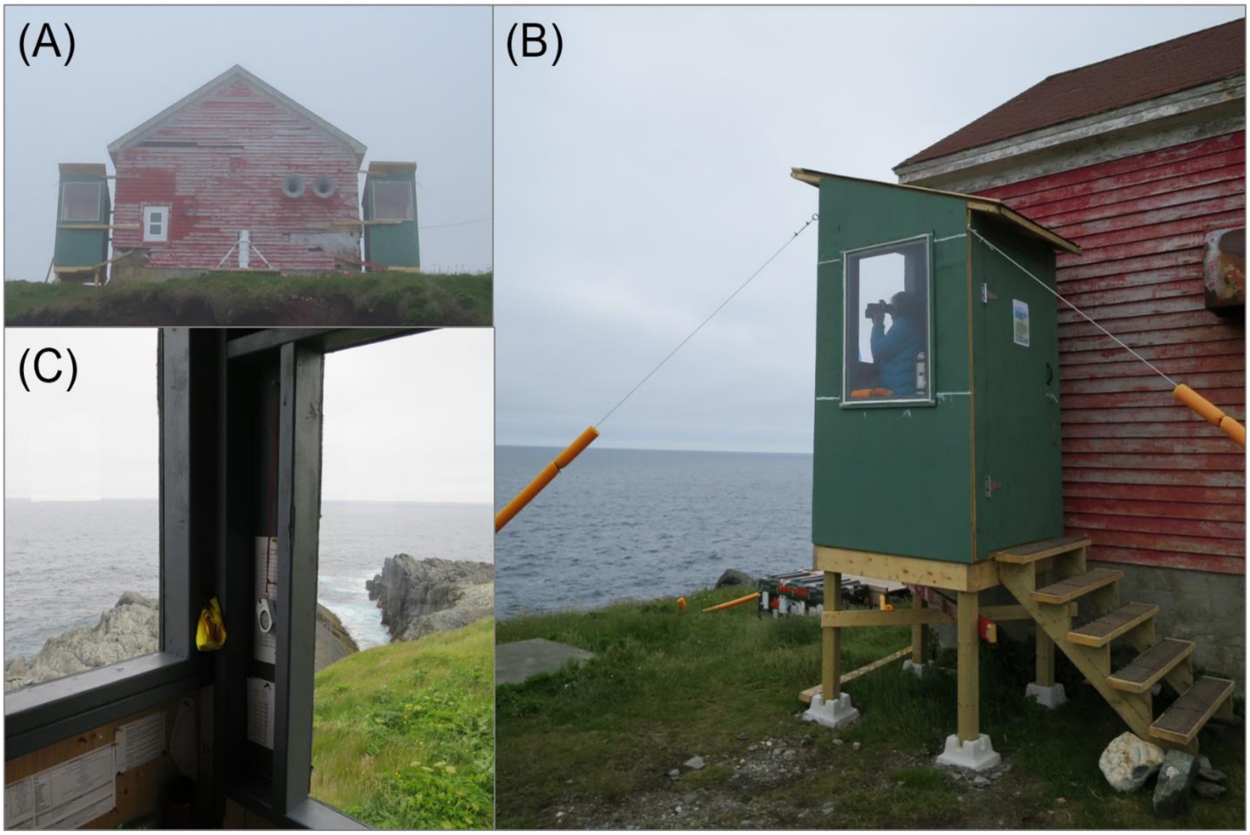

FIGURE 10. Data were collected from vantage points near the old foghorn building at Cape Race, NL. (A) Purpose-built observation booths for MMOs were located on either side of the old foghorn building. The FIRSTnavy sensor was situated on a level platform between the booths. (B) MMOs located marine mammals using reticle binoculars. (C) View from inside a booth.

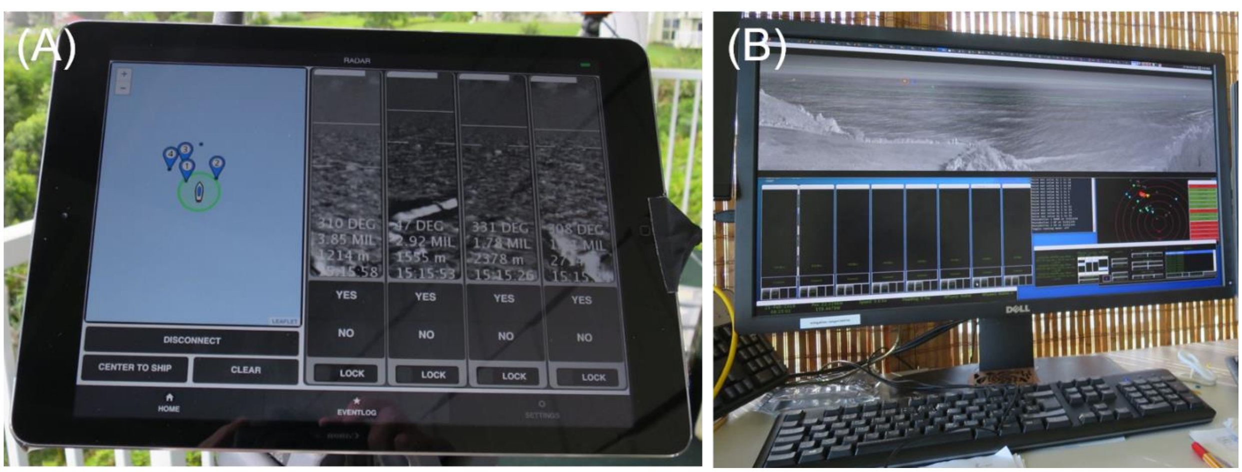

FIGURE 11. Example of the system used to relay automatic IR alerts to assisted MMOs at Cape Race, 2016. (A) An iPad relayed visual and audio alerts of potential marine mammal sightings from the (B) thermal (IR) system workstation located in the trailer onsite.

The following data were recorded for each sighting made by an MMO:

1. Local time.
2. Species: humpback whale, minke whale, fin whale, sei whale (*B. borealis*), North Atlantic right whale (*Eubalaena glacialis*), unidentified baleen whale, unidentified whale, white-beaked dolphin (*Lagenorhynchus albirostris*), unidentified dolphin, harbour porpoise, harbour seal, grey seal (*Halichoerus grypus*), unidentified seal.
3. ID reliability: certain, probable, possible.
4. Distance to sighting: number of reticles, clinometer degrees, or visual estimate.
5. Bearing to sighting: compass bearing determined using reticle binoculars.
6. Group size, including group size 1. The distance between individuals within a group varied according to the size of the species being observed such that larger species had more space between individuals within a group. The following group definitions were used:
   - Dolphins and killer whales – individuals within 10 m of each other.
   - Baleen whales – individuals within approximately two body lengths of each other. Whales observed lingering in the same areas (e.g., feeding) were categorized as being a single group even when individuals were more than 2 body lengths apart.
   - Seals – individuals within approximately 5 body lengths of each other.
7. Group ID: sequential numbers were assigned to identify repeat sightings of the same group (where group size ≥1); this was done only when the MMO was certain that the same group was re-sighted.
8. Cue that initially alerted MMO to the presence of the marine mammal: breach, body (e.g., back at surface), splash, flukes, pectoral fin, dorsal fin, head, "footprint", birds (diving or circling overhead), blow (categorized as being very weak and transparent, somewhat opaque, or very opaque). Subsequent cues were also recorded.
9. Was the MMO alerted to the presence of the marine mammal by the thermal (IR) system?

Observations were made during a wide variety of environmental conditions (e.g., daylight hours with excellent to impossible visibility, variable sea states with and without glare, and in a variety of fog "thicknesses"). During periods of reduced visibility (i.e., <500 m), a single MMO made observations from the south MMO booth. Due to the close proximity of the active foghorn to the north MMO booth, the north booth was not used when the foghorn was sounding.

3.2.3 Thermographic Data

The FIRSTnavy sensor was operated throughout the field period and recorded data during all hours when MMOs were on effort, as well as during some periods of darkness and heavy fog when no MMOs were on effort.

A night shift was run for four nights in order to collect additional thermographic data in darkness. A pair of field team members stayed in the trailer at Cape Race on these nights in order to ensure the security of the FIRSTnavy sensor.

3.3 Calibration Study

A calibration study was run in order to test the accuracy of all methods (reticle binoculars, theodolite, thermal (IR) system) used to localize marine mammals, and to calibrate the individual methods against a ground truth object on the water. This was done so that variability could be accounted for when matching sightings of marine mammals made using different localization methods.

On 8 July 2017, a local fishing vessel (FV *Three Sisters*) was chartered to follow a predetermined trackline in the area immediately offshore of our observation location at Cape Race (Figure 12). Location fixes were made concurrently by MMOs in each booth using reticle binoculars, by a pair of MMOs using a theodolite (Leica Builder 505; horizontal and vertical angles were recorded manually), and by using the thermal (IR) system. The vessel carried a handheld GPS device (Garmin GPSMap76Cx) in the wheelhouse, which was used to log the actual trackline followed. All location fixes were made on the vessel wheelhouse at the waterline. The trackline was chosen to mimic passing whales at different distances relative to the coastline, out to approximately 3 km.

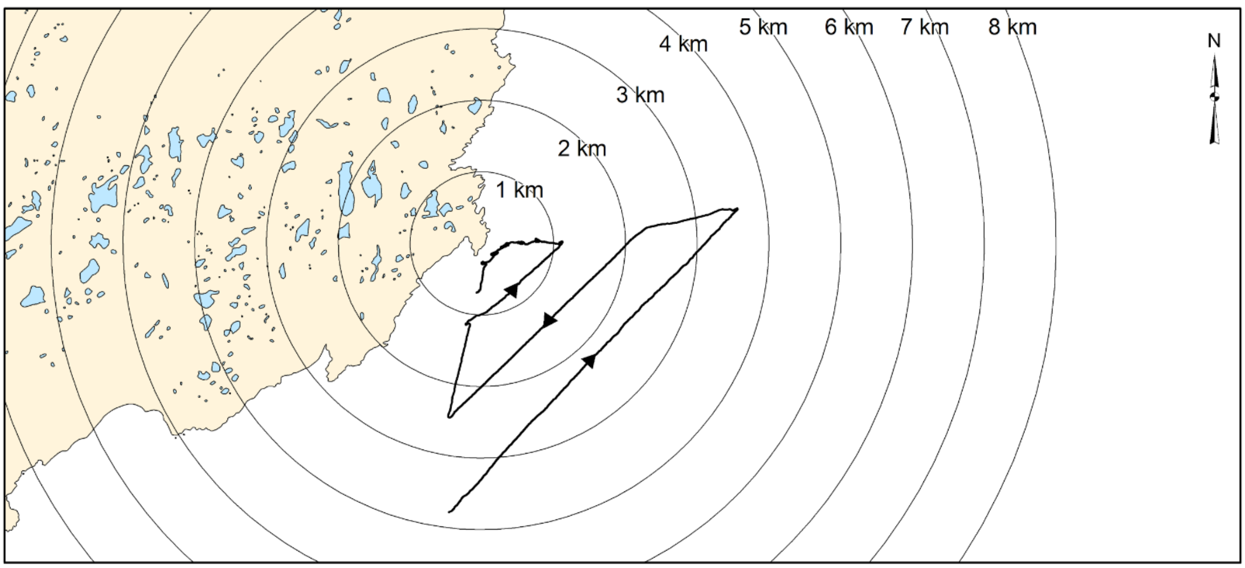

FIGURE 12. Approximate vessel trackline for the calibration study, Cape Race 2016.

3.4 Data Processing and Analysis

3.4.1 Visual Effort and Sightings

Visual observation effort, combined for both booths, was summed for the 2016 field season. All marine mammal sightings were categorized by species and sighting cue. Distance dependent sightability was plotted for humpback and minke whales, the only two species with a sufficient number of sightings for such an investigation. Daily sighting totals were compared for days with and without fog, and the relationship between visibility and sightability was examined.

3.4.2 Thermographic Data Processing

Our shore-based observation location presented us with a great data processing challenge in 2016. Due to the land-based location, there were two thermal signatures, northern gannets (*Morus bassanus*) and flying insects (observed within ~2 m of the camera), which posed significant difficulties for the thermal imaging detection algorithm. The northern gannet feeding activity in the area, including in the thermal (IR) system FOV, was especially high, resulting in an extremely large number of false positive thermal detections. This did not impede the detection of whale blows. However, the large number of false positive detections increased the frequency of alerts to the MMO to a level where the iPad used to deliver alerts to the MMO was sometimes not used because the alerts became more of a distraction than a helpful tool.

As all thermal (IR) data were recorded, we were able to reprocess the data and perform a retrospective analysis. To reduce the number of false positives to an acceptable level, we designed a second stage classifier based on a deep learning neural network[4] trained with true and false positive data from the first three days of the 2016 field season. This second stage classifier was used to filter out thermal anomalies in the remainder of the dataset. As a result, the number of false positives was reduced by a factor of 10, however, this also reduced the number of true positives by approximately a factor of 3. As these two thermal signatures are either not present (flying insects) or much less frequently present (gannets) on the open ocean, we did not try to reduce their influence further. We determined that spending additional personnel time trying to further address this challenge would not have been a prudent use of project funds.

After the majority of bird and insect detections were removed from the dataset, 6-second video snippets featuring thermal signatures resembling whale blows, but including both true and false signatures, were extracted by the optimized automatic detection algorithm.

3.4.3 Manual Review of Thermographic Data

After the thermographic data had been reprocessed, automatic IR detections were reviewed and classified as either being a thermal signature of a marine mammal or a false positive. Thermal anomalies classified as being created by a marine mammal were further classified as being a blow or body (which included aerial displays such as breaches and pectoral slaps). False positives were further classified as being birds, water (usually a breaking wave), insects, vessels, people (tourist visitors to Cape Race), items onshore (often grass or cables that secured MMO booths), sea, sky or unknown. Reviewers assigned a level of confidence to each classified image (certain, probable, or possible) and categorized the snippets as being sharp or blurry. Reviewers were naïve (i.e., did not know prior to video analysis what the video contained) during their review of the thermal (IR) data.

[4] Deep learning is a type of machine learning, a subfield of computer science that evolved from pattern recognition. It gives computers the ability to learn without being explicitly programmed (Wikipedia 2019).

3.4.4 Thermographic Detections of Marine Mammals

Consecutive true positive detections of marine mammals (i.e., detections of the same encounter that occurred within seconds of each other) were excluded from the dataset. Detection functions were generated to investigate thermal (IR) system detection performance for day and night-time hours and were compared with detection functions generated for sightings made by MMOs.

3.4.5 Comparison of Distances Estimated During the Calibration Study

Distances estimated by the three methods during the calibration study were compared, and limits used to guide the matching of visual sighting with thermal (IR) detections were determined.

3.4.6 Comparison of MMO Sightings and Automatic IR Detections of Marine Mammals

All visual sightings made by MMOs without the assistance of the IR system were pooled regardless of observation booth (north or south) used. To account for differences in FOV among the two MMO booths and thermal (IR) system, only detections made in the area of overlapping FOV (between bearings 50 and 235; Figure 13) were included. Thermal (IR) data collected at night were excluded. The following analyses were conducted on the resulting dataset consisting of only data collected simultaneously by experienced MMOs and the thermal (IR) system.

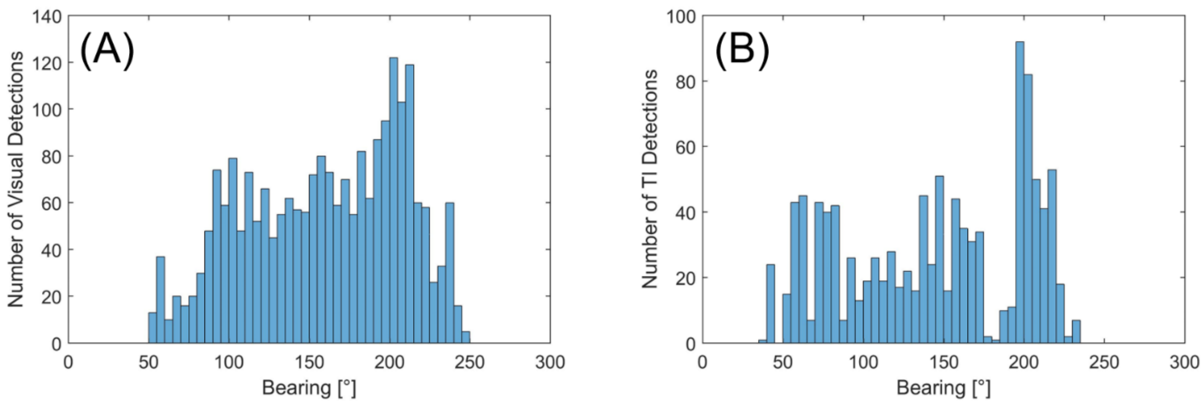

FIGURE 13. Bearings of marine mammal detections made by A) MMOs (both booths combined) and B) thermal (IR) system, Cape Race 2016.

- Average detection distances for sightings made by MMOs and the thermal (IR) system were compared across visibility and sightability categories.
- The performance of MMOs and the thermal (IR) system were compared using conditional probabilities (e.g., P(MMO|IR) = the probability that a marine mammal was detected by an MMO given that the same marine mammal was detected by the IR system). Because marine mammal detections made by MMOs are not typically simple point observations (i.e., an MMO may spot a marine mammal, but then take additional seconds or minutes to "confirm" the detection while waiting to observe a subsequent sighting cue produced by that marine mammal), a "buffer" was used when matching MMO sightings with IR detections. Detections were matched based on distance to detection using a buffer = 20 % of the detection distance.
- We evaluated conditional probability as a function of distance for baleen whales with and without minkes included, and for all marine mammal species observed. We also evaluated conditional probability on a daily basis.

Performance of the assisted MMOs was not investigated with this dataset because we felt that we could not accurately characterize aMMO performance given the large number of false positive detections (§ 3.4.2) and frequent troubleshooting required with the iPad application used to send IR alerts to aMMOs.

3.5 Results

Though data from the 2016 Cape Race field program were combined with data collected during non-ESRF projects to investigate the performance of the automatic whale detection system in a wide range of environmental conditions (Zitterbart et al. submitted), here we present results and analysis based only on data collected during the 2016 Cape Race field program.

3.5.1 Visual Observation Effort

There were approximately 454 hours of visual observation effort over the course of 31 days during the 2016 field season at Cape Race; this is the combined effort for both booths. Roughly one third of the visual observations were made with assistance from the thermal (IR) system (Table 3). The thermal (IR) system was used more often during the second half vs. the first half of the field season because the field team had to troubleshoot how to adapt the alerting system given the very large number of false alerts from diving seabirds and flying insects (§ 3.5.6).

3.5.2 Visual Sightings

A total of 1,923 marine mammal sightings were recorded by MMOs during the 2016 field season at Cape Race (Table 4 & Figure 14). Humpback and minke whales were the most commonly sighted species. Humpback whales were primarily detected by their blows, though MMOs also observed bodies (including dorsal fins) and aerial displays (breaching, fluking, slapping pectoral fins, and splashing; Figure 15). Minke whales were primarily detected as their backs or dorsal fins broke the surface of the water; minkes were rarely detected on account of their blows (Figure 16). Humpback whales were detectable at greater distances than minke whales (Figure 17). Other species of baleen whales (Figure 18), smaller cetaceans (Figure 19), and seals (Figure 20) were less commonly sighted.

TABLE 3. Marine mammal observer (MMO) effort (combined for two observation booths) at Cape Race, 2016. Some observations were made with assistance from the thermal (IR) system (i.e., aMMO).

Date	MMO Effort (h:mm:ss)	aMMO Effort (h:mm:ss)	Daily Total MMO Effort (h:mm:ss)
01-Jul-16	7:37:00		7:37:00
02-Jul-16	10:37:00	6:44:00	17:21:00
03-Jul-16	7:56:16		7:56:16
04-Jul-16	15:26:21	6:19:47	21:46:08
05-Jul-16	17:57:21		17:57:21
06-Jul-16	12:29:45		12:29:45
07-Jul-16	18:21:13	1:03:00	19:24:13
08-Jul-16	3:10:30		3:10:30
09-Jul-16	20:38:17		20:38:17
10-Jul-16	16:22:21	4:54:00	21:16:21
11-Jul-16	11:04:25		11:04:25
12-Jul-16	11:38:57		11:38:57
13-Jul-16	21:38:56		21:38:56
14-Jul-16	21:35:10		21:35:10
15-Jul-16	15:38:13	3:35:58	19:14:11
16-Jul-16	6:46:24	6:16:44	13:03:08
17-Jul-16	8:38:50		8:38:50
18-Jul-16	10:02:10	4:50:00	14:52:10
19-Jul-16		9:37:30	9:37:30
20-Jul-16	10:33:35	9:56:13	20:29:48
21-Jul-16	10:46:08	10:41:48	21:27:56
22-Jul-16	11:44:00	9:35:21	21:19:21
23-Jul-16	10:39:46	10:07:09	20:46:55
24-Jul-16	3:05:47	8:47:18	11:53:05
25-Jul-16	11:56:39	5:53:24	17:50:03
26-Jul-16	0:36:44	4:01:35	4:38:19
27-Jul-16	0:03:00	9:51:03	9:54:03
28-Jul-16		9:31:37	9:31:37
29-Jul-16	7:39:05	9:45:43	17:24:48
30-Jul-16	2:41:00	10:21:46	13:02:46
31-Jul-16	2:28:00	2:28:30	4:56:30
Effort Totals	**309:52:53**	**144:22:26**	**454:15:19**

TABLE 4. Marine mammals sighted, categorized by sighting cue, at Cape Race, 2016.

Species	Sighting Cue										Species Totals
	Blow	Body	Dorsal Fin	Head	Breach	Flukes	Pectoral Fin	Splash	Foot-print	Birds	
Baleen Whales											
N. Atlantic Right Whale		1									**1**
Fin Whale	2										**2**
Sei Whale	1										**1**
Minke Whale	13	595	106								**714**
Humpback Whale	711	177	15		26	21	13	28			**991**
Unidentified Baleen Whale	46	18						5	1		**70**
Other Cetaceans											
Unidentified Whale	10	14						10	1		**35**
White-Beaked Dolphin								1			**1**
Unidentified Dolphin		9	4		1			1		2	**17**
Harbour Porpoise		3	5								**8**
Seals											
Harbour Seal		13		65							**78**
Grey Seal				1							**1**
Unidentified Seal		1		3							**4**
Cue Totals	**783**	**831**	**130**	**69**	**27**	**21**	**13**	**45**	**2**	**2**	**1923**

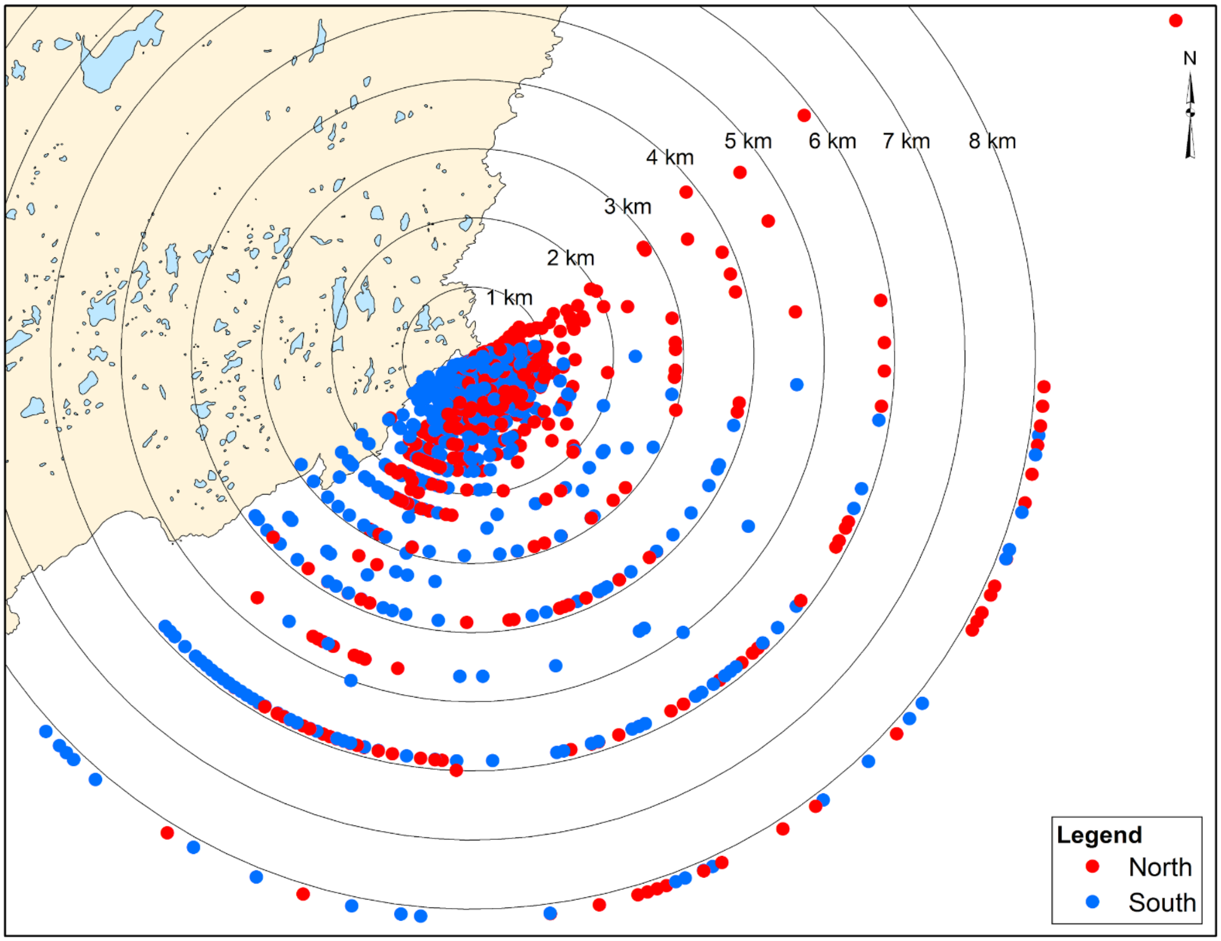

FIGURE 14. MMO sightings of marine mammals at Cape Race, 2016. Observations were made from "North" and "South" observation booths.

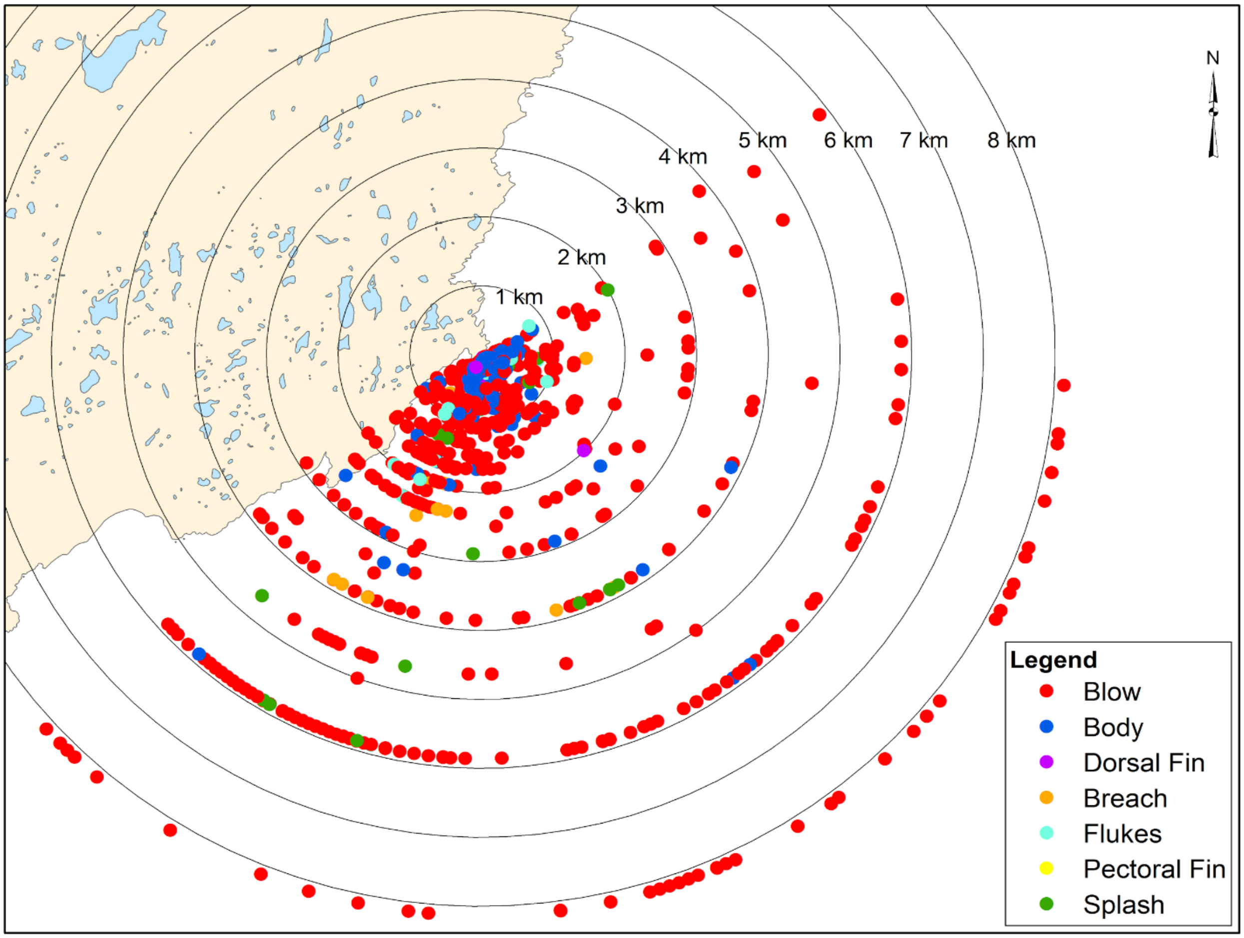

FIGURE 15. MMO sightings of humpback whales, categorized by sighting cue, at Cape Race, 2016.

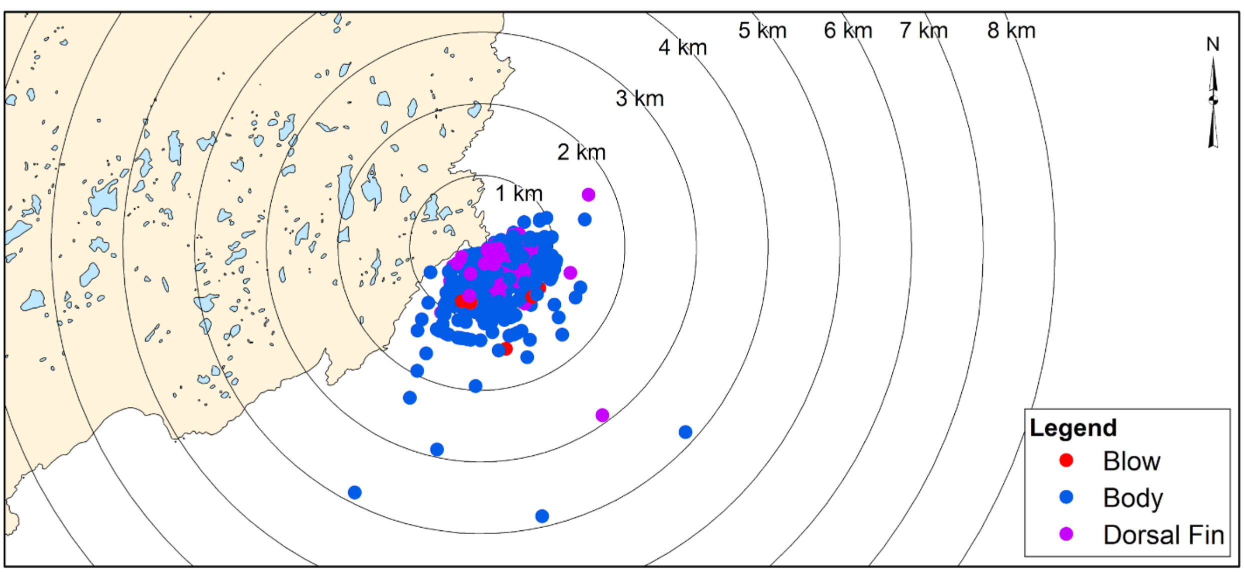

FIGURE 16. MMO sightings of minke whales, categorized by sighting cue, at Cape Race, 2016.

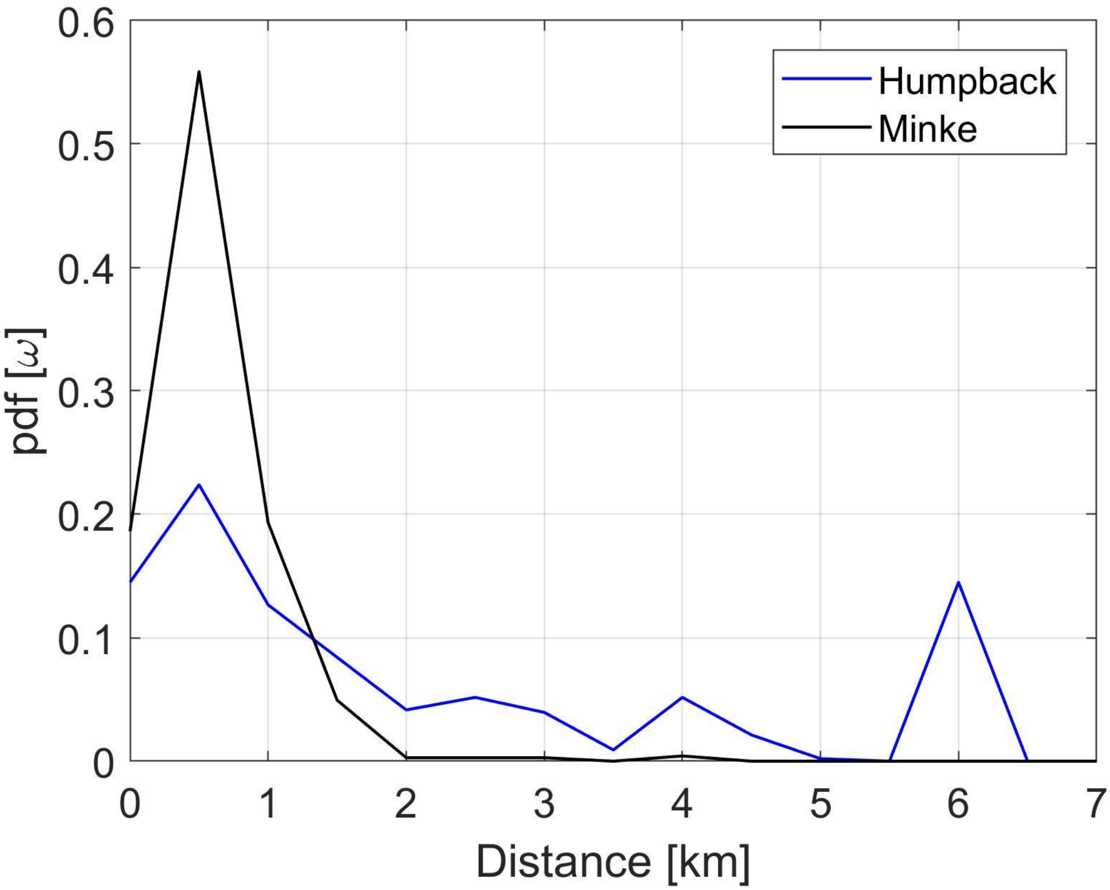

FIGURE 17. Distance dependent detectability (by MMOs) of humpback and minke whales, Cape Race 2016.

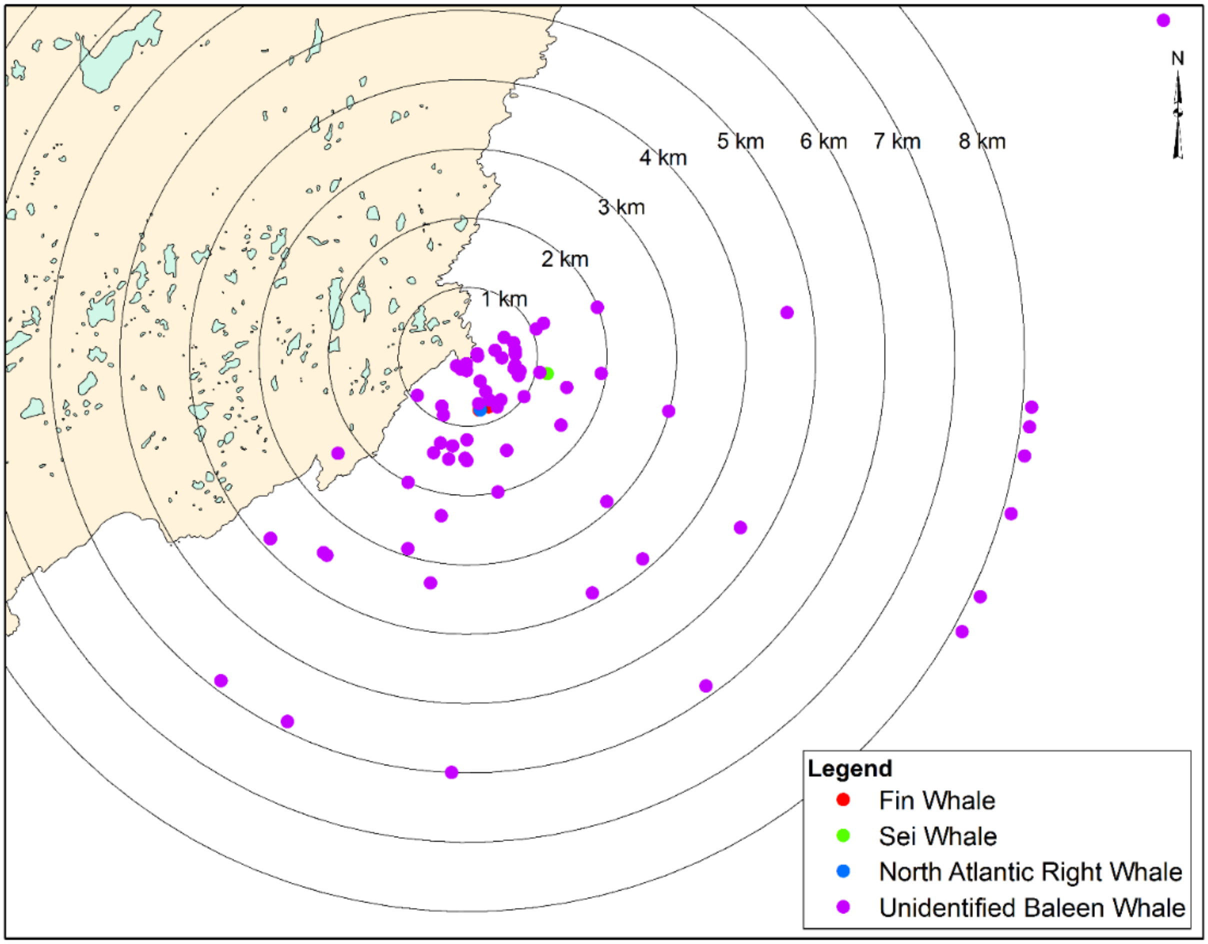

FIGURE 18. MMO sightings of baleen whales (excluding humpback and minke whales) at Cape Race, 2016.

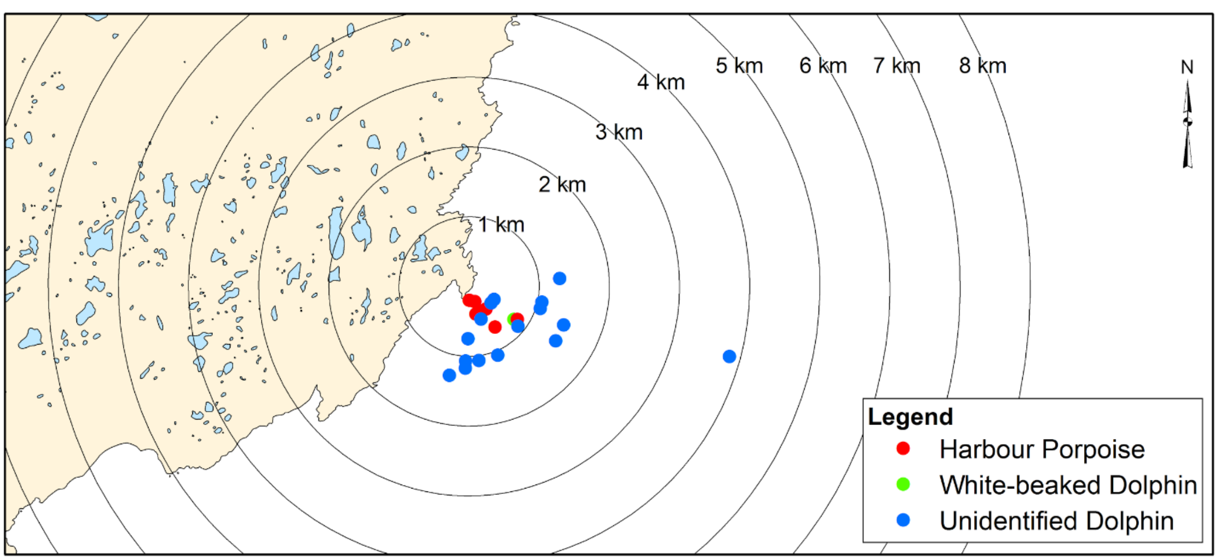

FIGURE 19. MMO sightings of odontocetes at Cape Race, 2016.

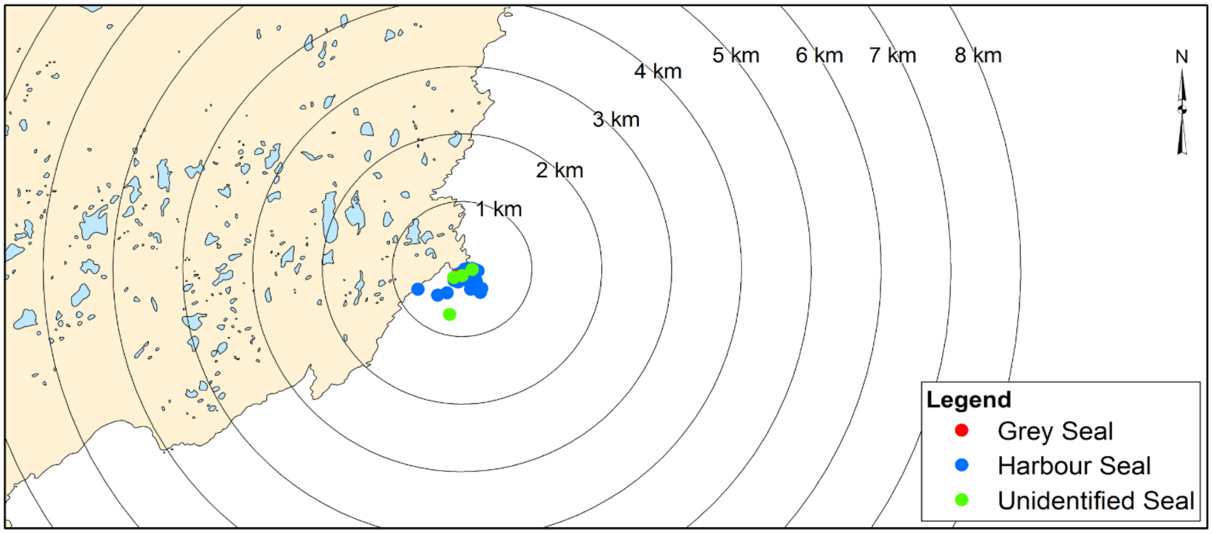

FIGURE 20. MMO sightings of seals at Cape Race, 2016.

3.5.3 Effect of Fog on Visual Sightings

Fog was recorded on 23 of the 31 observations days, and the lowest daily counts of marine mammal sightings (including zero sightings) occurred on days when fog was present (Table 5). The highest daily sighting count occurred on 10 July 2016, a day when conditions were predominantly recorded as being foggy. However, on this day visibility was recorded as ranging from 0.75–3 km, which was generally better than on many of the other days with fog (Table 5).

3.5.4 Correlation of Visibility with Sightability

Sightability was correlated with visibility as estimated by MMOs, but not as measured by the Vaisala FS11 visibility sensor (Figure 21). In general, visibility as estimated by MMOs can be used to distinguish between two regimes: (i) poor sightability, corresponding to visibilities <3.5 km, and (ii) good sightability, corresponding to visibilities >3.5 km (Figure 21).

3.5.5 Thermographic Data Collected

Approximately 400 hours of thermographic data were collected during the 2016 field season; 52 hours were collected during the "night shift".

TABLE 5. Summary of MMO marine mammal sightings and environmental conditions at Cape Race, 2016.

Date	MMO Effort (h:mm:ss)	Cues Sighted	Environmental Conditions	
			Precipitation*	Visibility (km)
01-Jul-16	7:37:00	0	F, M	0 – 2
02-Jul-16	17:21:00	11	F, M	0 – 10
03-Jul-16	7:56:16	0	F, M	0 – 0.2
04-Jul-16	21:46:08	17	F, N	1 – 10
05-Jul-16	17:57:21	22	F, R, M, H, N	0 – 10
06-Jul-16	12:29:45	10	F, H, N	0.05 – 10
07-Jul-16	19:24:13	162	R, H, N	8 – 10
08-Jul-16	3:10:30	61	N	10
09-Jul-16	20:38:17	172	H, N	9 – 10
10-Jul-16	21:16:21	235	F, M	0.75 – 3
11-Jul-16	11:04:25	39	F, R, D, M, H	0 – 5
12-Jul-16	11:38:57	135	N	5 – 10
13-Jul-16	21:38:56	77	H, N	5 – 10
14-Jul-16	21:35:10	136	R, D, H, N	6 – 10
15-Jul-16	19:14:11	102	F, H	0.15 – 8
16-Jul-16	13:03:08	20	F, M	0 – 2
17-Jul-16	8:38:50	8	F	0 – 3
18-Jul-16	14:52:10	8	F, M	0 – 10
19-Jul-16	9:37:30	0	F, M	0 – 0.4
20-Jul-16	20:29:48	136	F, R, M, H, N	0.5 – 10
21-Jul-16	21:27:56	36	H, N	2 – 9
22-Jul-16	21:19:21	122	H, N	5 – 10
23-Jul-16	20:46:55	58	F, H	3 – 9
24-Jul-16	11:53:05	17	F, M, H	0 – 6
25-Jul-16	17:50:03	80	F, M, H	0 – 10
26-Jul-16	4:38:19	23	F	0 – 0.3
27-Jul-16	9:54:03	6	F	0 – 0.4
28-Jul-16	9:31:37	3	F	0 – 0.2
29-Jul-16	17:24:48	69	F, H	0 – 9
30-Jul-16	13:02:46	96	F, M, H, N	0 – 8
31-Jul-16	4:56:30	62	F	0.05 – 3
Grand Total	**454:15:19**	**1923**		
***** F = fog, R = rain, D = drizzle, M = mixed precipitation, H = haze, N = none.				

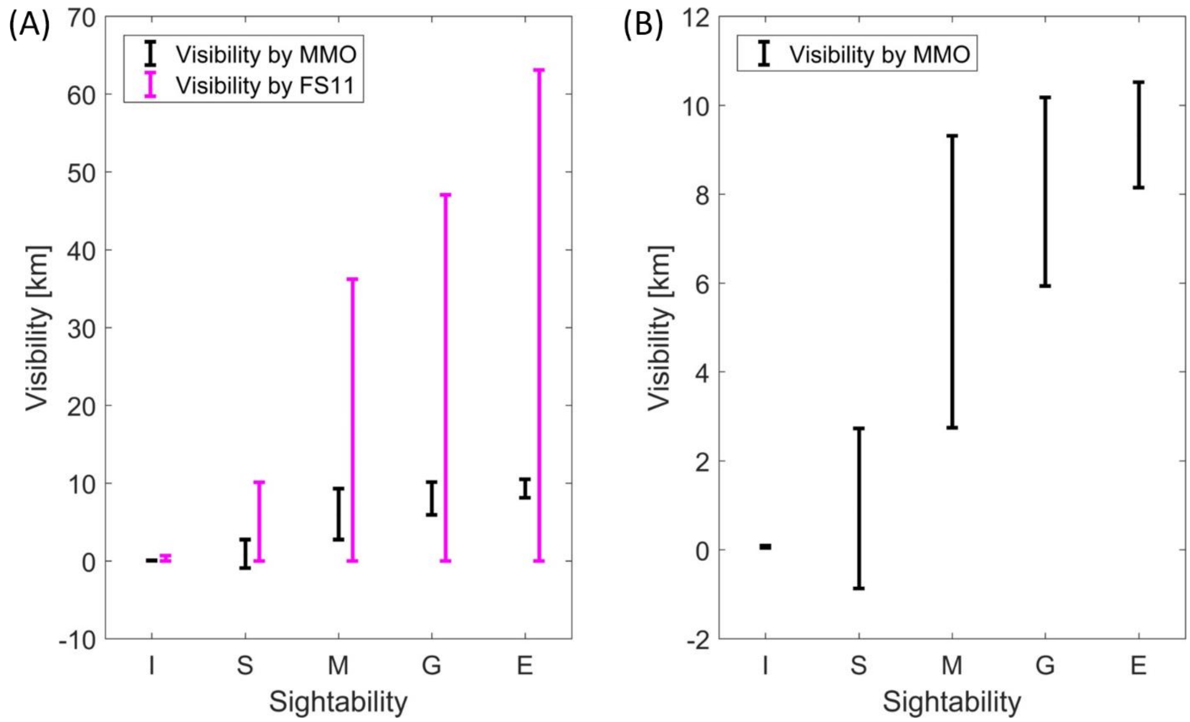

FIGURE 21. Comparison of visibility and sightability. (A) Visibility as estimated by MMOs and measured by the Vaisala FS11 visibility sensor. (B) Visibility as estimated by MMOs (note different scales on y-axis).

3.5.6 True and False Positive Automatic IR Detections

A total of 25,042 thermal (IR) data snippets were manually reviewed. The majority of these snippets were false positives (Figures 22 & 23); most of the bird false alerts were caused by diving northern gannets. Marine mammals accounted for 7.3 % of the automatic detections (Figures 24, 25 & 26, Table 6). Seals were occasionally detected by the thermal (IR) system.

In addition to birds not filtered out by the second stage classifier, water and waves (consistently breaking in a few locations nearshore) were the most commonly identified false positives (Table 6).

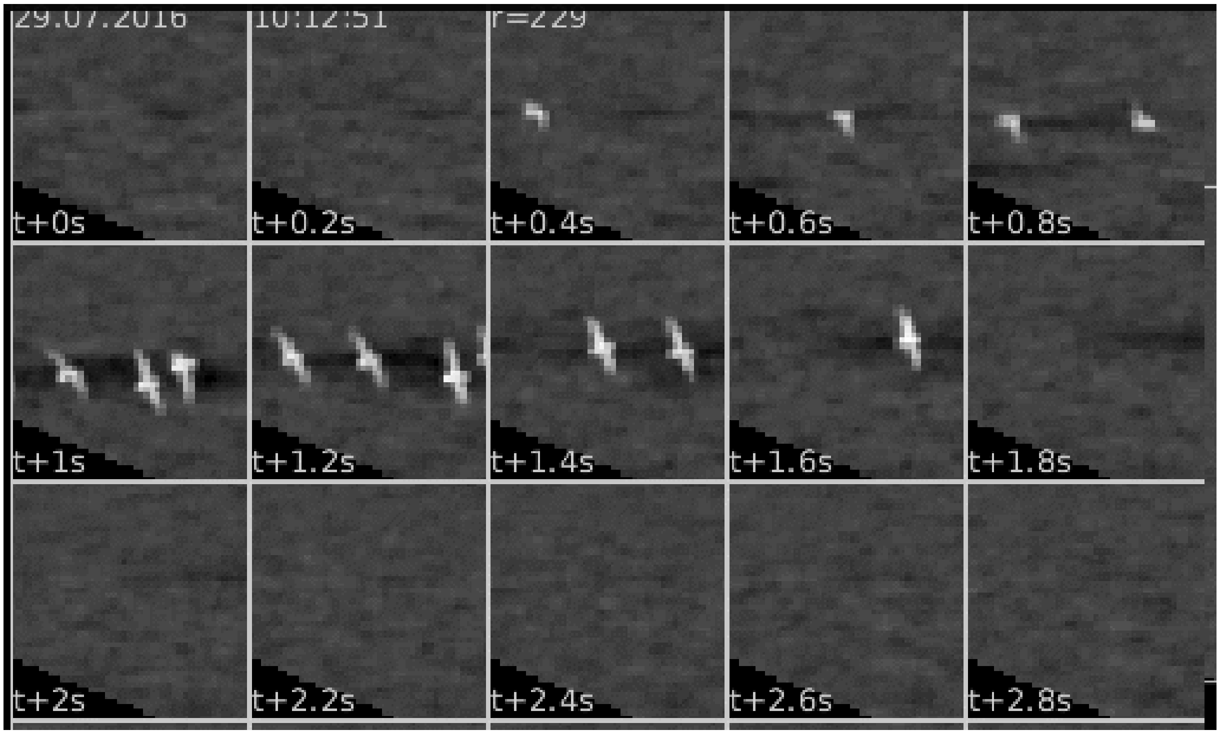

FIGURE 22. Northern gannets automatically detected by the thermal (IR) system at Cape Race, 2016.

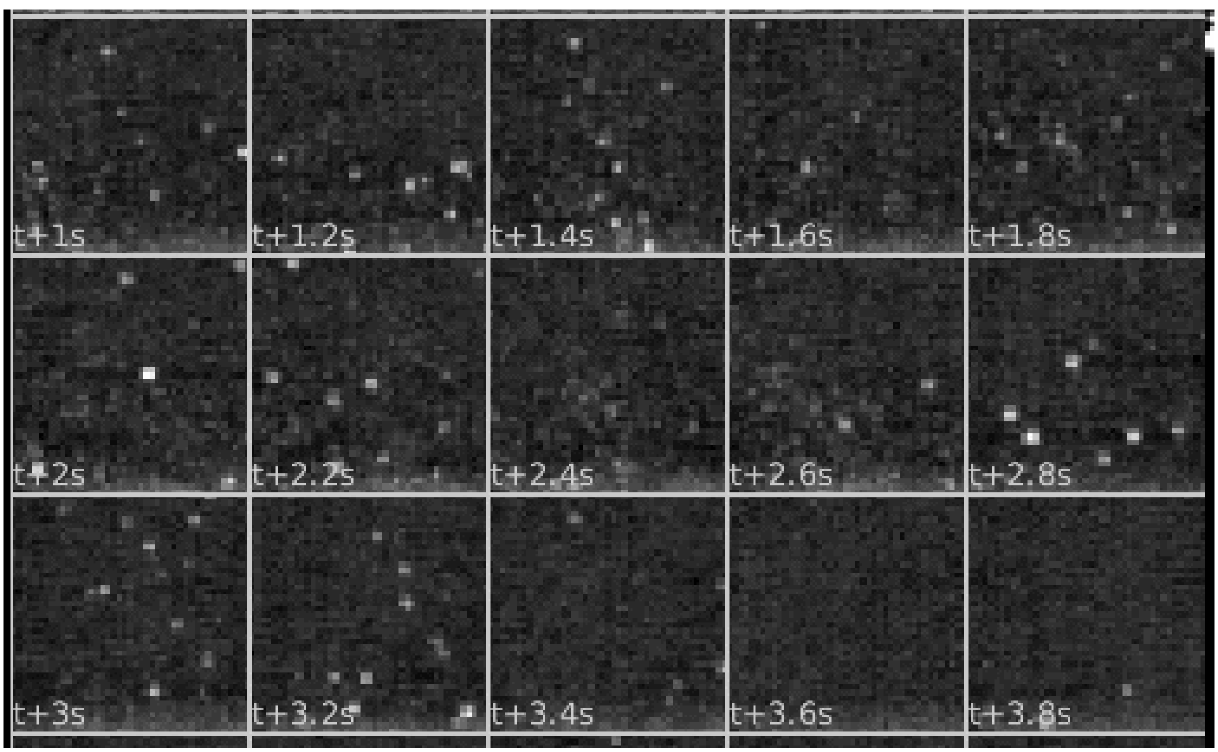

FIGURE 23. Flying insects automatically detected by the thermal (IR) system at Cape Race, 2016.

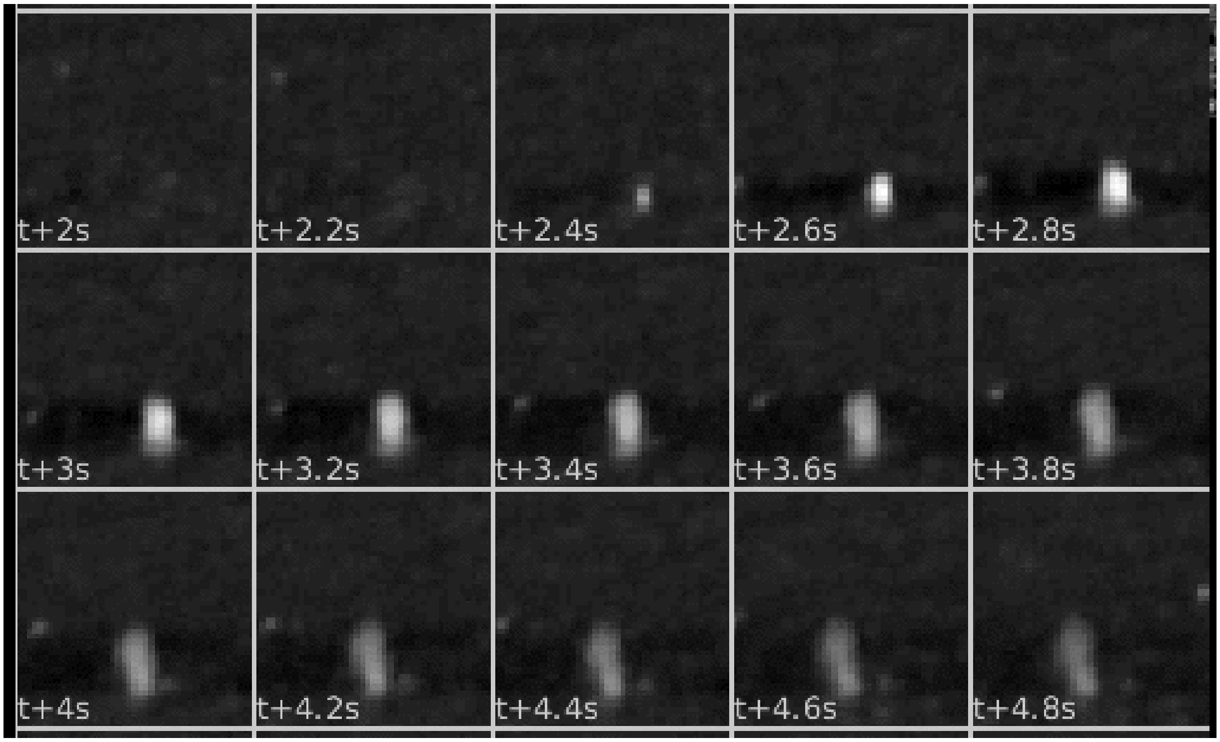

FIGURE 24. Whale blow automatically detected by the thermal (IR) system at Cape Race, 2016. Note the small white object in the images is a bird.

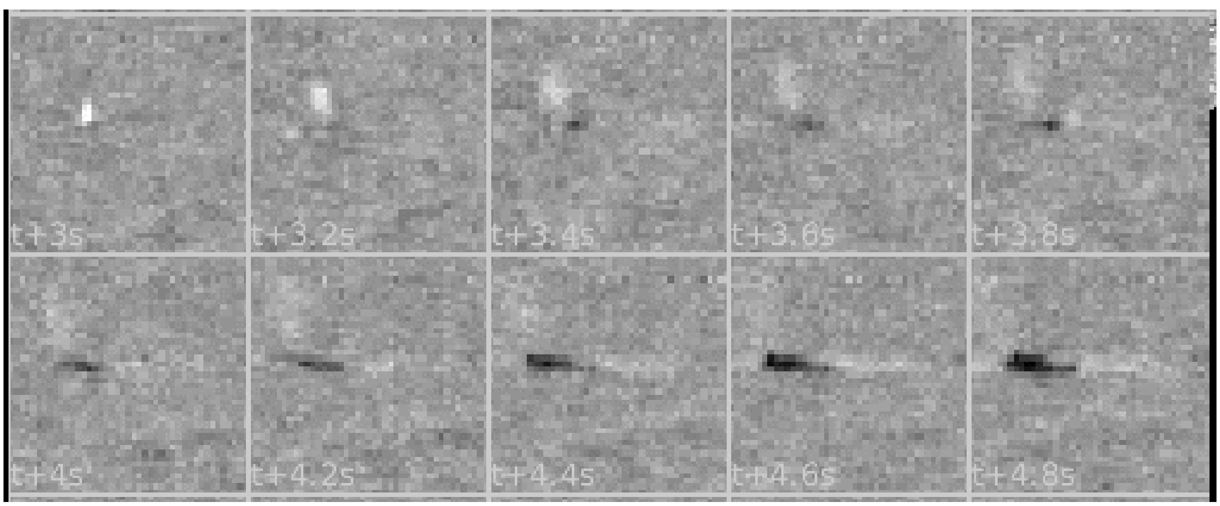

FIGURE 25. Whale body (preceded by a blow) automatically detected by thermal (IR) system at Cape Race, 2016.

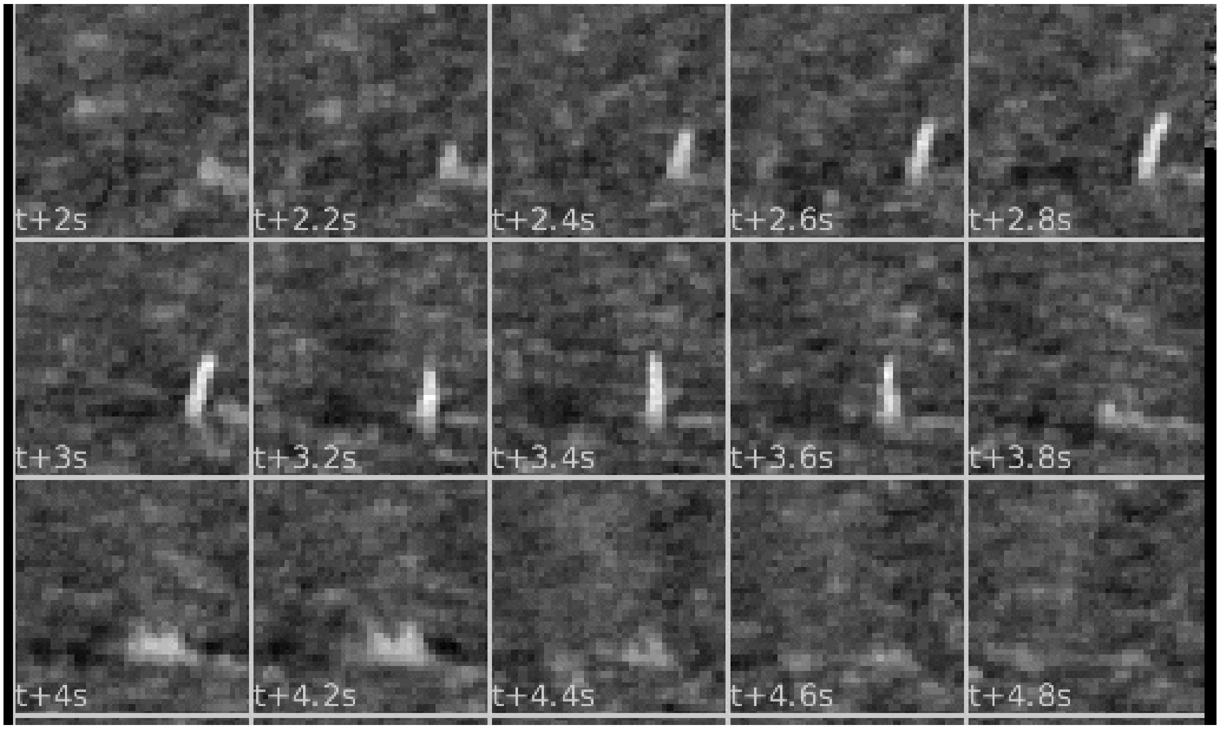

FIGURE 26. Humpback whale slapping a pectoral fin automatically detected by thermal (IR) system at Cape Race, 2016.

TABLE 6. Categories assigned to automatic IR detections during manual review of images.

Category	**Number of Detections**
Marine Mammals – True Positives	
Blow	1580
Body	242
False Positives	
Birds	7760
Insects	213
Waves and Water	8959
Vessel	923
Sky, Sea, Unknown	3587
Objects on Land (e.g., grass, people)	1778
Grand Total	**25,042**

3.5.7 Automatic IR Detections of Marine Mammals during Day and at Night

After excluding consecutive sightings, a total of 600 automatic IR detections were made during the day and 451 detections were made at night. The number of automatic IR detections (relative to effort) made at night greatly exceeded those made during the day.

The detection function describing the performance of the detection algorithm follows a shape that is expected from a point-transect distance sampling survey design (Figure 27). With increasing distance, the number of detections increases due to the increase in area monitored at that distance. We find the peak of the detection function differs for day vs. night; peaks were found at 0.5 km for MMOs and the IR system during the day, and at 2 km for the IR system at night (Figure 27).

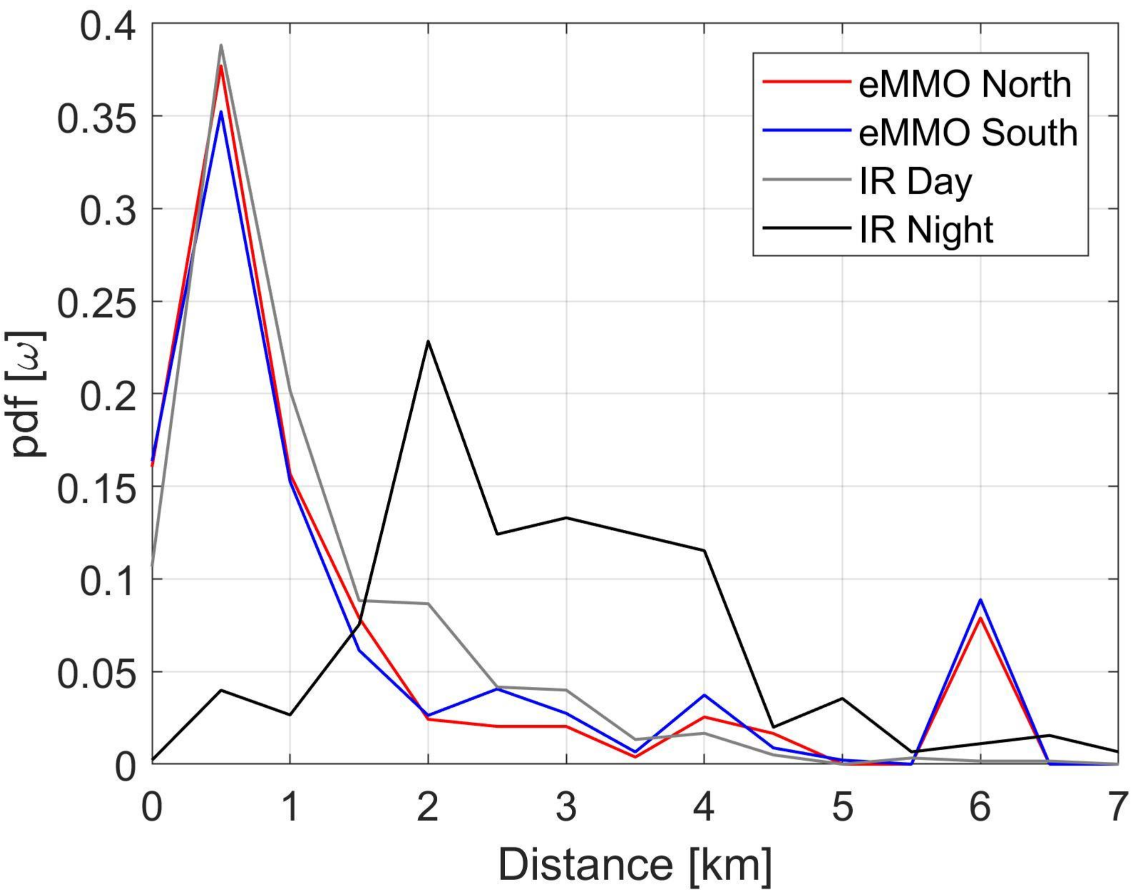

FIGURE 27. Detection functions for experienced MMOs and the thermal (IR) system at Cape Race 2016. MMOs made detections from two observation booths (north and south) during daylight hours. Detection functions for the thermal (IR) system were generated separately for daylight and darkness (night).

3.5.8 Results of the Calibration Study

The calibration study took approximately 2.5 hours to complete. A total of 628 localizations (north booth: 329, south booth: 299) were made by MMOs using reticle binoculars or estimating nearshore distances using clinometers or the naked eye (Figure 28). Three hundred and sixty localizations were made using the theodolite.

The calibration study shows very good accordance of all methods out to 1500 m. Further out, the different methods started to deviate; the theodolite underestimated the distance to the vessel, while the reticle binoculars often led to an overestimation of the distance.

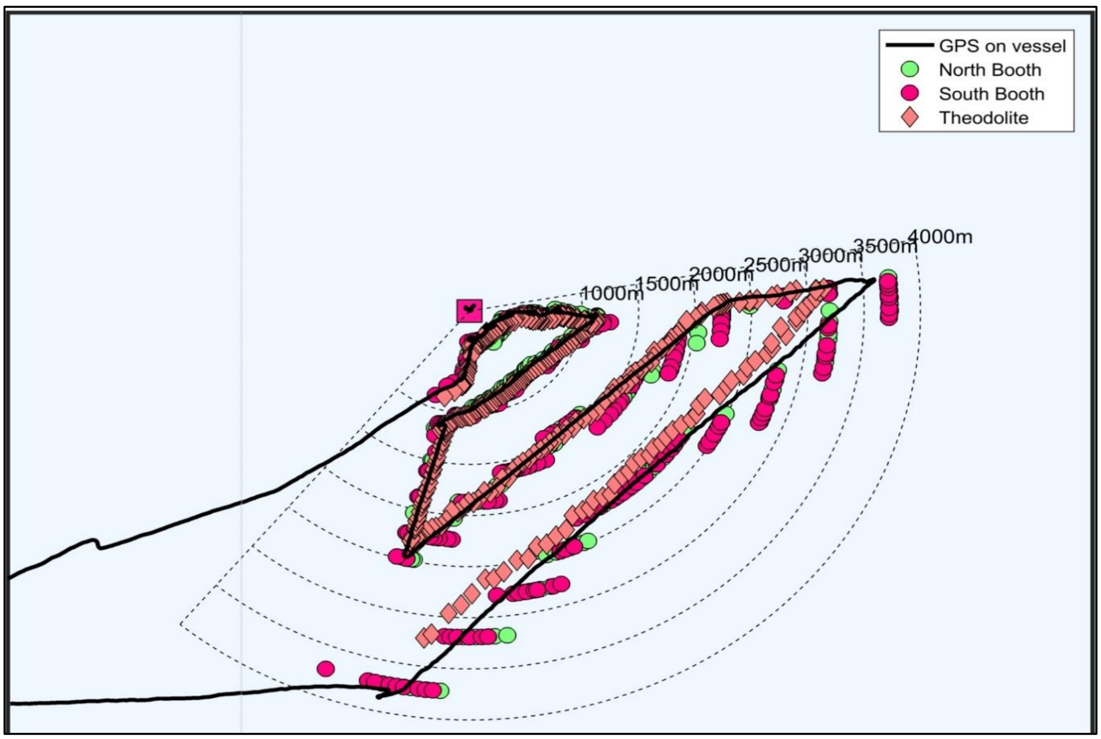

FIGURE 28. Sightings made using reticles binoculars (by MMOs in North and South observation booths) or a theodolite during the calibration study at Cape Race, 2016.

3.5.9 Average Detection Distances for MMOs and the thermal (IR) system

The average detection distance for MMOs and the automatic thermal (IR) system is <500 m in visibility conditions <5 km (as measured by the Vaisala FS11 visibility sensor; Figure 29A). In visibility conditions >7 km, the mean thermal (IR) detection distance increases to 2 km, and MMO detection distances increases to 1 km. In conditions with visibilities >10 km, the average MMO detection distance increases to 2 km (Figure 29A). When sightability was classified as impossible or severely impaired, both thermal imaging and MMOs had average detection distances <500 m. In moderate, good and excellent sighting conditions, detection ranges were comparable with the mean around 2 km (Figure 29B).

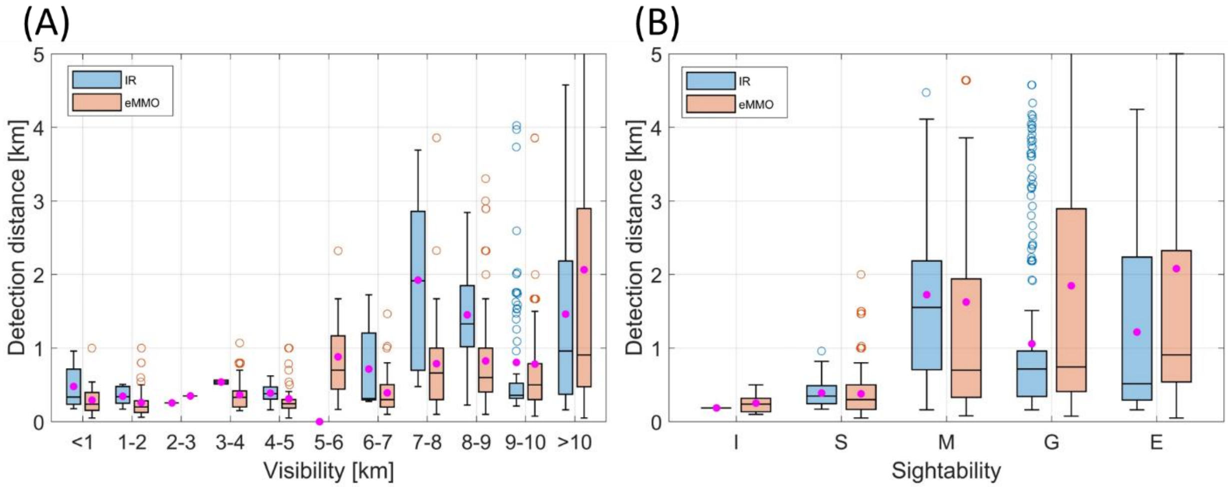

FIGURE 29. Influence of (A) visibility conditions as measured by the Vaisala FS11 sensor, and (B) human estimated sightability conditions (I = impossible, S = severely impaired, M = moderately impaired, G = good, E = excellent), on the average detection range of experienced MMOs and the thermal (IR) automatic detection system. Pink dots denote the mean value of the distribution in each bin.

3.5.10 Relative Detection Performance of MMOs and the Thermal (IR) System

The experienced MMOs marginally outperform the thermal (IR) system in detecting large baleen whales over the range of distances investigated (Figure 30A); the performance of the two methods is nearly identical for cues sighted at 0.5 km, though the difference gradually increases with increasing distance up to which cues sighted are considered such that the eMMO performs approximately twice as well as the thermal (IR) system at 6 km. The difference in performance between the two methods increases when cues produced by other species are included in the analysis such that the eMMO performs approximately three times as well as the thermal (IR) system (Figures 30B & 30C). The detection performance of the thermal (IR) system also increased, compared to the comparison for baleen whales without minkes, though the increase was slight. On a day-to-day basis, detection performance was found to be quite variable (Figure 31), though overall, the eMMO was found to outperform the thermal (IR) system.

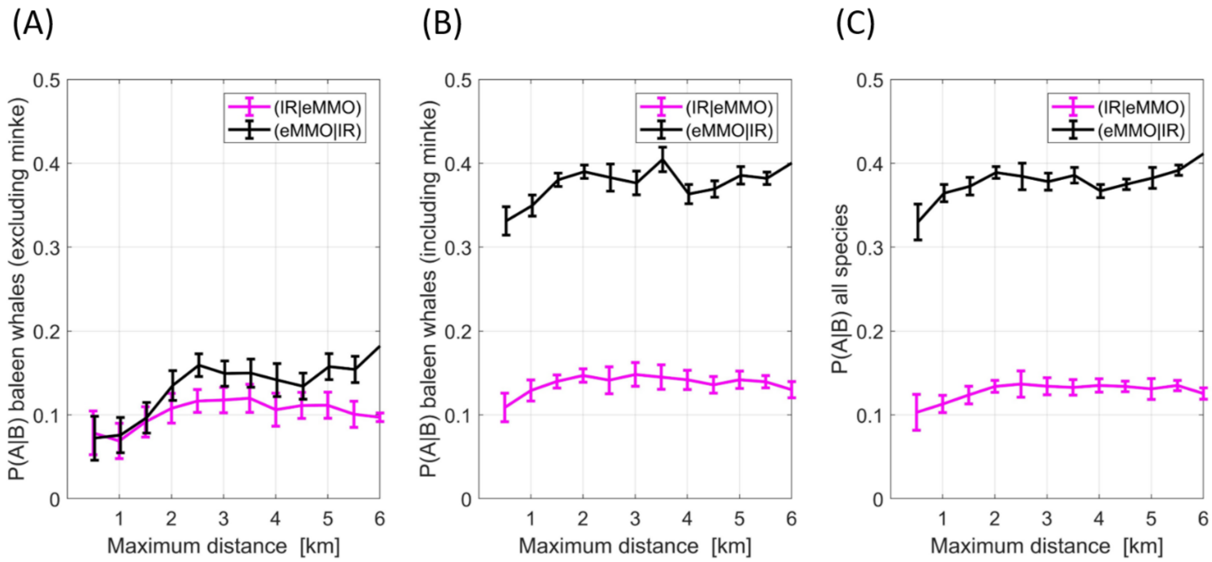

FIGURE 30. Conditional probabilities that detections were made by experienced MMOs and the thermal (IR) system at Cape Race, 2016 for (A) baleen whales (excluding minkes), (B) baleen whales (including minkes), and (C) all species of marine mammals.

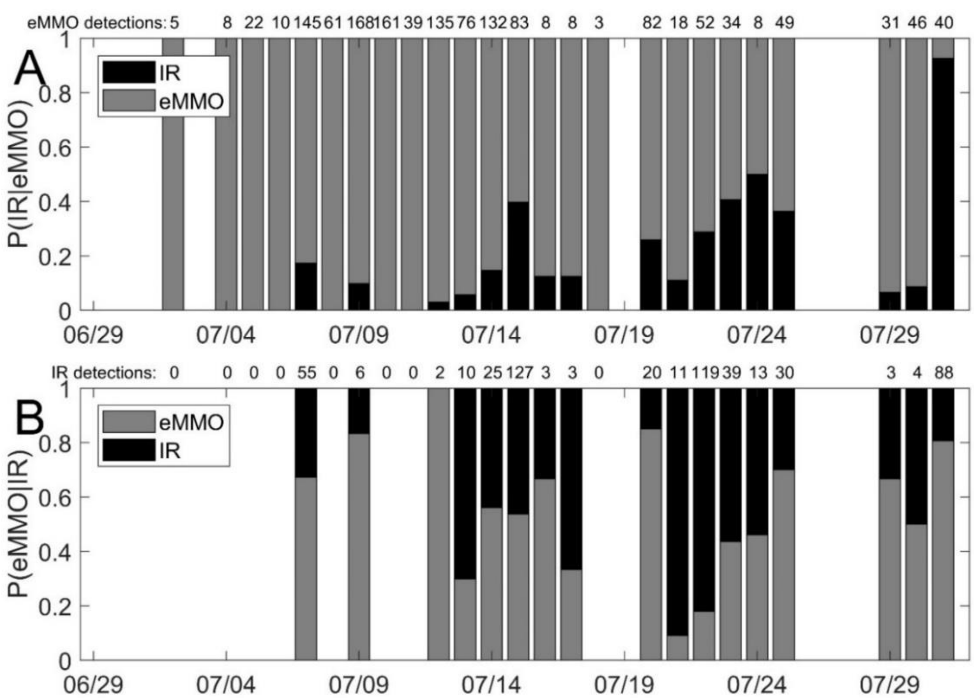

FIGURE 31. Daily conditional probabilities that (A) marine mammals were detected by the thermal (IR) system given that they were sighted by MMOs, and (B) marine mammals were sighted by MMOs given that they were detected by the thermal (IR) system, at Cape Race 2016.

3.6 Discussion

Data collected during the 2016 field season allowed us to investigate a number of factors that influence the performance of the thermal (IR) system despite the unexpectedly high number of false positive detections caused by plunge-diving northern gannets and flying insects at our shore-based field site.

3.6.1 False Positive Automatic IR Detections

Managing the number of false positive detections is a crucial aspect of any automatic detection system. In a mitigation setting, false negatives (i.e., missed whales) are problematic because they result in an increased risk of animals being exposed to potentially injurious levels of airgun sound. The automatic whale detection algorithm must be designed to balance the number of false positives with the number of false negatives.

Many of the false positive detections we encountered were due to our shore-based field site location (e.g., diving gannets and waves breaking along the shore to the south of our observation site). Thus, we did not attempt to modify the algorithm to remove these specific sources of false positive detections as they are less likely to occur when at sea, and because modifying the algorithm to remove them would be accompanied by an unwanted increase in false negatives. The software does incorporate an object-tracking algorithm designed to filter out individual birds (as well as vessels moving through the field of view) before creating a false alert. It was thought that because multiple diving gannets often flew through the FOV at the same time, that the tracking algorithm could not separate the tracks, and therefore, did not remove the birds from the detection process, leading to very high numbers of false alerts.

3.6.2 Detection Functions: Day vs. Night

The thermal (IR) system was shown to be capable of making automatic detections of whale during both daylight and darkness. Detection functions for the automatic thermal (IR) system (and for the MMOs) generally follow the shape expected for a point transect distance sampling design. With increasing distance from the observer, an increasingly large portion of the ocean surface is surveyed (i.e., between 1–2 km and 2–3 km, the observed area is doubled). Therefore, more animals and thus more detections are expected farther out. Number of detections is expected to increase linearly until reaching a peak. After this peak, the detection function follows a non-linear decay in the number of detected animals. The peak of the detection function marks the point where (assuming an equal distribution of animals throughout the observation area) the detection method starts to miss cues and can serve as an indicator up to which distance a method can be reliably utilized for mitigation purposes.

We found that the automatic IR detection function peaks at 0.5 km in daytime, and 2 km at night (see Figure 27). The position of a detection peak depends on several factors, including visibility, platform height and animal behaviour. We suggest that the difference between day and night peaks may be attributed to humpbacks tracking capelin (*Mallotus villosus*; Whitehead and Carscadden 1985 or other prey) in nearshore waters, and we make this suggestion only after eliminating other possible explanations for the difference. We determined that the different peaks could not be attributed to different visibility conditions during the day vs. night by plotting the Vaisala FS11 visibility measurements for day and night; visibility conditions were not found to be better during the night than the day (Figure 32). The position of the daytime IR detection peak also coincides with the detection peak for MMOs (see Figure 27). Because MMOs were shown to detect whales at distances up to 6 km (see Figure 27), this suggests that had whales been present at 2 km during the day, that they would have been detected by MMOs (and the IR system). Thus, the lack of a peak at 2 km during the daytime likely reflects a lack of whales at that distance. Additionally, the platform height at Cape Race remained constant throughout the study.

That thermal (IR) detections were made at distances of 3–4 km during both day and night indicates that this technology is capable of assisting with 24-hr monitoring of the area around a vessel at distances relevant for mitigation (DFO 2008; Verfuss et al. 2016).

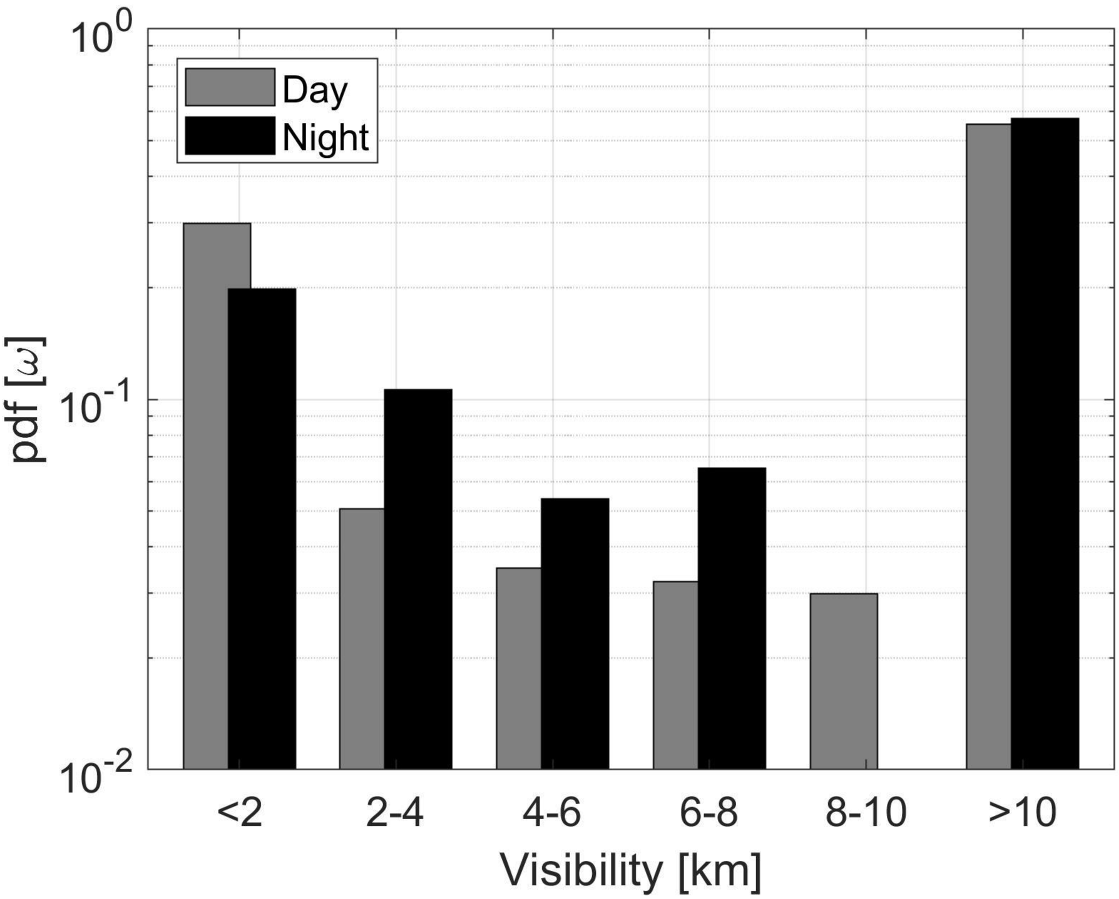

FIGURE 32. Visibility during day and night, as measured by the Vaisala FS11 sensor at Cape Race, 2016.

3.6.3 Thermal (IR) Detection Performance in the Fog

The potential for using thermal (IR) imaging to detect marine mammals in foggy conditions is limited (Verfuss et al. 2018); long wavelength infrared (LWIR) transmission loss increases with increasing relative humidity (Winchester and Gimmestad 1982), and not unexpectedly, LWIR detection range has been demonstrated to decrease with increasing fog (Beier and Gemperlein 2004). Unfortunately, the impact fog has on the ability of the IR system to detect whales is impossible to measure using perceptibility as a performance metric (§ 2.3.2) because human observers are also strongly impacted by fog. Instead, we conducted an analysis on how visibility (which is affected by fog) influences detection. We find that at low visibility (as would be caused by dense fog) the thermal (IR) imaging system was as impeded as a human observer (see Figure 29A; dense fog is, on basis of the average detection distance, comparable to visibility conditions <5 km as measured by the FS11). It is to be noted that the FS11 sensor measures visibility in units standardized for use in aviation. Therefore, FS11 measurements are generally greater than human-estimated visibility values. We use the FS11 measured visibility values here for comparability with other studies. Interestingly, we also find that when visibility is <10 km, the average detection distance is higher for the thermal (IR) system than for an experienced MMO. We suggest that this may be due to hazy or misty conditions, which IR radiation can penetrate better than visual spectrum light, leading to larger detection distances for the IR system (see Figure 29A). In operational terms this can be interpreted that at hazy or misty conditions the thermal (IR) detection systems would allow for greater detection ranges than an MMO, but in dense fog conditions, it is equally affected.

3.6.4 Relative Detection Performance of MMOs and the Thermal (IR) System

Detection performance for the thermal (IR) system is rather low (conditional probabilities ~0.1–0.15; see Figure 30), regardless of what species are included in the analysis. It is likely that these values underestimate the true performance capability of the system given that such a large number of false positive detections were encountered at our shore-based field site (§ 3.5.6). The large number of diving gannets in the area, potentially attracted to the same prey as the humpbacks, often resulted in the detection algorithm running into its computational limit (i.e., too few computational resources) and subsequently dropping frames, thus missing detections. As well, the additional classification step used to reduce the initial number of detections down to a set that could be subsequently reviewed by humans (§ 3.4.2) likely also resulted in some true positive detections being filtered out of the dataset. In 2016, we expected that the system would perform better at sea, as such large numbers of false positives are unlikely to occur on a ship-based scenario.

The improvement in MMO performance when additional species were included in the analysis (see Figures 30B & 30C) is thought to reflect the complex detection scenario present at Cape Race. Both relatively easy to spot (i.e., humpbacks) and more elusive marine mammals (e.g., minkes and harbour porpoise) are present at Cape Race. In order to remain vigilant for all species, MMOs must simultaneously employ different search strategies (i.e., observe both near and offshore waters, and employ multiple search images, e.g., blows and brief glimpses of dorsal fins). Including multiple marine mammal species in the analysis (see Figures 30B & 30C), results in the incorporation of all of the MMOs monitoring efforts, thereby increasing the MMO's overall performance. That the performance of thermal (IR) system did not increase to the same extent as the MMOs when additional species were included indicates that the thermal (IR) system is not as capable at detecting these more elusive species.

Detection performance of both MMOs and the thermal (IR) system was shown to be extremely variable over time (see Figure 31), and likely reflects the variability in daily visibility conditions (see Table 5). Low conditional probability values indicate that overall performance can be increased by employing multiple detection methods simultaneously. In fact, Verfuss et al. (2018) have suggested that modelling be used to estimate what overall detection performance may be achieved when multiple detection methods are employed together. The data we have collected at Cape Race will be useful in these efforts.

3.6.5 Delivery of Thermal (IR) Alerts to MMOs on a Seismic Vessel

Though we had hoped to make a comparison of marine mammal detections made by MMOs with and without assistance from the thermal (IR) system, we encountered a number of challenges in 2016 that prevented us from making this analysis in a robust manner: the iPad application developed to deliver thermal (IR) alerts to the MMO had a number of software bugs and was not always functional (§ 3.2.2), and the large number of false positive detections (birds and insects) resulted in such a large number of thermal (IR) alerts that the MMO using the iPad sometimes shut the application off because it was more distracting than helpful. This experience with the iPads also made us aware of other aspects of the alerting system that needed to be solved before embarking on our 2017 vessel-based field program. In particular, MMOs need to remain mobile (i.e., they cannot remain behind a computer at a fixed station) as they search for marine mammals using the naked eye or binoculars, and any signal used to alert and MMO to a new automatic IR detection must get the attention of the MMO without distracting or alarming the bridge crew (i.e., cannot broadcast an audible beep over the bridge area on a working seismic vessel).

4.0 COMPARISON OF THREE DETECTION METHODS AT SEA, 2017

The 2017 field program was designed to address the final two study objectives (§ 1.2): 2) to compare marine mammal detections made using three different methods (i.e., MMOs making visual observations, MMOs utilizing AIMMMS to make detections, and PAM); and 3) to evaluate the quality and accuracy of marine mammal detections made by MMOs and PAM operators with different levels of training and experience. The 2017 field program was conducted aboard a research vessel offshore Atlantic Canada in order to mimic a marine mammal monitoring program for a seismic survey as closely as possible. The research and results described here have been submitted for publication (Smith et al. submitted).

4.1 Survey Details

We chartered the RV *Leeway Odyssey*, a 38 m aluminum-hulled oceanographic research vessel, for the survey. The vessel is powered by two 125 KW CAT diesel generators, with a twin-shaft fixed-pitch propulsion system. The superstructure of the *Leeway Odyssey* was modified to accommodate our data collection needs: a custom-built observation booth was installed on the roof of the bridge to provide a vantage point for a second MMO to collect data independently of the MMO on the bridge (Figure 33A); a custom-built platform was welded onto the vessel for installation of the FIRSTnavy sensor (i.e., IR camera; Figure 33B); and for the PAM system, a winch and large block to deploy the THA tow cable and hydrophone array were installed on the back deck.

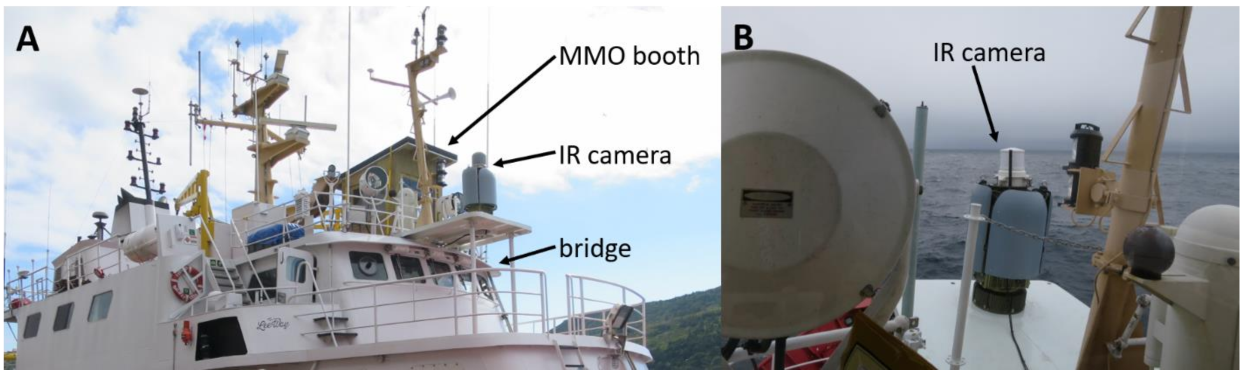

FIGURE 33. The RV *Leeway Odyssey* after installation of the marine mammal observer (MMO) booth on the roof of the bridge and the IR camera (A); a gimbal is necessary to stabilize the IR camera at sea (B). The hydrophone array was deployed from the back deck using an A-frame.

It should be noted that the *Leeway Odyssey* is relatively small in comparison with most commercial seismic vessels, the largest of which are currently ~100 m in length and 70 m in width (e.g., RV *Ramform Hyperion*). The relatively small size of the *Leeway Odyssey* affected maximum platform heights, the extent of roll and pitch, and the vessel noise produced.

The survey was conducted offshore of Nova Scotia and Newfoundland, Canada; the vessel departed Halifax on 30 July and returned on 23 August 2017. The tracklines were modified in response to changing sea conditions such that we were in a "following" sea as much as possible (Figure 34). Vessel speed was maintained as close to 5 knots as feasible to mimic typical seismic survey operation speeds but sped up occasionally to 10 knots at night-time.

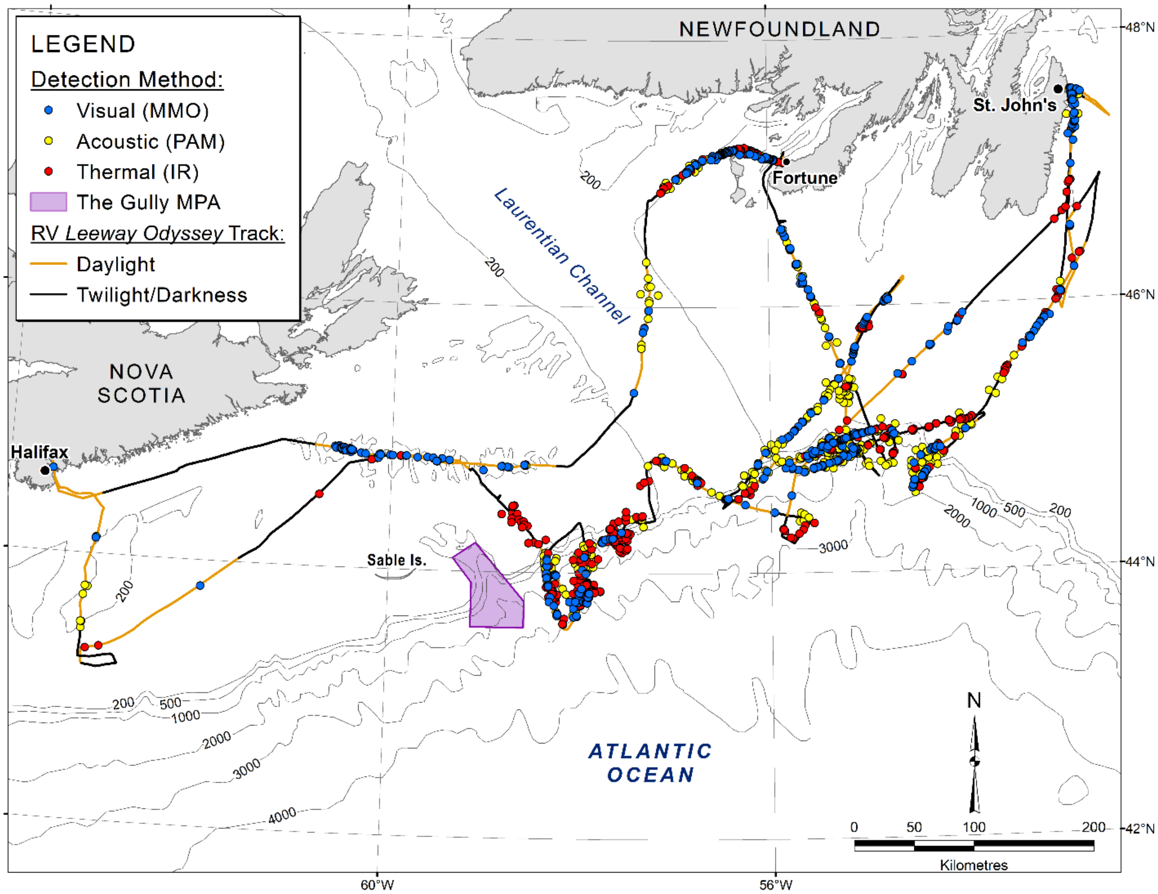

FIGURE 34. Marine mammal detections made aboard the RV *Leeway Odyssey* during 30 July to 23 August 2017. Markers for detections made using the towed hydrophone array (THA, i.e., passive acoustic monitoring) are plotted as mirror-image locations on either side of the vessel trackline when localized and are plotted on the trackline when not localized. Visual observations were made by marine mammal observers (MMOs) during daylight hours, the THA was deployed daily from ~04:00–24:00 (local time), and the IR system ran continuously throughout the survey.

4.2 Terminology Used to Describe Detections

In this study, we use different terms to refer to detections made using visual, acoustic and thermal (IR) methods.

- 'Visual detections' or 'sightings' (terms used interchangeably) are marine mammal detections made by MMOs. All MMO detections were geo-located instantaneously (i.e., bearing and distance were collected and then converted, using a computer algorithm, to provide a geographically-referenced location for the marine mammal).
- 'Acoustic detections' were made by PAM operators using semi-automated software. Localizations were also estimated using semi-automated software, however, not all acoustic detections were localized.

- 'Acoustic localization' refers to the subset of acoustic detections for which a geographic location (i.e., a localization) was estimated.
- 'Automatic (IR) detections' were made using the thermal (IR) system. These were categorized as 'true- or false-positive (IR) detections'. All automatic (IR) detections were geo-located.

Marine mammal detections made using the three different methods, or by MMOs and PAM operators with different levels of experience, were compared at either the level of individual detection or 'encounter'.

- (Single) detections are single cues (e.g., a blow or a dorsal fin) detected by MMOs or the thermal (IR) system. For the PAM system, the series of vocalizations (e.g., click trains or whistles occurring closely in time and/or space) considered by the operator to be made by an individual or group of marine mammals, and which typically resulted in a single localization, was considered a single detection.
- Encounters consist of detections separated in time and/or space from other detections. Repeat detections of the same individual or group of marine mammals were thus considered to belong to the same encounter. Encounters also include multiple groups of animals in the same general vicinity if they occurred within ~5–10 minutes of one another (e.g., groups of dolphins approaching and bow-riding the vessel, humpbacks engaged in surface activity). Detections were subjectively assigned to an encounter during post-processing. Additional factors considered when assigning detections to an encounter included species, group size, sighting location, travel behaviour, and any comments recorded by MMOs or PAM operators. Encounters sometimes included more than one species, e.g., some encounters included both dolphins (one or more species) and pilot whales; and sometimes included animals that were classified differently by different observers (e.g., humpback whale and unidentified baleen whale).

4.3 Data Collection

4.3.1 Visual Observations

Visual observations were made by one of three types of MMOs: experienced, inexperienced and assisted. Experienced MMOs (eMMOs) were biologists with 15+ years of relevant work experience. All eMMOs had previously surveyed marine mammals from shore, vessel and aerial platforms including during seismic surveys offshore Atlantic Canada. Four eMMOs were present during the survey. One inexperienced MMO (iMMO) was used in this study. The iMMO was enrolled in a marine biology undergraduate program, was familiar with marine mammals, but had very limited at-sea experience, and had never been formally trained or collected data using a protocol similar to those employed by MMOs as part of a seismic monitoring and mitigation program. The iMMO received a day of technical training immediately prior to vessel departure. The iMMO was mentored by an eMMO for the first day of data collection on the vessel to ensure that the data collection and recording protocols were understood. The assisted MMOs (aMMOs) were experienced MMOs provided with automatic detections from the IR system via a tablet computer (see Figure 11A); automatic detections were relayed directly to the aMMO, i.e., not verified by a human to remove false positives. The aMMO wore a bluetooth headset over which a "beep" was sounded in real-time to alert him or her to a detection. The aMMO could choose to ignore the alert (e.g., if already engaged in making a visual sighting), or view a six second video snippet of the automatic detection on the data collection tablet. The distance and bearing values to the automatic detection were supplied to the aMMO on the tablet computer along with the video snippet.

Visual observations were made concurrently from two locations: from the bridge and from a booth on the roof of the bridge (see Figure 33A). The approximate viewing heights were 5.4 m and 7.7 m ASL for the bridge and booth, respectively. Both had slightly obstructed views looking forward and to the sides of the vessel; neither had a view of the water directly behind the vessel. The MMO on the bridge was instructed to remain inside the bridge in order to maintain independence from the MMO on the roof. Observations were made from both locations during all daylight hours with the exception of times when environmental conditions made it unsafe for an MMO to ascend to the rooftop observation booth. Each day, individual MMOs made observations from both bridge and roof observation positions such that that the different types of observation effort (i.e., eMMO, iMMO and aMMO) were roughly evenly distributed between the two positions. Maximum observation shift length was 3 h.

MMOs scanned the water's surface for marine mammals with the naked eye or using Fujinon 7 x 50 mm reticle binoculars. Marine mammal sightings were recorded using *Mysticetus* software (by Mysticetus LLC, mysticetus.com). *Mysticetus* simultaneously recorded GPS data, automatically logged the time and vessel position, and calculated the geographic location of each sighting; calculations for sightings made from the bridge and roof accounted for the different viewing heights. MMOs identified marine mammals to the most specific taxonomic rank in which they were confident. The following additional information was collected for each sighting: cue, behaviour, direction and speed of travel, and group size. Individuals were considered to belong to a group if the approximate distance between individuals was within two body lengths for baleen whales; 10 m for dolphins and pilot whales; or five body lengths for seals. Re-sightings of the same individual or group were recorded as such. aMMOs also recorded if they received an IR alert that corresponded with a sighting, and if an alert was received before or after a visual sighting was made (i.e., did the alert assist the aMMO in making the sighting?).

4.3.2 Thermal (IR) Imaging System

The FIRSTnavy sensor and actively stabilized gimbal were installed on the vessel by RDE. The gimbal compensated for up to 12° roll and 10° pitch. The FIRSTnavy sensor had a viewing height of 7.8 m above the vessel's waterline, and a 234.5° field of view (FOV; from 239.1°, past 0°, to 123.6°, where the 0° reference is looking forward; see Figure 33). The FIRSTnavy sensor FOV was slightly obstructed by vessel superstructure ahead and to the sides of the vessel, similar to that of the MMOs. Data were transmitted to a workstation in a lab located aft of the bridge.

Data acquisition and processing were performed with custom developed software (*Tashtego*). An engineer oversaw the functioning of the IR system, and adjusted the detection threshold when the false positive rate increased such that the frequency of alerts was a hindrance to aMMOs. IR data were acquired and recorded 24 hours per day for the duration of the survey. The IR system simultaneously logged GPS data and heading.

4.3.3 Passive Acoustic Monitoring (PAM)

4.3.3.1 Towed Hydrophone Array

A towed hydrophone array (THA) was the primary type of PAM equipment used to make acoustic detections. Hardware details for the THA are provided in Appendix B.

The THA was monitored by both experienced PAM (ePAM) operators and inexperienced PAM (iPAM) operators. The three ePAM operators all had 6+ years' experience in the field of marine mammal research and PAM in particular, and all had expertise in real-time bioacoustics monitoring and data analysis. ePAM operators had significant (4+ years) experience using *PAMGuard* software (used record, visualize, and analyze acoustic signals in real-time) in real-time at sea and were familiar with THA operations and maintenance for research and/or seismic industry applications. The two iPAM operators were biologists with 15+ years' work experience with marine mammals, each had completed two or more shifts as an MMO in offshore seismic surveys, but neither had prior experience with THA or *PAMGuard*. Pre-survey training for the iPAM operators used course materials from the Bio-Waves' Passive Acoustics Technology Training Course. The training consisted of one day of independent review (materials on acoustic theory, digitizing sound, noise and filters, marine mammal sounds, and *PAMGuard* introductory training modules) and a one day in-person training session led by Bio-Waves staff (hardware review and hands-on practice with localization methods in *PAMGuard*). Additional at-sea training for iPAM operators occurred during the initial days of the survey whereby ePAM operators mentored iPAM operators.

The THA was monitored in real-time by ePAM operators at all times during array deployment (~04:00–24:00 daily; all times given as local daylight-saving time, LT); iPAM operators monitored the acoustic signal daily from ~05:00–21:00 during the last two weeks of the survey. Shift lengths ranged from 2–3.7 hours. The iPAM and ePAM operators worked independently of one another (i.e., at spatially separated stations within the acoustic lab) during data collection and analysis, with the exception of occasions when iPAM operators requested technical assistance from ePAM operators related to system or software function.

When on effort, PAM operators visually and aurally monitored the spectrogram and bearing time displays consisting of a frequency range of ~500 Hz to 250 kHz. *PAMGuard* automatic classifiers were parameterized to monitor for a range of echolocating species, and whistles and low frequency moans were monitored using a 50–50,000 Hz range spectrogram window. *PAMGuard* automatically plotted bearings and estimated localizations on a map onscreen. PAM operators viewed these maps and measured the perpendicular distances from the trackline to the individual or group of vocalizing marine mammal(s) using a measuring tool that was part of the *PAMGuard* software. These data were then used, via target motion analysis, to estimate the location of the vocalizing animal or group of animals. Mirror-image localizations (i.e., on either side of the trackline) were also estimated because we were not able to turn in order to resolve the left-right ambiguity. PAM operators also noted any detections that did not result in a localization. PAM operators classified each detection to the highest taxonomic level possible.

4.3.3.2 Sonobuoys

In order to detect, monitor and localize calls of large baleen whales that went undetected on the THA due to vessel and flow noise, additional passive acoustic data were collected using a directional frequency analysis and recording (DIFAR) sonobuoy system. The sonobuoy system was used to monitor low frequencies from ~20–500 Hz, which would include calls from blue and fin whale, as well as other species such as North Atlantic right and sei whales and low-frequency components of humpback whale calls and songs. Sonobuoys were usually deployed when large baleen whales were sighted and when ePAM operators were available to assist and monitor. When possible, sonobuoys were deployed in pairs to allow detected calls to be localized.

Hardware details for the sonobuoy system are provided in Appendix B. Because in-situ monitoring of the towed array was prioritized over real-time sonobuoy monitoring, post-cruise processing of acoustic data from sonobuoys was conducted.

4.3.4 Environmental Data

MMOs recorded, when on effort during daytime, the following environmental data every half hour and when conditions changed: precipitation type (rain, drizzle, fog, or none), Beaufort wind force, and visibility (estimated viewing distance to maximum of 10 km). Air and sea surface temperature (SST) data from the Banquereau Banks buoy at 44.240 N 57.100 W were downloaded from http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/waves-vagues/data-donnees/data-donnees-eng.asp?medsid=C44139; data were logged at 3-hour intervals.

4.4 Data Analysis

Analyses were performed in R 3.5.1 (R Core Team 2018) or Matlab R2017b (The MathWorks Inc. 2017), and results were deemed significant at $\alpha = 0.05$. Note that essentially two sets of IR data were analyzed: 1) the set of automatic IR detections (§ 4.4.1 and 4.4.3), and 2) the set of all detections made by the aMMO (§ 4.4.4; this dataset includes some detections that were not associated in real-time with an IR alert). Sonobuoy data are not included in any statistical analyses because sonobuoys were deployed infrequently, typically when baleen whales were sighted. Hence, sonobuoy effort is biased and overlaps with only a very small fraction of the times during which the other detection methods were employed.

4.4.1 Human Verification of Automatic IR Detections

In order to identify which IR automatic detections were true detections of marine mammals and which detections were false positives, all automatic detections were human-verified. Automatic detections were initially reviewed and classified by MMOs while still at sea (but not while making observations). Classification was subjective and based on the overall appearance of the thermal anomaly in a six-second video snippet (see Figure 8). The subset of detections initially classified as true positives was subsequently reviewed by three independent MMOs and assigned to the highest taxonomic rank in which two or more reviewers had confidence.

4.4.2 Post-processing of Sonobuoy Data

Sonobuoy recordings and ancillary data were reviewed and analyzed by an experienced analyst using *PAMGuard*'s mixed-mode after the survey was completed. Mixed-mode allows recorded GPS and audio data to be integrated and played back during analysis to allow localizations to be efficiently post-processed faster than real-time. Vocalizations were additionally reviewed by an outside expert (i.e., an experienced marine mammal bioacoustician employed by the National Oceanic and Atmospheric Administration, NOAA) who verified classification to the highest taxonomic rank possible. During the post-cruise analysis, localizations were estimated for whale calls received on two or more sonobuoys whenever possible.

4.4.3 Comparison of MMO, PAM (THA) and Automatic IR Detections

Duplicate true positive IR detections (i.e., multiple automatic detections made during a single blow or dorsal fin surfacing) were removed from the dataset: detections made within two seconds and five bearing degrees of one another were identified, and only the first of these was retained. Conditional probabilities (i.e., P(A|B) = the probability that marine mammals were detected using method 'A' given that those same marine mammals were also detected using method 'B') were then calculated to compare detections made using different methods. Detections made using different methods were judged to be of the same marine mammal(s) if they were made within five minutes of each other. Multiple detections made using one method matched to a detection made using another method were judged to be the equivalent of re-sightings of the same marine mammal(s), and therefore only the first of these was retained in the analysis. Conditional probabilities were calculated using only detections made during times when the methods being compared were employed (i.e., 'on effort') concurrently. Conditional probabilities were calculated for all cetacean detections combined, and separately for large whales and small cetaceans (which were mutually exclusive groups). Baleen and sperm whales were included in the 'large whale' category; these species were combined in order to use as much of the dataset as possible as many visual detections made by MMOs and automatic detections made by the IR system could not be classified more specifically. All other toothed whales were categorized as 'small cetaceans'.

Detections made using visual, passive acoustic and IR methods were summarized across categories of three environmental parameters (precipitation, visibility and Beaufort wind force). The proportions of detections made during different categories of environmental conditions were compared between detection methods using a Chi-square test of independence; post-hoc testing of pairs of groups was done using Bonferroni-adjusted p-values (package 'FIFER'; Fife 2014). Categories of environmental conditions were combined when there were fewer than two observations per category for a given method (e.g., rain, drizzle and fog were combined to create a 'precipitation' category). The distributions around the vessel of detections made by MMOs and the IR system were compared using Watson's U^2 test for homogeneity (package 'CircStats'; Lund and Agostinelli 2018).

4.4.4 Comparison of eMMO, aMMO and PAM (THA) Detections

The proportions of sightings or acoustic detections made to the level of species were compared for different observer types (eMMO, aMMO, and ePAM) with Fisher's exact test (McDonald 2014). One-sided tests were used for comparisons of MMOs and PAM operators because MMOs were expected to classify more detections to the species level. Post-hoc tests were assessed using a Bonferroni-modified alpha. The distributions around the vessel of initial localizations for encounters recorded by MMOs (experienced and assisted combined) and ePAM operators were compared using Watson's U^2 test for homogeneity. The initial localization distance for encounters recorded by eMMO, aMMO and ePAM were compared using an ANOVA and the DTK post-hoc test (package 'DTK'; Lau 2013). The timing of the initial localization for encounters recorded by MMOs (experienced and assisted combined) and ePAM operators was compared using a Wilcoxon signed rank test. Conditional probability was calculated as a means to compare encounters recorded by different observer types (eMMO, aMMO, and ePAM) during concurrent effort.

4.4.5 Comparison of eMMO and iMMO Detections

The proportions of sightings classified to species level were compared for eMMOs and the iMMO with Fisher's exact test; tests were made for all sighting distances combined, as well as sighting distances less than or equal to, and greater than, 500 m. This distance was selected because it is the minimum required radius for the safety zone specified in many mitigation guidelines (Weir and Dolman 2007; DFO 2008; DCE et al. 2015; Acosta et al. 2017). Encounters recorded by eMMOs and the iMMO were compared using: Watson's test (distributions around the vessel), and paired t-tests (initial localization distance, and timing of the initial localization). One-sided tests (Fisher's exact test and paired t-tests) were used because the eMMO was expected to perform better than the iMMO. Conditional probability was calculated to compare encounters made by eMMOs and the iMMO during concurrent effort; conditional probabilities were calculated for all sighting distances combined, as well as sighting distances less than or equal to, and greater than, 500 m.

4.4.6 Comparison of Detections Made by ePAM and iPAM Operators

The number of detections, as well as the number of detections localized, were compared for ePAM and iPAM operators. The proportions of detections classified to species level were compared with 1-sided Fisher's exact test. Initial localization distance and timing of encounters recorded by ePAM and iPAM operators were compared using a Wilcoxan signed rank test and a 1-sided paired t-test, respectively. Conditional probability was calculated to compare encounters made by ePAM and iPAM operators during concurrent effort.

4.5 Results

Environmental conditions were generally favorable for sighting marine mammals during the survey: there was limited precipitation (including fog), visibility was predominantly greater than 4 km, and sea state was generally at or below Beaufort wind force 4 (Table 7). Air temperature and SST recorded by the buoy at Banquereau bank averaged 18.49 ± 0.95 °C and 18.23 ± 0.71 °C, respectively (mean ± SD) during 31 July to 23 August 2017.

A total of 2,155 marine mammal detections (all methods combined) was made over the course of the survey (Tables 8 & 9; Figure 34). The majority of detections were made during daylight hours (Table 8).

TABLE 7. Environmental conditions recorded by observers during daylight hours while conducting a survey for marine mammals offshore of Atlantic Canada, August 2017.

Environmental Parameter	Parameter Categories and Percentage of Daylight Hours in Each Category
Precipitation	None (88.1%); Rain (1.8%); Drizzle (1.5%); Fog (7.8%); Mix (0.7%)
Visibility (km)	≤0.5 (3.2%); >0.5 & ≤1 (1.7%); >1 & ≤2 (2.6%); >2 & ≤4 (4.7%); 4+ (87.7%)
Wind Force	0—1 (5.4%); 2 (20.2%); 3 (26.6%); 4 (17.4%); 5+ (30.4%)

TABLE 8. Summary of all data collection effort and marine mammal detections made using visual, passive acoustic monitoring (PAM) and thermal (IR) imaging. Visual detections were made by marine mammal observers (MMOs), PAM detections were made using a towed hydrophone array (THA) or sonobuoys, and IR detections were made by the FIRSTnavy sensor. Daylight hours were from 05:30–20:00 LT; IR effort includes times when the gimbal's stabilization capability was exceeded and ~1 h each night when vessel speed was increased to 10 knots for engine maintenance. Duplicate IR detections have been removed.

Detection Method	Daylight			Twilight/Darkness		
	Effort (h)	No. of Detections	Detection Rate (/h)	Effort (h)	No. of Detections	Detection Rate (/h)
Visual Detections						
Experienced MMO	185.7	241	1.3	3.5	1	0.3
Inexperienced MMO	88.3	80	0.9	0.2	0	0
Assisted MMO	177.7	198	1.1	6.6	1	0.2
Two MMOs	56.6	93	1.6	2.4	0	0
Totals	508.4	612	1.2	12.7	2	0.2
PAM Detections						
Experienced PAM Operator (THA)	189.1	212	1.1	52.4	45	0.9
Inexperienced PAM Operator (THA)	143.5	224	1.6	10	27	2.7
Sonobuoys	10.4	13	1.2	4.3	0	0
Totals	343.0	449	1.3	66.7	72	1.1
Thermal (IR) Detections						
FIRSTnavy sensor	287.7	841	2.9	175.5	179	1.0
Grand Totals, All Methods Combined	1139.1	1902	1.7	254.9	253	0.7

TABLE 9. Visual sightings, by sighting cue, of marine mammals made during a marine mammal survey offshore of Atlantic Canada, August 2017.

Species	Sighting Cue									Species Totals
	Blow	Body	Dorsal Fin	Head	Breach	Fluke	Splash	Birds	Other	
Baleen Whales (Including Unidentified Whales)										
Blue Whale	9									9
Fin Whale	9	5							1	15
Minke Whale	2	2	3							7
Humpback Whale	55	7	1		4	3	3			73
Unidentified Baleen Whale	70	1	3				2		1	77
Unidentified Whale	27		2		1					30
Total	172	15	9	0	5	3	5	0	2	211
Toothed Whales										
Killer Whale			1							1
Long-Finned Pilot Whale	3	20	15		1		3		1	43
White-Beaked Dolphin	1	17	2				6			26
Atlantic White-Sided Dolphin		4	1				3			8
Short-Beaked Common Dolphin	5	66	29		4		43	2		149
Risso's Dolphin	1	2	7		1		3			14
Striped Dolphin							1			1
Sperm Whale	19	3					1			23
Northern Bottlenose Whale							1			1
Unidentified Dolphin	1	18	38		1		67	1	3	129
Unidentified Toothed Whale		1								1
Total	30	131	93	0	7	0	128	3	4	396
Unidentified Cetacean		1								1
Seals										
Grey Seal				1						1
Unidentified Seal				5						5
Total	0	0	0	6	0	0	0	0	0	6
Cue Totals	202	147	102	6	12	3	133	3	6	614

4.5.1 Thermal (IR) System Functioning

The *Leeway Odyssey* is the smallest vessel on which the AIMMMS has been installed to date, and the low platform height where the sensor was installed, 7.8 m compared more than 15 m on other installations, in combination with rolling seas, often exceeded the limit of mechanical stabilization by the gimbal. During these 'stall' times, the camera FOV differed from one frame to the next, thereby removing pixel correspondence between frames, which is the basis for the automatic detection algorithm. By design, the detection algorithm cannot function on a non-stabilized video feed, and therefore performed poorly (i.e., had an increased probability of making false positive detections and decreased probability of making true positive detections) during these times. This lack of stabilization was apparent to MMOs when manually reviewing video snippets because the horizon in the imagery did not remain stable, and the IR engineer noted that this occurred during a substantial proportion of time. As the system was not designed to log periods when the gimbal's limits were exceeded, these periods could not be excluded from any analyses that included detections made only during periods of concurrent effort.

The IR system made 9,189 automatic detections over the course of the survey; 1,501 (16.3 %) of these were initially classified as true positives (Table 10). After undergoing the second review, 74 of these were re-classified as false positives, thereby reducing the total number of true positives to 1,447 (15.5 %). The majority of false positives were birds and waves (41.4 % and 22.8 % of all automatic detections, respectively; Table 10). Shearwaters were the most problematic group of birds, in terms of causing false positive IR detections, encountered during the survey. Shearwaters were automatically detected when rafting on the water's surface during calm seas; landing or scooting across the water's surface prior to getting airborne during calm to moderate sea conditions; and soaring low over the water during windy, large wave conditions (when they would disappear behind the crest of a wave and then re-emerge). Shearwater species encountered were ~90 % Great Shearwaters (*Puffinus gravis*), followed by Sooty (*Ardenna grisea*), Cory's (*Calonectris borealis*), and Manx (*P. puffinus*), in descending order. Larger birds, such as gulls and gannets, were also detected on the water.

True positive IR detections received final broad classifications of whale (223, primarily large whales), dolphin (861, includes small and large dolphins and pilot whales) and cetacean (343). Whereas four automatic detections received a final classification (i.e., agreed upon by two or more reviewers) of sperm whale (Table 10), an additional 20 detections were classified to the species level (sperm whale (4), humpback whale (16)) during the second round of review. However, this level of confidence was held by only a single reviewer in each instance; thus, these detections were classified according to majority rule as baleen whale (6), baleen or sperm whale (5), unidentified whale (2), cetacean (1) and dolphin (6).

TABLE 10. Number and retrospective classification of automatic detections made by the rotating thermal (IR) imaging system. All automatic IR detections were reviewed by humans (marine mammal observers).

Initial Review of All Automatic IR Detections			
True Positives		*False Positives*	
Blow	209	Birds	3804
Body	851	Waves	2093
Dolphin	415	Sun, Cloud, Unknown	1117
Breach, Splash	26	Objects at Sea	592
		Land	82
Totals	1501		7688
Grand Total, Automatic Detections			9189
Second Review of Initial True Positive IR Detections			
True Positives		*False Positives*	
Sperm Whale	4	Bird	12
Baleen Whale	38	Wave	9
Baleen or Sperm Whale	75	Unknown	53
Unidentified Whale	106		
Dolphin, Small	21		
Dolphin, Large	3		
Dolphin	837		
Cetacean	343		
Totals	1427		74
Total, Number of Initial True Positives Reviewed A Second Time			1501

4.5.2 Real-time IR System Use by MMOs While Making Observations

Assisted MMOs were on effort for 177.7 h during the survey (see Table 8), and made a total of 146 sightings, excluding re-sightings, during this time (Table 11). Of these sightings, 144 were classified as cetaceans, and two were classified as unidentified seals; 21.3 % of these sightings (Table 11) were associated with an IR alert in real-time by the aMMO. Of the cetacean sightings, 15 were observed after receiving an IR alert, 16 were observed by the aMMO with the subsequent receipt of an IR alert, and 113 had no associated IR alert in real-time (Table 11). IR alerts associated with sightings were primarily for toothed whales (24 of 146 sightings; 16.4 %). Neither seal sighting was associated with an IR alert.

4.5.3 Summary of Sonobuoy Detections

A total of 13 detections of baleen whales were made over the course of 18 sonobuoy deployments (Table 8 & Table 12). The duration of recordings for individual deployments was influenced by sea state, and ranged from 6–74 minutes. One sonobuoy detection was classified as a humpback whale, and three detections were classified as fin whales. The remainder of detections were classified as 'Balaenopterid' or 'Unidentified baleen whale' (Table 12).

TABLE 11. Marine mammal sightings made by marine mammal observers (MMOs) when assistance by the thermal (IR) imaging system (in the form of automatic detection alerts) was provided. Note that not all detections made by assisted MMOs were the result of assistance from the IR system.

Marine Mammal	**Number of Initial Sightings Associated with IR Alerts**		
	Alert Preceded Sighting	**Alert Received After Sighting**	**No IR Alert Associated with Sighting in Real Time**
Baleen Whales (including unidentified whales)			
Blue Whale	0	0	2
Fin Whale	0	1	2
Minke Whale	0	0	1
Humpback Whale	0	1	11
Unidentified Baleen Whale	3	1	9
Unidentified Whale	0	1	3
Total No. and % of All Marine Mammals	3 (2.1 %)	4 (2.7 %)	28 (19.2 %)
Toothed Whales			
Long-Finned Pilot Whale	1	0	6
White-Beaked Dolphin	0	0	5
Atlantic White-Sided Dolphin	0	0	2
Short-Beaked Common Dolphin	6	3	33
Risso's Dolphin	1	2	0
Striped Dolphin	0	1	0
Unidentified Dolphin	3	5	36
Sperm Whale	0	1	3
Unidentified Odontocete	1	0	0
Total No. and % of All Marine Mammals	12 (8.2 %)	12 (8.2 %)	85 (58.2 %)
Unidentified Seal	0	0	2
Grand Totals and % of All Marine Mammals	15 (10.3 %)	16 (11.0 %)	115 (78.8 %)

TABLE 12. Baleen whale calls recorded with targeted deployments of Directional Frequency Analysis and Recording (DIFAR) sonobuoys. Certainty of species classification are categorized as: certain (CE), probable (PR), or possible (PO). Some baleen whales detected by sonobuoys were also sighted by marine mammal observers (MMO).

Date	Call Start Time	Localized	Species	ID Certainty	Also Detected by MMO
2017-08-13	19:29:59	N	Humpback Whale	PR	Y
			Unidentified Baleen Whale	CE	
2017-08-14	13:54:00	Y	Unidentified Baleen Whale	CE	N
2017-08-14	18:09:59	N	Unidentified Baleen Whale	CE	Y
2017-08-15	13:49:00	N	Balaenopterid	PR	N
2017-08-15	18:08:00	N	Balaenopterid	PR	N
2017-08-15	21:41:00	Y	Balaenopterid	PR	Y
2017-08-16	13:51:00	N	Balaenopterid	PO	N
2017-08-16	17:46:00	N	Balaenopterid	PR	Y
2017-08-16	21:58:00	N	Balaenopterid	PR	N
2017-08-19	12:09:00	N	Fin Whale	PO	Y
2017-08-19	16:20:00	Y	Fin Whale	PO	N
2017-08-19	18:44:00	Y	Fin Whale	PO	N

4.5.4 Comparison of Automatic IR Detections with Visual Sightings and Acoustic Localizations (THA)

Detections made by MMOs and the IR system were observed to overlap the most consistently. The probability that detections made by MMOs were also made by the IR system, and vice versa, ranged from 20 % to 34 %. Of these, the greatest probabilities were for MMOs detecting large whales also detected by the IR system (34 %, Table 13), and for the IR system detecting small cetaceans also detected by MMOs (26 %; Table 13). The largest overlaps in detections made using PAM (THA only) were for small cetaceans: there was a 25 % probability that a small cetacean detection by an MMO was also detected by the THA, and a 47 % probability that a small cetacean detection by the THA was also detected by an MMO (Table 13). In general, the overlap in detections among methods was lowest for PAM and the IR system (range in conditional probabilities: 3–13 %). The probability that large whales were detected by both the THA and another method was less than 10 % (Table 13).

The proportions of detections made during different categories of environmental parameters (Table 14) were found to differ significantly between the three detection methods (Chi-square tests, all $p < 0.001$). Precipitation was rarely experienced during the survey, and accordingly, the highest proportion of detections were made during times without precipitation. When precipitation was present (rain, drizzle and fog, combined), PAM methods resulted in the greatest proportions of detections, followed by detections made by MMOs, and then the IR system (Chi-square post hoc tests, all *adjusted* $p < 0.013$). The IR system made no detections during periods of rain or drizzle (Table 14A). A little more than half of the daytime survey effort occurred in conditions with BF ≤3 (Table 14B). A significantly higher proportion of detections (89 %) was made by the IR system during Beaufort wind force levels of 2–3 than by either MMOs or PAM methods (47 % and 35 %, respectively; all *adjusted* $p < 0.001$). The proportions of detections made by MMOs and using PAM methods were similar for all but one (3 vs. 5+, *adjusted* $p < 0.001$) comparisons of wind force levels (all other *adjusted* $p > 0.136$). MMOs estimated that visibility was 4 km or greater during most of the survey. A greater proportion of detections was made during these relatively good visibility conditions, compared to all lower levels of visibility combined, for all methods. When visibility was 4 km or less, PAM methods resulted in the greatest proportions of detections, followed by detections made by MMOs, and then the IR system (Table 14C; Chi-square post hoc tests, all *adjusted* $p < 0.002$).

TABLE 13. Conditional probabilities for cetaceans detected using visual, passive acoustic monitoring (PAM) and thermal (IR) imaging methods when all three methods were employed concurrently (including times when IR system gimbal stabilization capability was exceeded). Visual detections were made by marine mammal observers (MMOs), PAM detections were made using a towed hydrophone array (THA), and IR detections were made by the FIRSTnavy sensor. Large whales include baleen and sperm whales; small cetaceans include all other toothed whales.

	Conditional Probabilities		
	All Cetaceans	**Large Whales**	**Small Cetaceans**
P(MMO|IR)	0.21	0.34	0.20
P(MMO|THA)	0.28	0.09	0.25
P(IR|MMO)	0.28	0.21	0.26
P(IR|THA)	0.13	0.03	0.09
P(THA|MMO)	0.42	0.10	0.47
P(THA|IR)	0.12	0.04	0.11

TABLE 14. Effort and number of marine mammal detections made using visual, passive acoustic monitoring (PAM) and thermal (IR) imaging methods during concurrent effort, categorized according to environmental conditions recorded by marine mammal observers (MMOs). Visual detections were made by MMOs, PAM detections were made using a towed hydrophone array (THA), and IR detections were made by the FIRSTnavy sensor. Effort includes times when the IR system gimbal stabilization capability was exceeded.

Environmental Parameters and Categories	**Observation Effort (h)**	**Marine Mammal Detections According to Type of Effort**		
		Visual	**PAM (THA)**	**IR**
A. Precipitation				
None	178.8	426	399	741
Rain	0.3	1	1	0
Drizzle	1.4	2	6	0
Fog	16.7	16	34	7
B. Beaufort Wind Force				
0–1	10.6	1	0	0
2	39.1	67	60	137
3	60.4	140	94	526
4	27.8	131	135	71
5+	59.2	106	151	14
C. Visibility (km)				
≤0.5	6.9	6	12	2
≤0.5 & ≤1	3.8	2	13	1
>1 & ≤2	5.1	8	10	0
>2 & ≤4	7.1	4	12	0
≤4	*22.9*	*20*	*47*	*3*
4+	174.3	425	393	745
Totals	197.2	445	440	748

The radial distribution of detections around the vessel differed significantly for automatic IR detections as compared to visual sightings (Watson's $U^2_{1020, 614} = 1.75$, $p < 0.001$). Visual sightings were fairly evenly distributed around the front and sides of the vessel, whereas the greatest proportion of IR detections was made along the starboard side of the vessel. The small number of detections made towards the stern of the vessel were made only by MMOs (Figure 35A); these detections were not in the FIRSTnavy sensor's FOV.

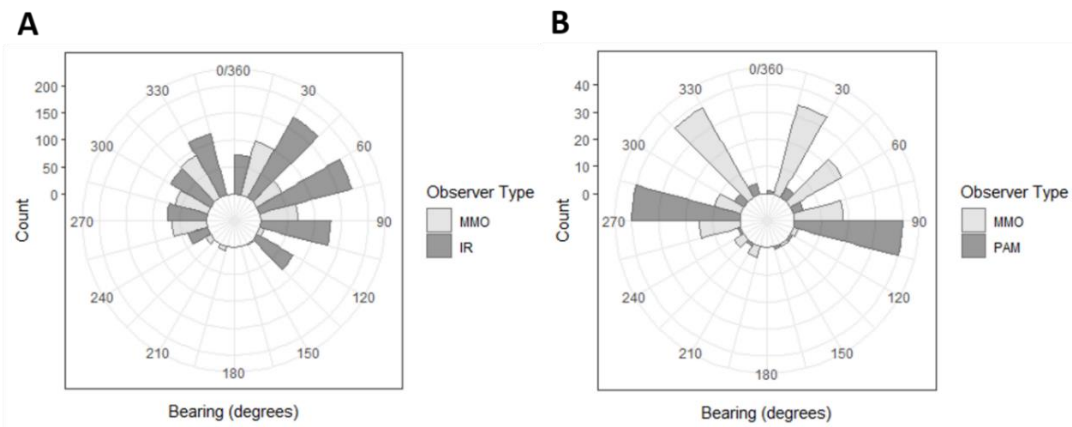

FIGURE 35. Distribution around the vessel of marine mammals: (A) sightings by marine mammal observers (MMOs) and the automatic detections by the IR system, (B) encounters recorded by MMOs and passive acoustic monitoring (PAM; using towed hydrophone array) observers during concurrent effort. Data are visualized using 30° bins, paired by observer type. Note that multiple detections of marine mammals were made during some encounters.

4.5.5 Comparison of Detections Made Visually (With and Without IR Assistance) and Passive Acoustically (Using the THA)

During 80 hours of concurrent effort by the eMMO, aMMO and ePAM (THA) operators, the eMMO and ePAM operator each made more detections than did the aMMO. The ePAM operator localized slightly more than half of these detections (Table 15). During the concurrent effort, MMOs (experienced and assisted) detected both baleen and toothed whales whereas the ePAM operator detected only toothed whales (Table 15). MMOs (experienced and assisted) classified more detections to the species level than did PAM operators (1-sided Fisher's exact tests, all $p < 0.001$). The only species classified by PAM operators during the concurrent effort was the sperm whale, whereas MMOs classified species of both baleen and toothed whales (Table 15). Single detections of a humpback whale and a Risso's dolphin (*Grampus griseus*) were made by an ePAM operator during the survey but were not made at a time when both the eMMO and aMMO were also on effort, and are therefore not included in this analysis of data from concurrent effort.

The distribution around the vessel of initial detections for encounters sighted by the MMOs (eMMO and aMMO combined) differed significantly from the distribution of encounters localized by PAM operators (Watson's $U^2_{196, 102} = 2.45$, $p < 0.001$); the majority of PAM localizations were made to either side of the vessel whereas most MMO sightings were made to the front and both sides of the vessel

(Figure 35B). Distance to the initial detections for encounters also differed significantly among observer types ($F_{2,128} = 20.98$, $p < 0.001$); experienced and assisted MMOs made detections at the same mean initial distance (0.8 km), which was significantly less than that for PAM operators (mean = 2.6 km). The initial detection time for encounters made by MMOs (eMMO and aMMO combined) was on average made marginally before PAM operators, though this difference was not significant (Wilcoxon signed rank test: $V = 73$, $p = 0.24$; mean difference in timing of detection = 59.4 s).

TABLE 15. Marine mammal detections made by experienced marine mammal observers (MMOs; with and without assistance from the IR system) and experienced passive acoustic monitoring (PAM; using towed hydrophone array (THA)) operators when all three methods were employed concurrently (including times when the IR system gimbal stabilization capability was exceeded). Note that the assisted MMO received automatic IR alerts from the IR system, but did not always use these alerts to make detections; and that PAM detections were not always localized.

Marine Mammal	**Observer type**		
	MMO Without IR Assistance	**MMO with IR Assistance**	**PAM, THA [No. of Detections Localized]**
Baleen Whales (Including Unidentified Whales)			
Blue Whale	1	1	0
Fin Whale	3	3	0
Minke Whale	0	1	0
Humpback Whale	1	1	0
Unidentified Baleen Whale	6	4	0
Unidentified Whale	4	0	0
Total (No. Classified to Species)	15 (5)	10 (6)	0
Toothed Whales			
Long-Finned Pilot Whale	4	5	0
Blackfish (i.e., Pilot or Killer Whales)	0	0	9 [3]
Atlantic White-Sided Dolphin	1	1	0
Short-Beaked Common Dolphin	36	37	0
Common Dolphin spp.	0	0	5 [1]
Unidentified Dolphin	29	18	49 [33]
Unidentified Dolphin or Beaked Whale	0	0	7 [4]
Sperm Whale	4	4	20 [13]
Total (No. Classified to Species)	74 (45)	65 (47)	90 [54] (20)
Seals			
Unidentified Seal	1	1	0
Grand Total No. Of Detections	90	76	90 [54]

Of the 72 toothed whale encounters detected (and geo-located) during concurrent effort by eMMOs, aMMOs and ePAM operators, 17 were recorded by all three types of observers (Table 16A). More toothed whale encounters were recorded by ePAM operators than by either eMMOs or aMMOs (Table 16A). It was also more likely that an ePAM operator would localize toothed whales also located visually by an MMO (56 %, Table 16B) than an MMO was to locate toothed whales also localized by ePAM (39 %; Table 16B). Twenty-eight toothed whale encounters were recorded by only ePAM operators during this time (Table 16A).

Twelve baleen whale encounters were recorded by MMOs during concurrent effort by eMMOs, aMMOs and ePAM operators. The probability that a baleen whale detected by an aMMO was also detected by an eMMO was marginally higher (62 %) than the probability that a baleen whale detected by an eMMO was also detected by an aMMO (56 %; Table 16B).

TABLE 16. Number and conditional probabilities for marine mammal encounters when experienced and assisted marine mammal observers (MMOs) and experienced passive acoustic monitoring (PAM) operators were on effort concurrently. Note that the aMMO received automatic IR alerts from the IR system, but did not always use these alerts to make detections. PAM detections were made using a towed hydrophone array (THA); only PAM detections that were localized are included. “All cetaceans” includes unidentified whales.

Observer Type	All Cetaceans	Baleen Whales	Toothed Whales
A. Number of Encounters Observed By			
eMMO only	16	4	8
aMMO only	11	3	8
THA only	28	0	28
eMMO and aMMO	12	5	7
eMMO and THA	2	0	2
aMMO and THA	2	0	2
eMMO and aMMO and THA	17	0	17
Total no. of encounters	88	12	72
Summary: THA with or without MMO	49	0	49
Summary: eMMO with or without other methods	47	9	34
Summary: aMMO with or without other methods	40	8	34
B. Conditional Probabilities (No. of Encounters Used in Calculation)			
P(eMMO|aMMO)	0.69 (44)	0.62 (12)	0.71 (60)
P(eMMO|THA)	0.39 (64)	-	0.39 (77)
P(aMMO|eMMO)	0.62 (44)	0.56 (12)	0.71 (60)
P(aMMO|THA)	0.39 (64)	-	0.39 (72)
P(THA|eMMO)	0.40 (77)	0	0.56 (77)
P(THA|aMMO)	0.45 (72)	0	0.56 (72)

4.5.6 Experienced vs. Inexperienced MMOs

Over the course of the 43.5 hours of effort during which eMMO and iMMO shifts overlapped, eMMOs consistently made more sightings than the iMMO, and made significantly more sightings to the level of species than did the iMMO at sighting distances within 500 m of the vessel (1-sided Fisher’s exact test, p = 0.032; Table 17). At distances greater than 500 m, the proportion of sightings classified to species level did not differ significantly between eMMOs and the iMMO (p = 0.223; Table 17).

Of the 33 marine mammal encounters that occurred during concurrent effort by the eMMOs and the iMMO, 14 were recorded by both observer types (Table 18A). Of the other 19 encounters, i.e., those recorded by only one observer type, the iMMO made fewer encounters than the eMMOs at initial sighting distances >500 m. Also, the iMMO was less likely to encounter marine mammals also encountered by eMMOs than the eMMOs were to encounter marine mammals also encountered by the iMMO at initial sighting distances >500 m (Table 18B). At initial sighting distances ≤500 m, the eMMOs and the iMMO were equally likely to encounter the same marine mammals (Table 18B).

TABLE 17. Sightings classified to species level when experienced and inexperienced marine mammal observers (MMOs) were on effort concurrently. The classification during the initial sighting was used when the same individual/group of marine mammals was re-sighted.

Observer type	**Number of Sightings Classified to Species**					
	All Distances		**≤500 m**		**>500 m**	
	Yes	**No**	**Yes**	**No**	**Yes**	**No**
eMMO	39	8	24	1	15	7
iMMO	15	9	12	5	3	4
p-values for 1-sided Fisher's exact text	0.055		0.032		0.223	

TABLE 18. Number and conditional probabilities for marine mammal encounters recorded when experienced and inexperienced marine mammal observers (MMOs) were on effort concurrently. Two of the 14 encounters observed by both the eMMO and the iMMO were excluded from the analyses by distance bins (last two columns) because they were initially sighted at distances <500 m by the iMMO and at distances >500 m by the eMMO.

	Initial Sighting Distance		
	All Distances	**≤500 m**	**>500 m**
A. Number of Encounters Observed By			
eMMO only	14	4	10
iMMO only	5	4	1
eMMO and iMMO	14	8	4
Total no. of encounters	33	16	15
B. Conditional Probabilities (No. of Encounters Used in Calculation)			
P(iMMO|eMMO)	0.50 (33)	0.67 (16)	0.29 (15)
P(eMMO|iMMO)	0.74 (33)	0.67 (16)	**0. 80 (15)**

The distribution of bearings to initial sightings (for a given encounter) did not differ between the eMMOs and the iMMO (Watson's $U^2_{28, 19} = 0.10$, $p > 0.10$). Also, for the 14 encounters recorded by both MMO types (Table 18A), the differences in initial sighting distance and time did not differ significantly between the eMMOs and iMMO. Mean initial sighting distances were 0.8 ± SD 1.1 km and 0.5 ± 0.8 km, respectively (1-sided paired t_{13}= 1.51, p = 0.078). The eMMOs made initial sightings an average of 91.9 seconds before the iMMO (1-sided paired t_{13} = -0.59, p = 0.28).

In general, the iMMO classified fewer baleen whales to species level than the eMMO. Of the 14 encounters recorded by both eMMO and iMMO, all were classified to the species level by the eMMO. Four of the encounters were of baleen whales: a fin and minke whale identified by the eMMO were classified by the iMMO as 'unidentified baleen whale', and a fin and humpback whale identified by the eMMO were eventually also classified as the same species by the iMMO during a re-sighting (i.e., iMMO species ID did not match eMMO species ID during initial sighting). The iMMO was generally better at identifying smaller toothed whales than larger baleen whales. Of the 10 toothed whale encounters recorded by both eMMO and iMMO, six had matching identifications (dolphins or pilot whales); two had multiple species identified by the eMMO, not all of which were identified by the iMMO; one had multiple species identified by the eMMO that did not include the species identified by the iMMO; and the final encounter was identified as an 'unidentified dolphin' by the iMMO.

4.5.7 Experienced vs. Inexperienced PAM

Over the course of the 142.7 hours of observation effort during which ePAM and iPAM shifts overlapped, iPAM operators made more detections than ePAM operators, and localized a somewhat larger proportion of those detections (Table 19A). iPAM operators also made more detections (with or without localizations) per encounter than did ePAM operators, and iPAM operators recorded slightly fewer encounters overall (Table 19B).

TABLE 19. Acoustic detections of marine mammals made using a towed hydrophone array when experienced and inexperienced passive acoustic monitoring (PAM) operators were on effort concurrently. Encounters primarily consist of repeat detections of the same individual or group of marine mammals; detections were subjectively assigned to encounters during post-processing of data.

	Observer Type	
	ePAM	**iPAM**
A. Number of Detections		
Localized	94	156
Not localized	50	65
Total (proportion of total localized)	144 (0.65)	221 (0.70)
Categorized to species level (proportion of total)	29 (0.20)	39 (0.18)
B. Number of Encounters		
Localized	81	88
Not localized	44	31
Total	125	119
Mean no. of detections per encounter	1.2	1.9

ePAM and iPAM operators did equally well at classifying marine mammal acoustic detections to species level (1-sided Fisher's exact test, $p = 0.32$; Table 19A). Sperm whales were the only species-level classification assigned by iPAM operators; ePAM operators made single classifications of humpback whale and Risso's dolphin in addition to sperm whales (Table 20). Considering only encounters recorded by both ePAM and iPAM operators, the two were comparable in terms of marine mammal classifications made. The same identification was reached in 59 of the 84 shared encounters: unidentified dolphin (39); sperm whale (17); blackfish, i.e., pilot whales, killer whales or Risso's dolphins (2); unidentified dolphin and blackfish (1). Of the remaining 25 shared encounters, ePAM and iPAM operators assigned higher taxonomic level classifications in 18 and 5 encounters, respectively. The lower-level taxonomic classifications were primarily 'unidentified dolphin'.

For the 57 encounters that were localized by both ePAM and iPAM operators, the differences in initial sighting distance and time did not differ significantly between the PAM operators. Mean initial sighting distances were 1.7 ± 1.6 km and 1.3 ± 1.0 km, for ePAM and iPAM operators respectively (Wilcoxon signed rank test: $V = 977.5$, $p = 0.232$). The iPAM operators made initial sightings an average of 197.8 s before the ePAM operators (1-sided paired $t_{56} = 2.41$, $p = 0.99$).

TABLE 20. Marine mammals detected and classified by experienced and inexperienced passive acoustic monitoring (PAM) operators when on effort concurrently. Detections were made using a towed hydrophone array; not all detections were localized.

Marine Mammal Group/Species	**Number of Detections**	
	Experienced PAM Operator	**Inexperienced PAM Operator**
Baleen Whale		
Humpback Whale	1	0
Toothed Whales		
Pilot Whale spp.	9	0
Blackfish (i.e., Pilot Whale or Killer Whale)	9	18
Common Dolphin spp.	9	0
Risso's Dolphin	1	0
Unidentified Dolphin	88	164
Sperm Whale	27	39
Totals	143	221
Grand Totals	144	221

Of the 160 marine mammal encounters recorded during concurrent effort by the ePAM and iPAM operators, 41 were recorded by ePAM, 35 were recorded by iPAM, and 84 were recorded by both observers. The probability that an encounter recorded by an iPAM operator was also recorded by an ePAM operator was marginally less (67 %) than the probability that an encounter recorded by an ePAM operator was also recorded by and iPAM operator (71 %).

4.6 Discussion

4.6.1 General Functioning of the Thermal (IR) System

Our results clearly demonstrate the potential for using a vessel-mounted thermal (IR) imaging system to detect marine mammals in real-time, in thermal conditions encountered offshore Atlantic Canada in summertime. Previously, this system had been deployed only in relatively cooler polar and subpolar oceans (Zitterbart et al. 2013), and it was unknown if, and how well, it would detect marine mammals during seismic surveys conducted in relatively warmer ocean surface waters.

Our study also provides the first evidence of routine detection of small cetaceans (dolphins and pilot whales) via the rotating thermal (IR) system. IR detections were made mostly during encounters when dolphins were observed approaching, swimming within ~ hundreds of meters of, and bow-riding our research vessel. Dolphins have previously been detected via handheld binocular thermal imaging system (Sagem MATIS; Baldacci et al. 2005), whereas large whales have previously only been detected with the rotating thermal (IR) imaging system we used (Zitterbart et al. 2013; Zitterbart et al. submitted). The relatively low camera height (7.8 m in this study; ~15.5 m in Baldacci et al. 2005) compared to the 26–28.5 m camera height used in previous studies with the FIRSTnavy sensor facilitated detections closer to the vessel. A minimum detection distance of ~90 m was possible when the FIRSTnavy sensor was mounted at 28.5 m (unpublished data), due to the vertical field of view of the camera.

The ability to classify IR detections only into broad marine mammal categories points to the utility of this system as a 'bell-ringer' for MMOs (Zitterbart et al. 2013) instead of a stand-alone detection system. Currently, this species-level classification of IR detections is done by humans, and not attempted in the IR imaging design. Classification of IR detections can be done in near real-time if the IR system is constantly monitored by a technician experienced in classifying these images. Although screening out false positives before sending IR alerts to an MMO would be beneficial in reducing distractions caused by false alerts, a further real-time classification of true positive IR detections to the species level would not provide the MMO with much additional advantage in terms of improving mitigation ability.

4.6.2 Real-time IR System Use by MMOs While Making Observations

During our 2017 survey, the IR camera system resulted in 84.5 % of the automatic detections being false positives (§ 4.5.1). This equates to ~5.6 false positives per hour averaged over the entire survey, which is comparable to the rate experienced during seven expeditions in the Arctic and Southern oceans (Zitterbart et al. 2013). Sometimes, multiple alerts, often caused by birds or bow-riding dolphins, resulted in several false positive detections within a single minute. It is during these periods of high false positive rates that we faced our greatest challenge in using this system to alert MMOs to the presence of marine mammals. During these encounters, aMMOs found the alerting system to be more of a distraction than a help, and once they became overwhelmed by the high rate of alerts, they simply turned the alerting system off.

The potential for birds to cause false positive detections (also identified by Zitterbart et al. 2013 as the primary cause of false positive detections during their seven expeditions) highlights the importance of considering non-target species in a study area when planning the deployment of thermal (IR) systems. For example, during the shore-based field program at Cape Race in 2016, plunge-diving northern gannets contributed to such a high false positive rate that additional data screening was required (§ 3.4.2). Alternatively, the improvement of the bird tracking algorithm might help reduce the false positive rate in spite of birds being present in high numbers.

4.6.3 Relative Performance of our PAM Systems Compared to THA Used During Seismic Surveys

The comparison of acoustic detections made during our study with those made during seismic surveys illustrates some of the challenges with employing "conventional" THA for mitigation purposes. In making these comparisons, it must be noted that we may have experienced different noise characteristics produced by our relatively small research vessel might have had an effect on the results compared to noise produced by a large seismic vessel; and towing the THA 300 m behind the vessel. We also did not have an active sound source (i.e., airgun array) present, which would have resulted in some percentage of time during which the airgun pulses were present reducing the ability to detect calls.

The encounter rates during our study (~1 encounter per hour, for both ePAM and iPAM monitoring the THA; calculated using detection rates in Table 8 and the mean no. of detections per encounter,

Table 19) were ca. 5.5 times higher than those during a recent seismic program on the Scotian Shelf during May through September (RPS Energy Canada 2014). During the RPS program, the combined detection rate for the six vessels involved in acoustic monitoring was 0.183 detections per hour. Though a decrease in cetacean calling behaviour in the presence of seismic sound has been observed (Blackwell et al. 2015), RPS Energy Canada (2014) reported a greater number of acoustic detections when the sound source was active vs. when the source was inactive (749 and 269 detections during ~4179 and ~1424 h of acoustic monitoring, respectively). Assuming our acoustic encounter rates reflect a higher detection rate because of a lower signal-to-noise ratio, comparisons of the number of THA detections with visual and IR detections made during this study likely overestimate the expected relative performance of a THA during a seismic survey.

The performance of our THA in detecting baleen whales appears to be comparable to that of other THAs used for monitoring programs during seismic surveys. The THA that we deployed was capable of detecting baleen whale vocalizations above ~500 Hz to 1 kHz (from 1 kHz and above on the APC hydrophones, and 500 Hz and above on the HTI-99-UHF hydrophones), but only when the signal-to-noise ratio of the call was well above the ship engine and cavitation noise levels. We made a single detection of a humpback whale using our THA on 13 August 2017. During a five-month monitoring program for a seismic survey conducted offshore Nova Scotia, conventional THAs were used during the first four months supplemented by an ultralow frequency THA designed specifically to detect blue and fin whales in the last month (RPS Energy Canada 2014). A single acoustic detection of an unidentified baleen whale was reported during this program; the whale was detected only acoustically by the PAM operator and was not "seen on the low frequency computer" (RPS Energy Canada 2014). As well, PAM data collected over the course of 76 seismic surveys conducted in UK Continental Shelf (UKCS) waters during 1995–2010 resulted in 772 acoustic detections: only one detection was confirmed as a baleen whale, and a second detection "may" have been a minke whale (Stone 2015b). The majority of the PAM data were collected during the UKCS surveys using THA deployed behind vessels, though some data were from "stationary platform deployments" (Stone 2015b).

The relatively high baleen whale encounter rate (~1.3 encounters per hour) achieved during our sonobuoy deployments is very likely inflated because our sonobuoy deployment effort usually accompanied visual detections of baleen whales. Therefore, detection rates would be biased upwards compared to a systematic or random sampling design. However, the sonobuoy detections in our study serve to illustrate that baleen whales were vocalizing within the detection range of our vessel but were not detected using the THA system. Work is underway however, to develop and test beamforming methods that use multiple hydrophones in seismic streamers (~8 km in length) and have been demonstrated to detect baleen whales during times when no baleen whales were detected using conventional THA systems. (Abadi et al. 2015; Abadi et al. 2017). Beamforming modules in *PAMGuard* have been developed for use with tetrahedral towed arrays and are currently undergoing field tests (Norris et al. 2018).

4.6.4 Detections Made During Darkness, Low Visibility Due to Precipitation, and High Sea State

During this survey, marine mammal detections were made in darkness using both acoustic (THA) and IR methods (see Table 8). IR detection rates were lower in darkness than during daylight (see Table 8). In contrast, the same IR system was deployed during seven vessel-based expeditions to the Arctic and Southern oceans, and in those areas the IR system performed better at night (Zitterbart et al. 2013), though direct comparisons are difficult due the rolling limitations during this study. Any comparison of night vs. day-time detection rates must take into account the potential for there to be circadian patterns in behaviour of the species encountered, as this will affect their availability for detection. For example, circadian patterns have been documented in sperm whale social behaviour at the surface (Watkins et al. 1999), and humpback whale surface feeding behaviour (Friedlaender et al. 2009). Similarly, the detection rate for the THA (ePAM and iPAM combined) during this survey was lower at night than during the day. But again, these detection rates could reflect circadian patterns in cetacean vocal behaviour (e.g., Baumgartner and Fratantoni 2008; Simon et al. 2010; Klinck et al. 2012b; Wang et al. 2016). Note that though our results do not reliably characterize day vs. night vocal behaviour of marine mammals in the study area because PAM data were not collected from 00:00–04:00 LT.

The poor performance of the IR system in rain, fog and drizzle (see Table 14A) during this survey is not surprising given the known effects of fog on LWIR transmission loss and detection range (Winchester and Gimmestad 1982; Beier and Gemperlein 2004), and past performance of IR systems used to detect marine mammals during periods of precipitation or fog (Baldacci et al. 2005; Zitterbart et al. submitted). In areas where fog is routinely present, e.g., offshore Newfoundland and Labrador during much of the summer, an IR system is impacted as much as a human and cannot be relied on to improve overall marine mammal detection rates.

The relatively poor performance of the IR system in this study in sea states greater than BF3 (see Table 14B) can be attributed to the inability of the gimbal to adequately stabilize the camera when in rough seas. This corresponds with reports by Baldacci et al. (2005) that their IR system (which was mounted on a tripod without stabilization on the deck of their research vessel), was ineffective in sea states greater than 2 or 3. In studies where IR detections were made of marine mammals during Beaufort 5, the camera was mounted either on a much larger and more stable research vessel or on the shore (Zitterbart et al. 2013, submitted). Given that ~45 % of this survey was conducted in sea states greater than Beaufort 3, and that seismic survey vessels are generally much larger and more stable than the research vessel used in this study, comparisons of the number of IR detections with visual and acoustic detections made during this study certainly underestimate the expected relative performance of the IR system in higher sea states as it would be deployed during a seismic survey.

4.6.5 Species Detected and Classified Using Different Methods

Experienced MMOs were found to identify both toothed and baleen whales, and to classify to species level for approximately half of the marine mammals detected (≥55 %, see Table 17). In comparison, detections made using either PAM or thermal (IR) methods resulted in a species-level classification less frequently. Approximately 20% of acoustic detections made using the THA were classified to species level; the majority of which were sperm whales, though one humpback whale and one Risso's dolphin were also identified (see Tables 19 & 20). Three of the 13 sonobuoy detections, and 4 of the 1,427 true positive IR detections were also classified to species level (i.e., as fin whales, shown in Table 12; and as sperm whales, in Table 10). It is possible that the number of species-level classifications may be increased for both PAM and IR systems. Acoustic classifications may be improved with the use of automatic species classifiers in *PAMGuard* (e.g., ROCCA or other classification algorithms; Oswald et al. 2013), which can be developed when sufficient vocalization data are available for species of interest in the study area. In addition, the use of an IR camera with a higher focal length may facilitate the ability to classify IR video snippets to species level. Whether or not classifying marine mammals to species level has any consequence for effective mitigation depends on how mitigation guidelines are written. In countries where mitigation action is taken only for particular species (e.g., Canada; DFO 2008) or additional mitigation actions apply to species of concern (e.g., New Zealand; DOC 2013), the ability to classify detections to species level is more critical, and should be considered when selecting detection methods for monitoring.

4.6.6 Maximizing Marine Mammal Detections

Our results suggest different marine mammal detection methods can be combined to complement one another in order to maximize the number of marine mammals detected. During periods of darkness, PAM and thermal (IR) methods are capable of detecting marine mammals (unlike MMOs), and more detections can be obtained by using both PAM and IR methods concurrently than by either method alone. The THA detected odontocetes almost exclusively, and the IR system detected both small cetaceans and large whales (sperm and baleen whales). Because there was little overlap in the marine mammals detected by these methods when employed concurrently (see Table 5), employing both methods at once will result in a greater proportion of marine mammals that are present in an area being monitored, to be detected overall (i.e., fewer missed mammals). For mitigation purposes, reducing the number of missed animals is very important. However, real-time classifications of marine mammals detected using PAM and thermal (IR) methods will likely be primarily above the species level.

During periods of high sea state and reduced visibility due to precipitation, total number of detections can be increased by using both visual (i.e., MMOs) and PAM methods concurrently rather than either method alone; recalling that PAM detections resulted in localizations slightly more than 50 % of the time. The performance of the two methods (in terms of proportion of detections made) was roughly comparable during periods of high sea state. However, in the presence of precipitation, PAM methods resulted in a higher proportion of detections than MMOs. Precipitation, when occurring, was generally light during the survey, and the performance of PAM methods during periods of heavy rain would likely be less than what we experienced, due to increased noise from rain. PAM methods were found to have more detections in common with MMOs than with IR, but in all cases the probability that the same animals would be detected via both methods was less than 47 % (see Table 13). Given that THA localizations were made at mean distances >1.0 km (i.e., beyond the commonly used 500 m safety zone), our results suggest the utility in using a THA in a mitigation setting may be that in some jurisdictions, similar to the IR system, it can function as a "bell-ringer" to alert observers to the presence of marine mammals in the general area. During a five-month monitoring program for a seismic survey conducted offshore Nova Scotia, acoustic detections resulted in 36 ramp-up delays (primarily dolphin species), but no shut downs (RPS Energy Canada 2014).

Overlap in detections made via visual and acoustic methods has also been investigated using data collected during seismic surveys conducted in UKCS waters, 1995–2010 (Stone 2015b). It was found that 52 % of detections were made only by MMOs, 20 % of detections were made only using PAM methods, and the remaining 28 % of detections were made by both MMOs and PAM; illustrating that greater numbers of marine mammals can be detected overall when more than one method is employed.

An added benefit of having MMOs and PAM operators working concurrently is that detections made using one method can be shared in near real-time across monitoring methods. This can make it possible, for example, for some visual sightings to have their classification confirmed acoustically, and for some acoustic detections to be classified to the species level and geo-located (e.g., resolving the mirror-image ambiguity). The relatively high proportion (~53 %) of acoustic detections classified to species level during seismic surveys in UK waters was possible because of visual confirmation (Stone 2015b).

During periods of good visibility, the IR system could be used to enhance the number of detections made by MMOs. Though detections made by MMOs and the IR system overlapped the most consistently, the probability that the same animals were detected by both methods did not exceed 34 %, leaving room to increase the total number of detections (and the proportion of the mammals present that are detected) by employing both methods at once. However, this potential increase is currently hampered by our inability to quickly screen out false alerts from the automatic IR detections delivered to MMOs in real-time in order to make verified sightings. This constraint is evident in the marginally lower detection rate for the aMMO compared to the eMMO (see Table 8); by the relatively large probability that marine mammals will be detected by both aMMO and eMMO concurrently (see Table 16B); and by IR alerts being associated with a marine mammal detection ~21 % of the time (see Table 11).

The potential increase in sightings that may result from assistance by the IR system is likely underestimated using results from our study. It is highly likely that additional true positive detections of marine mammals would have been made by the IR system if the gimbal was able to stabilize the FIRSTnavy sensor during the higher sea-states. Assuming an estimate of 50 % gimbal stall time, twice as many IR detections might have been made from a larger, and more stable, vessel. As well, mounting the FIRSTnavy sensor in a higher location on the vessel, as has been done previously (Zitterbart et al. 2013), likely would also have resulted in higher detection rates because the sensor FOV would be less obstructed by vessel infrastructure, and the detection range of the sensor would be greater with increased height.

The number of marine mammals detected can also be increased by increasing the total number of MMOs on effort together. It is known from previous work that detection rates tend to be higher with two MMOs on duty than with one MMO (Moulton and Lawson 2002; Holst et al. 2018). In this study, the second highest detection rate achieved was for two MMOs on effort together. Automatic detections made by the IR system resulted in the highest overall marine mammal detection rate, but very likely also included the highest proportion of re-sightings across all methods. In a seismic mitigation context, "tracking" of marine mammals as they approach the safety zone is important, in which case, repeated re-sightings of marine mammals would be beneficial.

Employing complementary detection methods simultaneously is recognized as a means to achieve increased detection performance overall (e.g., DFO 2015; Verfuss et al. 2018). In fact, Verfuss et al. (2018) suggest the use of a modelling framework to explore which combinations of methods, in addition to consideration of target species' behaviour, should be used to develop better monitoring strategies and regulations. Keeping in mind the caveats mentioned above, the data collected in this survey could be useful in such a modelling effort.

4.6.7 Influence of MMO Experience Level

In light of the small sample size available to compare the detection performance of inexperienced vs. experienced MMOs, these results should be taken as an indication, as opposed to a broad characterization, of the manner in which level of experience may influence 'quality' of detections made. Our results suggest that iMMOs may be less effective at employing mitigation actions for marine mammals during seismic surveys than eMMOs. Compared to the eMMOs, the iMMO detected fewer marine mammals (see Tables 17 & 18). As well the iMMO may potentially allow marine mammals to approach the sound source more closely and be exposed to a greater number of airgun pulses, than eMMOs. However, this last supposition is based on differences in initial sighting distance and time for the eMMOs vs. the iMMO which were not found to be statistically significant. We highlight these results here to indicate that this requires further study.

Sighting data from 1,121 seismic surveys within the UKCS between 1995 and 2010 were analyzed, and it was found that dedicated MMOs with relevant marine mammal experience prior to becoming an MMO had better detection skills than those without prior experience (Stone 2015b). Sighting rates for experienced MMOs were 3 x higher in all weather, and 2.5 x higher in good weather, compared to inexperienced MMOs. Experienced MMOs were also found to detect animals at greater distances (approximately 1 km vs. 1.5 km, in all or good weather) than inexperienced MMOs (Stone 2015b).

Our results also suggest that because the iMMO was less likely to classify marine mammals to species level than the eMMOs, that iMMOs will be less effective at employing mitigation actions when they are prescribed only for specific marine mammals. For example, in Atlantic Canada, 'target' species typically include (as a minimum) the blue, North Atlantic right, and Scotian Shelf population of northern bottlenose whales (*Hyperoodon ampullatus;* DFO 2008), and mitigation action is taken when

MMOs positively identify one of these prescribed species and determine that it has or is about to enter the safety zone. However, in cases where mitigation guidelines actions apply to all cetaceans (e.g., JNCC 2017), the ability of MMOs to classify marine mammals to species level is not essential.

4.6.8 Influence of PAM Operator Experience Level

We interpret many of the differences in iPAM and ePAM operator performance observed during this study as reflecting a difference in approach to detection taken by the two observer types, as opposed to differences in their capabilities. This difference in approach became obvious when the data were analyzed. We characterize the approach taken by ePAM operators as being conservative: ePAM operators made fewer detections per encounter, (see Table 19), and 'waited' until marine mammals were abeam of the vessel to localize, which impacted both the distance to, and timing of, detections. In contrast, iPAM operators were relatively quick to record detections, and generally estimated locations using fewer bearings. The accuracy of the localizations made by the iPAM operators is very likely not as good as that of the ePAM operators but was not tested in this study. The ability to accurately localize vocalizing marine mammals relative to the safety zone is essential for effective mitigation and requires further investigation.

The results of this study suggest that ePAM and iPAM operators are comparable in detecting vocalizing marine mammals and classifying them to species level (see Table 19A). Similar results were obtained by Stone (2015b) in their analysis of acoustic detections made during seismic surveys within the UKCS between 1995 and 2010. They found no apparent correlation between the number of acoustic detections made and experience level of PAM operators. They also found that of the 772 acoustic detections made, many were classified as unidentified dolphins or cetaceans; and noted that though a number of the detections were classified to species level, this was often the result of visual confirmation (i.e., MMO sightings). In this study, both ePAM and iPAM operators classified the majority of detections as combined species groups (see Table 20). Very few species were identified using PAM during our study compared to the UK study, though we specifically instructed MMOs and PAM operators to work independently of one another, eliminating the opportunity for visual verification. In our study, both ePAM and iPAM operators readily classified vocalizations made by sperm whales, however, more 'challenging' species classifications were made only by ePAM operators (see Table 20). The consequence of not classifying a marine mammal to species level depend on how mitigation guidelines are written. For example, JNCC guidelines (2017) apply to all cetaceans, and Canadian guidelines (DFO 2008) are written such that all "non-identified" cetaceans must be assumed to be target species for which mitigation action is taken. Both examples imply that having PAM operators capable of classifying acoustic detections to species level is not essential for effective mitigation to be employed.

In this study, MMO detection performance and the approach taken to making detections by PAM operators were both shown to differ with experience level. These differences highlight the importance of having properly trained and experienced field personnel for marine mammal monitoring and mitigation duties during seismic surveys. Not only do field personnel need to possess appropriate skills and knowledge, but they must also be trained to consistently make detections following the same protocol. The importance of having properly trained and experienced personnel is made clear in some mitigation guidelines that specify criteria for an MMO or PAM operator to be considered 'qualified' (e.g., JNCC 2017; NMFS 2018).

5.0 SUMMARY AND CONCLUSIONS

To the best of our knowledge, this study is the first attempt to compare marine mammal detections made using multiple methods in real-time at sea. Our choice of research vessel was dictated by logistics, budget and vessel availability, and though the reduced stability it offered means that the potential performance of the IR system is underestimated by our results, correction factors can be estimated to account for this if the data are used in future modelling efforts (as suggested by Verfuss et al. 2018). Despite the challenges we experienced, this project has advanced our knowledge of visual, acoustic and thermal (IR) methods to detect marine mammals at sea.

Our results support the idea that employing more than one marine mammal detection method concurrently will result in improved marine mammal detection performance overall. PAM and thermal (IR) methods were effective during darkness; PAM and visual methods complemented each other during periods of high sea state and low visibility due to precipitation (including fog); and that thermal (IR) methods can be used to enhance visual methods during periods of good visibility, as well as in hazy or misty conditions. Employing PAM and visual methods during periods of good visibility also resulted in more detections than if only visual methods were used.

The vessel-mounted IR camera system was capable of detecting marine mammals in the thermal regime of Atlantic Canada during summer. The IR camera system resulted in 84.5 % of the automatic detections being false positives. This equates to ~5.6 false positives per hour averaged over the entire survey. Sometimes, multiple alerts, often caused by birds or bow-riding dolphins, resulted in several false positive detections within a single minute. It is during these periods of high false positive rates that further work in false alert suppression is necessary.

Detection of marine mammals via PAM, for mitigation purposes, appears to be primarily influenced by hardware and software (as opposed to level of PAM operator training and experience). Thirteen detections of baleen whales were made over the course of ~14 hours of sonobuoy effort compared to a single humpback whale detected using the towed hydrophone array over the course of the entire survey. We are not suggesting that sonobuoys be used for monitoring during seismic surveys but included them in our study to illustrate the importance of matching in-situ capabilities of the PAM hardware, to the extent possible, to detect and localize vocalizations of the key species of interest.

The types of marine mammals detected, and the extent to which they were classified to the species level, varied depending on which detection methods were used. MMOs and the thermal (IR) system effectively detected both baleen and toothed whales at the water's surface, though species were only reliably identified by MMOs. The species detected using PAM methods differed depending whether a towed array (which detected toothed whales almost exclusively) or sonobuoys (which detected baleen whales) were used. Most PAM detections were not classified to species level.

Experience level seemingly influenced MMO detection performance in a number of ways. The inexperienced MMO effectively monitored less of the viewable area around the vessel, detected fewer marine mammals, and was less likely to identify these marine mammals to species level, relative to experienced MMOs. The inexperienced MMO was also generally slower to detect marine mammals and made initial sightings of marine mammals when they were closer to the vessel, relative to the experienced MMOs. These findings suggest that the implementation of mitigation measures related to minimizing the amount of sound exposure (i.e., from an airgun array or other sound source) may be less effective when monitoring is conducted by MMOs that are inexperienced. As well, ensuring that PAM operators are experienced and familiar with vocalizations of marine mammals likely to be encountered in the study area, and have had adequate hands-on experience making detections prior to monitoring for mitigation purposes, should improve the consistency and quality of PAM detections made.

6.0 RECOMMENDATIONS FOR FUTURE STUDY

Our primary recommendation is that future comparisons of detection methods, when the intended application is for monitoring purposes during seismic surveys, be made from aboard a seismic survey vessel(s). This will ensure that a stable and appropriately high platform is available for the IR camera, and that PAM detections will be made in the appropriate acoustic setting, i.e., with airgun array pulses. Securing adequate bunk space for research team members aboard an active seismic survey vessel may be a challenge. One of the advantages we had in chartering our own research vessel is that we had the ability to modify the survey trackline. This allowed us to 'target' areas with marine mammals in an attempt to maximize the number of detections to be analyzed. Conducting future research aboard a working seismic vessel will come with the risk of not encountering adequate marine mammals for detection along the predetermined seismic survey lines and may therefore require a longer study period.

One of the weaknesses in our 2017 dataset is the lack of distance estimates for the IR detections. This prevented us from calculating detection functions for the IR system, which would have allowed for a more meaningful comparison of methods. Ensuring that future studies provide a stable platform for the IR system will avoid this issue.

Large numbers of false positive automatic IR detections were encountered during both the shore- and vessel-based field programs. As such, the thermal (IR) system requires further study and refinement so that the delivery of IR alerts to MMOs in real-time can be effectively used as a monitoring tool.

Our results also suggest that PAM methods other than the use of conventional THA coupled with target motion analysis be explored as a means to detect marine mammals acoustically during seismic surveys. Future studies that explore beamforming methods, cross-bearing methods or hydrophone arrays capable of localizing marine mammals in 3-D space in the context of mitigation would be useful.

7.0 ACKNOWLEDGEMENTS

This project was funded by the Environmental Studies Research Fund (Project #2014-03S). We thank our ESRF Regional Research Manager, D. Taylor, and the ESRF Secretariat for their support during this project. LGL contributed in-kind salary support for HRS and VDM during manuscript preparation and final reporting. AWI contributed in-kind salary support for DPZ during project implementation and OB during manuscript preparation and final reporting. Part of DPZ's salary time for detector development was supported by ONR Award N000141310856 to AWI.

2015: We thank M. Holst, A. Davis, J. Beland, K. Keats, J. Coffey (MMOs, LGL) and M. Flau (engineer, AWI) for data collection. We thank B. Mactavish, S. Penney-Belbin, A. Murphy, R. Martin, S. Besaw, J. Christian, P. Abgrall, T. Elliott (all LGL), and A.L. Bohlen (AWI) for assistance with logistics related to mobilizing, maintaining and de-mobilizing our remote field site. We thank Dawn Boutilier (DFO) for assisting us in obtaining the Licence to Occupy our field site at the Cape Race Lightstation Lands; the ModSpace delivery crew for delivering and setting up our office trailer at Cape Race; W. Costello for hooking up our electricity; and E. Evans for building the platform for the thermal (IR) camera and assisting with trailer modifications. We thank the people of Portugal Cove South for their assistance and hospitality, in particular D. and J. Shepherd, the staff at the Irish Loop Express, the Myrick Wireless Interpretive Centre at Cape Race, and the lighthouse keepers.

2016: We thank J. Beland, M. Holst, A. Davis, B. Mactavish (MMOs, LGL); and M. Flau (engineer, AWI) for data collection, and B. Mactavish and K. Knowles for covering the night shift. We thank C. Jones, S. Penney-Belbin, A. Murphy, R. Martin, S. Besaw, J. Christian, P. Abgrall, T. Elliott (all LGL) for assistance with logistics related to mobilizing, maintaining and de-mobilizing our remote field site. We thank Barbara McDonald (DFO) for assisting us in obtaining the Licence to Occupy our field site at the Cape Race Lightstation Lands; Alantra Leasing Inc. for delivering our office trailer to Cape Race; Weirs Construction for delivering the jersey blocks to Cape Race in thick fog and helping us to properly anchor the trailer; W. Costello for hooking up our electricity; and I. Goudie for building the MMO booths, installing the platform for the thermal (IR) camera, and assisting with trailer modifications. We P. Coombs for skippering the vessel used for the calibration study. We thank all the people of Portugal Cove South for their assistance and hospitality, in particular A. Coombs, D. and J. Shepherd, the staff at the Irish Loop Express, the Myrick Wireless Interpretive Centre at Cape Race, the Mistaken Point Ecological Reserve, and the lighthouse keepers.

2017: We thank P. Abgrall, A. Davis, M. Flau, C. Jones, E. Griffiths, P. Gruden, T. Lang, N. Riddoch, M. Sanchez and T. Thomas for data collection. We thank T. Norris and E. Ferguson (Bio-Waves Inc) for assistance in developing PAM methodology, field logistics, and analysis of PAM data; M. Oswald and B. Miller for sonobuoy post-processing; and D. Cholewiak (NMFS) for sonobuoy call review. We also thank D. Steckler and P. Donlan for support with *Mysticetus* software, and in particular, system configuration during vessel mobilization; the entire staff at LeeWay Marine for their efforts in preparation of our survey, as well as the *Leeway Odyssey* crew; Rheinmetall Electronics GmbH for camera installation and demobilization; H. Moors-Murphy (DFO) for advice on survey planning; and R. Martin, S. Besaw, and T. Jarvie (LGL) for help with contracts and logistics.

We thank S. Besaw, C. Jones, and G. Rayner for assistance with final report preparation.

8.0 LITERATURE CITED

Abadi, S.H., W.S.D. Wilcock, M. Tolstoy, T.J. Crone, and S.M. Carbotte. 2015. Sound source localization using data recorded by hydrophone streamers during seismic surveys. J. Acoust. Soc. Am. 138, 3951–3963. http://dx.doi.org/10.1121/1.4937768.

Abadi, S.H., M. Tolstoy, and W.S.D. Wilcock. 2017. Estimating the location of baleen whale calls using dual streamers to support mitigation procedures in seismic reflection surveys. PloS ONE 12(2): e0171115. https://doi:10.1371/journal.pone.0171115.

Acosta, A., N. Niño-Rodríguez, M.C. Yepes, and O. Boisseau. 2017. Mitigation provisions to be implemented for marine seismic surveying in Latin America: a review based on fish and cetaceans. Aquat. Biol. 26, 199–216. https://doi.org/10.3354/ab00688.

Baldacci, A., M. Carron, and N. Portunato. 2005. Infrared detection of marine mammals. Technical report SR-443. NATO Undersea Research Centre.

Barlow, J., M.C. Ferguson, W.F. Perin, L. Balance, T. Gerrodette, G. Joyce, C.D. MacLeod, K. Mullin, D.L. Palka, and G. Waring. 2006. Abundance and densities of beaked and bottlenose whales (family Ziphiidae). J. Cetac. Res. Manage. 7, 263–270.

Baumgartner, M.F. and D.M. Fratantoni. 2008. Diel periodicity in both sei whale vocalization rates and the vertical migration of their copepod prey observed from ocean gliders. Limnol. Oceanogr. 53, 2197–2209.

Beier, K. and H. Gemperlein. 2004. Simulation of infrared detection range at fog conditions for enhanced vision systems in civil aviation. Aerospace Science and Technology. 8, 63–71.

Blackwell, S.B., C.S. Nations, T.L. McDonald, A.M. Thode, D. Mathias, K.H. Kim, C.H. Greene, and A.M. Macrander. 2015. Effects of Airgun Sounds on Bowhead Whale Calling Rates: Evidence for Two Behavioral Thresholds. PLoS ONE 10(6): e0125720. https://doi.org/10.1371/journal.pone.0125720.

Boebel, O. and D.P. Zitterbart. 2014. Exploring the thermal limits of IR-based automatic whale detection (ETAW). Accessed at http://www.onr.navy.mil/reports/FY14/mbboebel.pdf on 10 April 2019.

Compton R., L. Goodwin, R. Handy, and V. Abbott. 2008. A critical examination of worldwide guidelines for minimising the disturbance to marine mammals during seismic surveys. Marine Policy. 32, 255–262.

Cuyler, L.C., R. Wiulsrød, and N.A. Øritsland. 1992. Thermal infrared radiation from free living whales. Mar. Mam. Sci. 8, 120–134.

DCE (Danish Centre for Environment and Energy), GNIR (Greenland Institute of Natural Resources), EAMRA (Environmental Agency for the Mineral Resource Activities, Greenland). 2015. Offshore seismic surveys in Greenland: Guidelines to best environmental practices, environmental impact assessments and environmental mitigation assessments. 48 pp. Available at: https://www.govmin.gl/images/Documents/Terms_guidelines_hydrocarbons/Environmental_Impact_Assessment_Guidelines.pdf.

DFO (Fisheries and Oceans Canada). 2008. *Statement of Canadian practice with respect to the mitigation of seismic sound in the marine environment*. Fisheries and Oceans Canada. 5 pp.

DFO. 2015. Review of mitigation and monitoring measures for seismic survey activities in and near the habitat of cetacean species at risk. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2015/005.

DFO. 2018. Oceanographic Conditions in the Atlantic Zone in 2017. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2018/039.

DOC (Department of Conservation, New Zealand). 2013. Code of conduct for minimising acoustic disturbance to marine mammals from seismic survey operations. Available at: https://www.doc.govt.nz/Documents/conservation/native-animals/marine-mammals/seismic-survey-code-of-conduct.pdf.

Erbe, C., R. Dunlop, and S. Dolman. 2018. Effects of noise on marine mammals. p277–309 In H. Slabbekoorn, R.J. Dooling, A.N. Popper and R.R. Fay (eds.), Effects of Anthropogenic Noise on Animals, Springer Handbook of Auditory Research 66. Springer Science+Business Media, LLC.

Fife, D. 2014. fifer: A collection of miscellaneous functions. R package version 1.0. https://CRAN.Rproject.org/package=fifer.

Friedlaender, A., E.L. Hazen, D.P. Nowacek, P.N. Halpin, C. Ware, M.T. Weinrich, T. Hurst, and D. Wiley. 2009. Diel changes in humpback whale (Megaptera novaeangliae) feeding behavior in response to sand lance (Ammodytes spp.) behavior and distribution. Mar. Ecol. Prog. Ser. 395, 91–100. https://doi:10.3354/meps08003.

Gisiner, R.C. 2016. Sound and marine seismic surveys. Acoustics Today 12, 10–18.

Gordon, J., D. Gillespie, J. Potter, A. Frantzis, M.P. Simmonds, R. Swift, and D. Thompson. 2003. A review of the effects of seismic surveys on marine mammals. Mar. Technol. Soc. J. 37, 16–34.

Haver, S., H. Klinck, S. Nieukirk, H. Matsumoto, R. Dziak, and J. Miksis-Olds. 2017. The not-so-silent world: measuring Arctic, Equatorial, and Antarctic soundscapes in the Atlantic Ocean. Deep Sea Res. Part 1 Oceanogr. Res. Pap. 122, 95–104.

Holst, M., V.D. Moulton, T. Gerwing, and R.E. Harris. 2018. Review of Marine Mammal Monitoring and Mitigation Data and Outcomes for 2-D Seismic Surveys in the Canadian Beaufort Sea, 2006–2012. LGL Report FA0150. Report by LGL Limited, St. John's, NL for Fisheries and Oceans Canada, Winnipeg, MB. 31 p.

Horton, T.W., A. Oline, N.D. Hauser, T.M. Khan, A. Laute, A. Stoller, K. Tison, and P. Zawar-Reza. 2017. Thermal imaging and biometrical thermography of humpback whales. Front. Mar. Sci. 4:424. https://doi.org/10.3389/fmars.2017.00424

JNCC (Joint Nature Conservation Committee). 2017. Guidelines for Minimizing Acoustic Disturbance to Marine Mammals from Seismic Surveys. Jncc.defra.gov.uk Seismic Survey. [online] Available at: jncc.defra.gov.uk/pdf/jncc_guidelines_seismicsurvey_aug2017.pdf [Accessed 13 Dec 2018].

Klinck H., S.L. Nieukirk, D.K. Mellinger, K. Klinck, H. Matsumoto, and R.P. Dziak. 2012a. Seasonal presence of cetaceans and ambient noise levels in polar waters of the North Atlantic. J. Acoust. Soc. Am. 132, EL176; https://doi.org/10.1121/1.4740226

Klinck, H., D.K. Mellinger, K. Klinck, N.M. Bogue, J.C. Luby, W.A. Jump, G.B. Shilling, T. Litchendorf, A.S. Wood, G.S. Schorr, and R.W. Baird. 2012b. Near-real-time acoustic monitoring of beaked whales and other cetaceans sing a SeagliderTM. PLoS ONE 7(5): e36128. doi:10.1371/journal.pone.0036128

Lau, M.K. 2013. DTK: Dunnett-Tukey-Kramer Pairwise Multiple Comparison Test Adjusted for Unequal Variances and Unequal Sample Sizes. R package version 3.5. https://CRAN.R-project.org/package=DTK

Lund, U. and C. Agostinelli. 2018. CircStats: Circular Statistics, from "Topics in Circular Statistics" (2001). S-plus original by Ulric Lund and R port by Claudio Agostinelli. R package version 0.2-6. https://CRAN.R-project.org/package=CircStats

McDonald, J.H. 2014. Handbook of biological statistics (3rd ed.). Sparky House Publishing, Baltimore, Maryland.

Miksis-Olds, J. and S.M. Nichols. 2016. Is low frequency ocean sound increasing globally? J. Acoust. Soc. Am 139, 501–511.

Moulton, V.D. and J.W. Lawson. 2002. Seals, 2001. p. 3-1 to 3-48 In: W.J. Richardson and J.W. Lawson (eds.), Marine mammal monitoring of WesternGeco's open-water seismic program in the Alaskan Beaufort Sea, 2001. LGL Rep. TA2564-4. Rep. from LGL Ltd., King City, Ont., for WesternGeco LLC, Anchorage, AK, BP Exploration (Alaska) Inc., Anchorage, AK, and Nat. Mar. Fish. Serv., Anchorage, AK, and Silver Spring, MD. 95 p. Available from the AFSC/MML Library, Seattle.

NMFS (National Marine Fisheries Service). 2018. Takes of marine mammals incidental to specified activities: taking marine mammals incidental to geophysical surveys in the Atlantic Ocean. Fed. Regist. 83(235, 7 Dec):63268–63381.

Norris, T., G. D'Spain, and G. Gillespie. 2018. The 3-D-V Array: A volumetric, digital towed hydrophone array system capable of bearing and location estimation in 3-D space. (Abstract for an oral presentation at the 6th International Meeting on the Effects of Sound in the Ocean on Marine Mammals (ESOMM) and the 3rd Programme Review Meeting of the E&P Sound & Marine Life Joint Industry Programme (JIP) in The Netherlands, The Hague. September 2018).

Nowacek, D.P., C.W. Clark, D. Mann, P.J. Miller, H.C. Rosenbaum, J.S. Golden, M. Jasny, J. Kraska, and B.L. Southall. 2015. Marine seismic surveys and ocean noise: time for coordinated and prudent planning. Front. Ecol. Environ. 13, 378–386.

Oswald, J.N., S. Rankin, J. Barlow, M. Oswald, and M.O. Lammers. 2013. Real-time Call Classification Algorithm (ROCCA): software for species identification of delphinid whistles. pp. 245–266 In: Detection, classification and localization of marine mammals using passive acoustics, 2003-2013: 10 years of international research. DIRAC NGO, Paris, France.

Perryman W.L., M.A. Donahue J.L. Laake, and T.E. Martin. 1999. Diel variation in migration rates of eastern Pacific gray whales measured with thermal imaging sensors. Mar. Mam. Sci. 15(2), 426-445.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Richardson, W.J., C.R. Greene, C.I. Malme, and D.H. Thomson. 1995. Marine mammals and noise. San Diego, CA. Academic Press.

Romano, T.A., M.J. Keogh, C. Kelly, P. Feng, L. Berk, C.E. Schlundt, D.A. Carder, and J.J. Finneran. 2004. Anthropogenic sound and marine mammal health: measures of the nervous and immune systems before and after intense sound exposure. Can. J. Fish. Aquat. Sci. 61(7), 1124–1134.

RPS Energy Canada. 2014. Wildlife Observation Report BP Tangier 3D WATS Seismic Survey Halifax, Nova Scotia 17 May to 14 September 2014. 46 pp. + appendices

Simon M., K.M. Stafford, K. Beedholm, C.M. Lee, and P.T. Madsen. 2010. Singing behavior of fin whales in the Davis Strait with implications for mating, migration and foraging. J Acoust. Soc Am. 128, 3200–10. pmid:21110615.

Smith H.R., D.P. Zitterbart, M. Flau, T.F. Norris, E.L. Ferguson, C. Jones, O. Boebel, and V.D. Moulton. Submitted. Comparisons of marine mammal detections made using visual, acoustic and thermal IR imaging methods, and made by observers with different levels of experience, offshore of Atlantic Canada.

Southall, B.L., A.E. Bowles, W.R. Ellison, J. Finneran, R. Gentry, C.R. Green, C.R., Kastak, D.R. Ketten, J.H. Miller, P.E. Nachtigall, W.J. Richardson, J.A. Thomas, and P.L. Tyack. 2007. Marine mammal noise exposure criteria. Aquat. Mamm. 33, 411-521, DOI 10.1578/AM.33.4.2007.411.

Southall, B.L., J.F. Finneran, C. Reichmuth, P.E. Nachtigall, D.R. Ketten, A.E. Bowles, W.T. Ellison, D.P. Nowacek, and P.L. Tyack. 2019. Marine mammal noise exposure criteria: updated scientific recommendations for residual hearing effects. Aquat. Mamm. 45, 125–232, DOI 10.1578/AM.45.2.2019.125.

Stone, C.J. 2015a. Marine mammal observations during seismic surveys from 1994-2010. JNCC report, No. 463a.

Stone, C.J. 2015b. Implementation of and considerations for revisions to the JNCC guidelines for seismic surveys. JNCC report, No. 463b.

The MathWorks Inc. 2017. Matlab R2017b. Natick, Massachusetts, United States.

Verfuss UK., D. Gillespie, J. Gordon, T.A. Marques, B. Miller, R. Plunkett, J. Theriault, D. Tollit, D.P. Zitterbart, P. Hubert, and L. Thomas. 2016. Low Visibility real-time marine mammal monitoring techniques review, Report Number SMRUOGP2015-002, provided to OGP, June 2016.

Verfuss UK., D. Gillespie, J. Gordon, T.A. Marques, B. Miller, R. Plunkett, J. Theriault, D. Tollit, D.P. Zitterbart, P. Hubert, and L. Thomas. 2018. Comparing methods suitable for monitoring marine mammals in low visibility conditions during seismic surveys. Mar. Pollut. Bull. 126, 1–18.

Wang, D., 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. 2016. Vast assembly of vocal marine mammals from diverse species on fish spawning ground. Nature. 531, 366–370.

Watkins, W.A., M.A. Daher, N.A. DiMarzio, A. Samuels, D. Wartzok, K.M. Fristrup, D.P. Gannon, P.W. Howey, and R.R. Maiefski. 1999. Sperm whale surface activity from tracking by radio and satellite tags. Mar. Mam. Sci. 15, 1158–1180.

Weir C.R. and S.J. Dolman. 2007. Comparative review of the regional marine mammal mitigation guidelines implemented during industrial seismic surveys, and guidance towards a worldwide standard. J. Int. Wildl. Law Policy. 10, 1–27.

Whitehead H. and J.E. Carscadden. 1985. Predicting inshore whale abundance -- whales and capelin off the Newfoundland coast. Can. J. Fish. Aquat. Sci. 42:976-981.

Wikipedia. 2019. Deep learning. Accessed at https://en.wikipedia.org/wiki/Deep_learning on 10 April 2019.

Wikipedia. 2019. Machine learning. Accessed at https://en.wikipedia.org/wiki/Machine_learning on 10 April 2019.

Winchester, L.W. Jr., and G.G. Gimmestad. 1982. Propagation of visible and infrared radiation by fog, rain, and snow. Keweenaw Research Center, Michigan Technological University Report No. 12685 for the U.S. Army Tank-Automotive Command Research and Development Center. 97 pp.

Wright, A., F. Robertson, and A.M. Cosentino. (eds) 2016. Incorporating new mitigation technologies into guidelines for seismic surveys and other underwater acoustic activities: Producing performance standards. Marine Species and Threats, Department of Conservation, Wellington, New Zealand. 27 p.

Zitterbart, D.P., L. Kindermann, E. Burkhardt, and O. Boebel. 2013. Automatic round-the-clock detection of whales for mitigation from underwater noise impacts. PLoS ONE, 8 (2013), p. e71217.

Zitterbart, D.P., H.R. Smith, M. Flau, S. Richter, E. Burkhardt, J. Beland, L. Bennet, A. Cammareri, A. Davis, M. Holst, E. Kniest, C. Lanfredi, H. Michel, M. Noad, K. Owen, and O. Boebel. submitted. Scaling the laws of thermal imaging-based whale detection.

APPENDIX A – BEAUFORT WIND FORCE SCALE

Interrelationships of wind speed, Beaufort wind force, and wave heights on the open sea (Table 5.1 from Richardson et al. 1995).

Wind Speed		Beaufort wind force	World Meteorological Organization Terms	Wave Height		Description
Knots	**m/s**			**ft**	**m**	
<1	<0.5	0	Calm	0	0	Glassy like a mirror
1-3	0.5-1.5	1	Light air	0-1	0-0.2	Ripples with the appearance of scales but no whitecaps or foam crests
4-6	2.1-3.1	2	Light breeze	1-2	0.2-0.5	Small wavelets, crests have a glassy appearance but do not break (no whitecaps)
7-10	3.6-5.1	3	Gentle breeze	2-3.5	0.5-1	Smooth large wavelets, crests begin to break, occasional/ scattered whitecaps
11-16	5.7-8.2	4	Moderate breeze	3.5-6	1-2	Slight; small fairly frequent whitecaps
17-21	8.7-10.8	5	Fresh breeze	6-9	2-3	Moderate waves becoming longer, some spray, frequent moderate whitecaps
22-27	11.3-13.9	6	Strong breeze	9-13	3-4	Rough, larger waves, longer-formed waves, many large whitecaps. Some airborne spray.
28-33	14.4-17.0	7	Near gale	13-19	4-5.5	Some foam from breaking waves is blown into streaks along wind direction.
34-40	17.5-20.6	8	Gale	18-25	5.5-7.5	Well-marked streaks of foam are blown along wind direction.
41-47	21.1-24.2	9	Strong gale	23-32	7-10	High waves whose crests sometimes roll over.
48-55	24.7-28.3	10	Storm	29-41	9-12.5	Very high waves with overhanging crests.
56-63	28.8-32.4	11	Violent storm	37-52	11.5-16	Exceptionally high waves.

APPENDIX B – PAM HARDWARE

B.1 Towed Hydrophone Array (THA)

- THA consisted of four hydrophone elements, pre-amplifier circuit boards, and pressure sensors, housed in a mineral oil-filled flexible tube.
- Two high-frequency hydrophones (High Tech, Inc. HTI-99-UHF elements; effective frequency response from 2 Hz to 250 kHz), were spaced 0.5 m apart, and positioned mid-way between a pair of mid-frequency hydrophones (APC International 42-1021 elements; effective frequency response from ~1–100 kHz) that were spaced 3 m apart.
- Custom designed hydrophone pre-amplifiers and pressure sensors (Kellar 7SE) were integrated into the array. Pre-amplifiers for the high frequency hydrophones were designed and integrated by the supplier (HTI) with 38 dB of gain and a high-pass corner frequency of 500 Hz. Preamplifiers for the mid-frequency hydrophones had 40 dB of gain, a 250 Hz high pass filter, and a 35 kHz low pass filter.
- The tow-cable was weighted with lead rope that was attached a few meters aft of the array to sink it below the water's surface.
- The array was deployed 300 m behind the stern of the vessel using a large block and a hydraulic winch and drum system.
- Pre-amplified analog signals from the hydrophones were passed from the array up a copper-wire cable into an acoustic processing system (APS) for signal conditioning, digitization and subsequent monitoring and recording.
- The APS was powered by a 12 V DC battery bank, that was independent of the vessel's power system, in order to avoid electrical noise which is typically introduced by the vessel's power.
- Additional measures taken to minimize noise in the acoustic signal included electronic shielding of all acoustic analog cables from the array to the APS and avoiding the use of high-power electronics and radio-wave emitting (e.g., VHF radio) devices in the acoustics lab on the vessel.
- In the APS, analog signals were filtered to reduce low-frequency noise using an adjustable high-pass filter and amplifier (Magrec; up to 20 dB gain).
- Signals were then split and sent to independent monitoring systems for the experienced (ePAM) and inexperienced (iPAM) operators.
- For ePAM operators, signals were digitized using a SAIL DAQ sound card (SMRU Instrumentation) sampling at 500 kHz (i.e., high frequency signal) and an RME digital audio-interface (model FireFace UCX), sampling at 192 kHz (i.e., mid frequency signal).
- For iPAM operators, only mid-frequency signals were digitized using an RME sound card (model BabyFace) sampling at 192 kHz; iPAM operators did not monitor the high frequency signal.
- PAMGuard software was used to record, visualize, and analyze digitized signals in real-time (www.pamguard.org). In addition to bioacoustic data collection, PAMGuard also simultaneously logged GPS data, and provided forms for operators to enter

information, such as on/off effort, filter and amplifier settings, configurations used, notes about encounters, and other ancillary data.

B.2 Sonobuoys

- Directional Frequency Analysis and Recording (DIFAR) sonobuoys.
- The sonobuoy system consisted of specialized sonobuoy radio receivers (WinRadio model WR-G39WSBe), and omni-directional (Cushcraft Ringo-Ranger ARX2) vertical pole antennas that were mounted on the ships superstructure, to receive the VHF radio signals transmitted from deployed sonobuoys.
- Signals were digitized using an RME sound card (model BabyFace), sampling at 48 kHz.
- Acoustic data from the sonobuoys were recorded using the PAMGuard DIFAR module, which was operated on a separate, dedicated laptop.

APPENDIX C – RECOMMENDATIONS

RECOMMENDATIONS FOR THE MONITORING AND MITIGATION OF SEISMIC SURVEY SOUND BASED ON THE FINDINGS OF A STUDY COMPARING THREE MARINE MAMMAL DETECTION METHODS

Submitted by:

LGL LIMITED
environmental research associates

Submitted to:

Environmental Studies Research Fund

31 March 2019

LGL Reference: FA0049-03

Recommendations for the Monitoring and Mitigation of Seismic Survey Sound Based on the Findings of a Study Comparing Three Marine Mammal Detection Methods

Prepared by

Heather R. Smith and Valerie D. Moulton

LGL Ltd., environmental research associates
388 Kenmount Road, PO Box 13248, Stn. A
St. John's, Newfoundland and Labrador A1B 4A5

Prepared for

Environmental Studies Research Fund
444 7th Avenue S.W.
Calgary, Alberta, T2P 0X8

31 March 2019
LGL Reference FA0049-03

Suggested format for citation:

Smith, H.R. and V.D. Moulton. 2019. Recommendations for the monitoring and mitigation of seismic survey sound based on the findings of a study comparing three marine mammal detection methods. Environmental Studies Research Funds Report No. 219, Appendix C. LGL Rep. FA0049-03. Rep. by LGL Limited, St. John's, NL, for Environmental Studies Research Fund, Calgary, AB. 18p.

TABLE OF CONTENTS

Page

LIST OF SYMBOLS, ABBREVIATIONS AND ACRONYMS

2-D	two-dimensional
APS	acoustic processing system
DAQ	data acquisition (device)
dB	decibel
DFO	Fisheries and Oceans Canada
DIFAR	directional frequency analysis and recording (sonobuoy)
EA	environmental assessment
ePAM	experienced passive acoustic monitoring (operator)
ESRF	Environmental Studies Research Fund
FIRST	fast infrared search and track (reconnaissance sensor)
FOV	field of view
GPS	global positioning system
Hz	hertz
K	kelvin
IHA	incidental harassment authorization
iPAM	inexperienced passive acoustic monitoring (operator)
IR	infrared
JNCC	Joint Nature Conservation Committee (UK)
kHz	kilohertz
km	kilometre
LGL	LGL Limited, environmental research associates
m	metre
MMO	marine mammal observer
n	sample size
NL	Newfoundland and Labrador
NMFS	National Marine Fisheries Service (US)
NOAA	National Oceanic and Atmospheric Administration
PAM	passive acoustic monitoring
PSO	protected species observer
RDE	Rheinmetall Defence Electronics GmbH
RME	RME Intelligent Audio Solutions
SAIL	St. Andrews Instrumentation
SARA	*Species at Risk Act*
SMRU	Sea Mammal Research Unit
Statement	*Statement of Canadian Practice with respect to the Mitigation of Seismic Sound in the Marine Environment*
THA	towed hydrophone array
UHF	ultra high frequency
UK	United Kingdom
UKCS	United Kingdom Continental Shelf
US	United States
x	times (multiples)
VHF	very high frequency

1.0 INTRODUCTION

The *Statement of Canadian Practice with respect to the Mitigation of Seismic Sound in the Marine Environment* (the "Statement"; DFO 2008) outlines the minimum requirements for planning and conducting seismic surveys that use air source (i.e., airgun) arrays in ice-free, marine waters in Canada. The purpose of the Statement is to mitigate the potential impacts of seismic sound on marine fauna. This report recommends changes to some components of the Statement based on the results of a recently completed study in which we compared marine mammal detections made using visual, acoustic and thermal (infrared, IR) imaging methods; and made by observers with different levels of experience; offshore of Atlantic Canada (Smith et al. submitted). Relevant information from other sources is referenced in support of recommendations.

2.0 BACKGROUND

During a marine seismic survey, acoustic energy is generated by an air source array towed behind a survey vessel and the data are used to map the geology underlying the seafloor. Sounds are produced at frequencies that overlap with those used by many baleen and toothed whales (e.g., Richardson et al. 1995; Menze et al. 2017). Impulsive sounds generated during seismic surveys have been documented to elicit behavioural responses in marine mammals and could result in hearing impairment or injury (Richardson et al. 1995; Gordon et al. 2003; Romano et al. 2004; Southall et al. 2007, 2019; Nowacek et al. 2015; Stone 2015a; Blackwell et al. 2015; Erbe et al. 2018).

Growing awareness about anthropogenic sound in the marine environment, and in particular, concern over the impacts of seismic sound on marine fauna, prompted consideration of how this issue should be regulated in Canada. In 2004, Fisheries and Oceans Canada (DFO) led a review and synthesis of available evidence of the physical, physiological and behavioural impacts of seismic sound on marine life. This review process involved federal and provincial government advisors, as well as national and international scientific experts, and culminated in the production of a peer-reviewed report summarizing impacts on invertebrates, fish, sea turtles and marine mammals (DFO 2004).

Mitigation guidelines used world-wide were also reviewed, and those actions deemed most effective and appropriate for use in Canadian waters were compiled by the federal and provincial governments to produce the Statement (DFO 2008). Mitigation measures are generally designed to minimize the chance of marine mammals incurring hearing impairment and are often triggered when marine mammals are detected entering or are about to enter a safety zone around the sound source (Weir and Dolman 2007; Compton et al. 2008). Detections are most commonly made by visual observation (naked eye or with binoculars). Marine mammal observers (MMOs) scan the ocean's surface for sighting cues including respirations (i.e., "blows") and dorsal fins. MMO effectiveness is reduced by darkness, high sea states, and during low visibility conditions such as fog (Todd et al. 2015; Verfuss et al. 2018). The ability of an MMO to make detections is also reduced when MMOs lack monitoring experience, are fatigued, or are looking in another direction when a marine mammal cue is produced. Passive acoustic monitoring (PAM) and thermal (IR) imaging systems have been identified as being suitable methods to supplement visual observation (Verfuss et al. 2018), and some mitigation guidelines encourage or require the use of PAM to complement visual monitoring for marine mammals during seismic surveys, especially at night (summarized in Compton et al. 2008).

The Statement requires that monitoring for marine mammals be done by qualified MMOs stationed on seismic vessels; and that visual monitoring be done in preparation for, and during times when, the air source array is active, and the safety zone is visible. The Statement also requires that PAM technology be used prior to starting up the air source array when the safety zone is not visible and when the survey is being carried out in an area identified as critical habitat (for endangered or threatened cetaceans on Schedule 1 of the *Species at Risk Act*, SARA) or where vocalizing cetaceans are expected to be encountered if that species has been identified during the environmental assessment (EA) process as a species for which there could be significant adverse effects. Furthermore, the Statement allows that alternate technologies (e.g., IR imaging) may be proposed if they can provide equally effective levels of protection (DFO 2008). In recognition of the continued growth in relevant scientific knowledge and improvements in mitigation technology and practices, the Statement undergoes regular review, during which modification and amendments to the Statement are considered (DFO 2019). However, the Statement has not been updated since it was first published in 2008.

3.0 NEW INFORMATION CONSIDERED

This report considers results from our recently completed study, funded by the Environmental Studies Research Fund (ESRF), that compared marine mammal detections made using visual, acoustic and thermal (IR) imaging methods; and made by observers with different levels of experience (Smith et al. submitted). In short, during our ESRF-funded study, data were collected offshore of Nova Scotia and Newfoundland, Canada, during 31 July to 23 August 2017, from aboard a 38 m research vessel. Marine mammal detections were made using three methods concurrently and independently: MMOs made visual observations using the naked eye or binoculars; acoustic detections were primarily made using a towed hydrophone array, though sonobuoys were also sometimes deployed when baleen whales were sighted; and automatic detections were made using an IR imaging system that comprised a rotating IR camera and detector/classifier previously tested in Atlantic Canada (Zitterbart et al. submitted). [See supplemental material for specifications of the PAM and IR imaging systems used in our study.] Visual and acoustic (towed hydrophone array) detections made by MMOs and PAM Operators with different levels of experience were compared. The effectiveness of transmitting automatic IR detections in real time as alerts to an MMO to assist in making visual sightings was also investigated.

The following results and conclusions from Smith et al. (submitted) are relevant in the context of recommending changes be made to some components of the Statement:

1. Experience level seemingly influenced MMO detection performance in a number of ways. The inexperienced MMO (n = 1) was found to effectively monitor less of the viewable area around the vessel, detect fewer marine mammals, and classified fewer marine mammals to species level, than the experienced MMOs (n = 4). The inexperienced MMO was also slightly slower to detect marine mammals and made initial sightings of marine mammals

when they were marginally closer to the vessel, than the experienced MMOs. It was concluded that mitigation measures reliant on an observer detecting marine mammals in a specified safety zone are likely to be less effectively employed when monitoring is conducted by inexperienced MMOs.

2. Limited data showed that the second highest detection rate achieved was for two MMOs on effort together (second only to the automatic detections made by the IR system). Making the comparison between single and multiple MMOs on effort together was not a goal of this study, thus data available to make this comparison were limited.

3. Detections made using the towed hydrophone array were almost exclusively of vocalizing toothed whales. These detections were localized 65 to 70 % of the time, and localizations were made at mean distances > 1.0 km (i.e., beyond the commonly used 500 m safety zone). In contrast, detections made using sonobuoys were of vocalizing baleen whales. This was despite the towed hydrophone array being capable of detecting baleen whale vocalizations above 500 Hz.

   As well, experienced and inexperienced PAM Operators monitoring towed hydrophone arrays were found to be comparable in terms of detecting vocalizing marine mammals and classified roughly the same proportion of detections to species level. However, whereas both experienced and inexperienced PAM Operators readily classified vocalizations made by sperm whales (*Physeter macrocephalus*), more 'challenging' species classifications (e.g., Risso's dolphins, *Grampus griseus*) were made only by the experienced PAM Operators. Some differences were apparent in the approach to detection taken by the experienced and inexperienced PAM Operators, highlighting the importance of having operators who are experienced and trained to consistently make detections following the same protocol.

   It was concluded that it is important to match the in-situ capabilities of the PAM hardware to detect and localize vocalizations of the key species of interest. It was also concluded that it is important to ensure that PAM Operators are experienced and familiar with the vocalizations of marine mammals likely to be encountered in the study area, and to ensure that PAM Operators have adequate hands-on experience making detections for mitigation purposes in order to improve the consistency and quality of PAM detections made.

4. The vessel-mounted rotating IR camera system was capable of detecting marine mammals in Atlantic Canada during summer. Automatic IR detections were made during daylight and darkness, but the system did not perform well in the presence of precipitation, including fog. Also, the IR camera system resulted in a very high number of false positives (84.5 %), which were mostly attributable to the detection of seabirds. Automatic IR alerts delivered to an "assisted MMO" did result in that MMO making some marine mammal detections, however, the assisted MMO was found to make fewer detections than an experienced MMO who did not receive automatic IR alerts. It was determined that the IR alerting system may actually have been more of a distraction to the assisted MMO given

the high number of false positives. As such, it was concluded that this system requires further study and refinement so that the delivery of IR alerts to MMOs in real-time can be effectively used as a monitoring tool. As well, consideration must be given to the size of the vessel on which an IR system is deployed, in addition to the gimbal stabilization employed with the camera system, as performance will be influenced by camera height above sea level and gimbal stabilization capacity.

5. The types of marine mammals detected, and the extent to which they were classified to the species level, varied depending on which detection methods were used. MMOs and the thermal (IR) system effectively detected both baleen and toothed whales at the water's surface, though species were only reliably identified by MMOs. PAM methods detected cetaceans that were underwater. The species detected using PAM methods differed depending whether a towed array (which detected toothed whales almost exclusively) or sonobuoys (which detected baleen whales) were deployed. Most PAM detections were not classified to species level (e.g., pilot whale spp., unidentified dolphin, Balaenopterid).

6. Employing more than one marine mammal detection method concurrently resulted in improved marine mammal detection performance overall. Our results suggest that PAM and thermal (IR) methods worked effectively during darkness; PAM and visual methods complemented each other during periods of high sea state and low visibility due to precipitation (including fog); and that thermal (IR) methods can be used to enhance visual methods during periods of good visibility. Employing PAM and visual methods during periods of good visibility also resulted in more detections than if only visual methods were used.

4.0 RECOMMENDATIONS AND ADDITIONAL RELEVANT INFORMATION

1. Recommend that criteria be developed, and referenced in the Statement, to define what is required for an MMO to be considered "qualified". The Statement makes reference to a "qualified MMO" but provides no guidance on what this means. Given that the results of our study suggest that mitigation measures will be less effectively implemented when monitoring is conducted by MMOs that are inexperienced, we recommend that "qualified" be defined in terms of both experience and training.

   This recommendation is also supported by the following relevant information.

   - Sighting data from 1,121 seismic surveys within the United Kingdom Continental Shelf (UKCS) between 1995 and 2010 were analyzed, and it was found that dedicated MMOs with marine mammal experience prior to becoming an MMO had better detection skills than those without prior experience (Stone 2015b). Sighting rates for experienced MMOs were 3 x higher in all weather, and 2.5 x higher in good weather, compared to inexperienced MMOs; and experienced MMOs were found to detect animals at greater distances than inexperienced MMOs (approximately 1.5 km vs. 1 km, in all or good weather).

- Mitigation guidelines developed by the United Kingdom (UK) require that all MMOs be formally trained on an approved course, and have some experience spotting marine mammals (JNCC 2017). Experienced MMOs, defined as having a minimum of 20 weeks' experience implementing JNCC guidelines in UK waters, and experienced at identifying UK marine mammal species, are essential only in areas considered of importance for marine mammals; inexperienced MMOs may be used elsewhere (JNCC 2017). Because JNCC mitigation actions apply to all cetaceans (JNCC 2017), the ability of MMOs to identify marine mammals to the species level is not critical, though it may limit the utility of MMO collected data for other purposes. JNCC guidelines are used as the basis of, or have been adopted as the standard, by other countries (Parsons et al. 2009).
- The incidental harassment authorizations (IHAs) recently issued for geophysical surveys in United States (US) mid-Atlantic waters require that marine mammal monitoring and mitigation measures be conducted by trained and qualified protected species observers (PSOs), and that resumes for individual PSOs be submitted to the National Marine Fisheries Service (NMFS) for review and approval (NOAA 2018). Though NMFS does not approve specific training programs, they require that all PSOs be trained biologists, recommend topics to be covered during training courses, and specify minimum qualifications for individuals. Additionally, each source vessel taking part in these surveys is required to carry a minimum of one experienced visual PSO, defined as having a minimum of 90 days at-sea experience during seismic surveys (NOAA 2018).

2. Recommend considering the requirement that two MMOs be on duty together – where one of the MMOs must be an experienced MMO. This would simultaneously increase the probability of detecting marine mammals and provide an opportunity for less experienced MMOs to gain experience during seismic surveys under the mentorship of a more experienced MMO. This recommendation is supported by the results of our study which suggest higher detection rates are achieved when two MMOs are on effort together.

   This recommendation is also supported by the following relevant information.

   - Sighting rates of bowhead whales (*Balaena mysticetus*) were significantly higher when more observers were on watch during marine mammal monitoring conducted during 2-D seismic surveys in the Canadian Beaufort Sea, 2006–2008 (Holst et al. 2018).
   - It is known from previous work that detection rates of seals were higher with two MMOs on duty than with one MMO (Moulton and Lawson, 2002).
   - In recognition that the probability of detection is higher when monitoring is done by more than one visual observer, dedicated cetacean surveys typically use teams of two or more observers (e.g., Hammond et al. 2017; Yano et al. 2018).
   - The IHAs recently issued for geophysical surveys in US mid-Atlantic waters require that a minimum of two PSOs be on duty conducting visual observations during survey operations in all daylight hours (NOAA 2018).

3. Recommend considering that PAM be required in addition to visual monitoring by MMOs during the pre ramp-up watch, as well as during all times when the air source array is active, in order to maximize the potential for detecting cetaceans (including, but not restricted to those listed as endangered or threatened on Schedule 1 of the SARA). The results of our study showed that using PAM in combination with visual methods resulted in improved marine mammal detection performance overall. The results of our study also showed that the towed hydrophone array successfully detected many species of toothed whales and would thus be useful for enhancing detection performance during the pre-ramp-up watch when the detection of marine mammals within the safety zone, regardless of species, would lead to a ramp-up delay.

   This recommendation is also supported by the following relevant information.

   - DFO's (2015) recent review of monitoring and mitigation measures for seismic surveys near cetacean habitat recognized that a combination of detection methods would likely be required to achieve a target detection probability.
   - In their recent review of methods suitable for monitoring marine mammals in low visibility conditions during seismic surveys, Verfuss et al. (2018) concluded that combinations of two or more detection methods may be necessary to provide required detection levels.
   - The IHAs recently issued for geophysical surveys in US mid-Atlantic waters require the use of a towed PAM system, in addition to MMOs, immediately prior to (i.e., prior to ramp-up), and at all times when, airguns are active (NOAA 2018).

4. Recommend that when required, in-situ capabilities of the PAM hardware must be matched to the extent possible to the key species of interest. As noted previously, the Statement requires that PAM technology be used prior to starting up the air source array when the safety zone is not visible <u>and</u> when the survey is being conducted in an area identified as critical habitat (for endangered or threatened cetaceans on Schedule 1 of the SARA) <u>or</u> where vocalizing cetaceans are expected to be encountered if that species has been identified during the EA process as a species for which there could be significant adverse effects. However, the Statement provides no guidance on PAM system requirements and capabilities. The results of our study suggest that "conventional" towed hydrophone arrays have limited ability to detect vocalizing baleen whales, though they can be used to successfully detect many species of toothed whales.

   This recommendation is also supported by the following relevant information.

   - Acoustic detection data from 76 surveys that used PAM during seismic surveys (towed hydrophone array; primarily used in deep water areas) within the UKCS between 1995 and 2010 were analyzed, and during this time, only two baleen whale detections were made (one confirmed and one probable), despite numerous visual sightings of baleen whales (Stone 2015b).

- During a recent seismic program on the Scotian Shelf during May through September, a single detection of a baleen whale (i.e., a humpback, *Megaptera novaeangliae*) was made using a towed hydrophone array. Though unspecified, this detection was likely made using an ultralow frequency towed hydrophone array that was designed specifically to detect blue and fin whales (RPS Energy Canada 2014).
- Mitigation guidelines developed by the UK acknowledge that current PAM systems are not suitable for detecting baleen whales (JNCC 2017). However, JNCC guidelines do include a list of expected characteristics of PAM systems used, including being able to detect the range of frequencies of marine mammal vocalisations expected to be present in the survey area.
- The IHAs recently issued for geophysical surveys in US mid-Atlantic waters require the use of towed PAM systems, and that seismic operators provide PAM plans that demonstrate that the hardware and software planned for use are appropriately sensitive for the operation (NOAA 2018). NMFS also acknowledges that current PAM technology has some limitations, including the masking of signals by noise from the vessel, sound source, or flow.

5. Recommend that criteria be developed, and referenced in the Statement, to define what training and qualifications will be required for PAM Operators. The Statement indicates that PAM can be used to identify cetaceans to species level, and to determine the location of the vocalizing cetaceans relative to the safety zone. However, the Statement provides no guidance on training, qualifications, or previous experience required of personnel operating PAM systems. The results of our study indicate that experienced PAM Operators were able to make more “challenging” species-level classifications than inexperienced PAM Operators. Our results also highlight differences in the approach to localizing vocalizing marine mammals taken by PAM Operators with different levels of training and experience, and the need to ensure that PAM detections are made in a predictable and consistent manner in order to reliably identify and localize marine mammals.

   This recommendation is also supported by the following relevant information.

   - Mitigation guidelines developed by the UK do not require specific training courses of PAM Operators, though at a minimum, they are expected to be capable of assembling and deploying PAM equipment, configuring associated software, identifying acoustic signals and interpreting bearing and range information (JNCC 2017).
   - The IHAs recently issued for geophysical surveys in US mid-Atlantic waters require that all acoustic PSOs (i.e., PAM Operators) be trained and qualified, and that resumes for individual PSOs be submitted to the NMFS for review and approval (NOAA 2018). Though NMFS does not approve specific training programs, they require that all PSOs be trained biologists, recommend topics to be covered during training courses, and specify minimum qualifications for

individuals. Additionally, each source vessel taking part in these surveys is required to carry a minimum of two experienced acoustic PSOs, defined as having a minimum of 90 days experience working in that role (NOAA 2018).

6. Recommend that the rotating IR system we tested NOT be required at this time. The results of our study showed that an IR system has the potential to enhance visual observations during daylight hours, and was capable of detecting both toothed and baleen whales. However, the rotating IR system tested needs more research and development in order to deal with the large number of false positive detections made, and to better develop a system by which IR alerts are delivered to MMOs in real time in order to increase monitoring effectiveness. The results of our study also showed that the IR system did not result in more effective detection during periods of low visibility due to precipitation, including fog.

   This recommendation is also supported by the following relevant information.

   - During the recent commenting period on applications for incidental harassment authorizations (IHA) for geophysical surveys in the Atlantic Ocean, it was suggested that NMFS consider requiring the use of thermal detection systems to supplement visual observations (NOAA 2018). NMFS responded that this would be impracticable given that such systems currently have known limitations with respect to performance during a range of environmental conditions and species detected and acknowledged that commercial applications did not exist.

5.0 REFERENCES

Blackwell, S.B., C.S. Nations, T.L. McDonald, A.M. Thode, D. Mathias, K.H. Kim, C.H. Greene, and A.M. Macrander. 2015. Effects of Airgun Sounds on Bowhead Whale Calling Rates: Evidence for Two Behavioral Thresholds. PLoS ONE 10(6): e0125720. https://doi.org/10.1371/journal.pone.0125720.

Compton, R., L. Goodwin, R. Handy, and V. Abbott. 2008. A critical examination of worldwide guidelines for minimising the disturbance to marine mammals during seismic surveys. Marine Policy. 32, 255–262.

DFO (Fisheries and Oceans Canada). 2004. Review of scientific information on impacts of seismic sound on fish, invertebrates, marine turtles and marine mammals. DFO Can. Sci. Advis. Sec. Habitat Status Report 2004/002.

DFO. 2008. Statement of Canadian practice with respect to the mitigation of seismic sound in the marine environment. Fisheries and Oceans Canada. 5 p.

DFO. 2015. Proceedings of the National Peer Review of Mitigation and Monitoring Measures for Seismic Survey Activities in and near the Habitat of Cetacean Species at Risk. DFO Can. Sci. Advis. Sec. Proceed. Ser. 2015/033.

DFO. 2019. Statement of Canadian practice with respect to the mitigation of seismic sound in the marine environment. Fisheries and Oceans Canada. Retrieved from http://www.dfo-mpo.gc.ca/oceans/publications/seismic-sismique/index-eng.html on 11 March 2019.

Erbe, C., R. Dunlop, and S. Dolman. 2018. Effects of noise on marine mammals. p. 277–309 In H. Slabbekoorn, R.J. Dooling, A.N. Popper and R.R. Fay (eds.), Effects of Anthropogenic Noise on Animals, Springer Handbook of Auditory Research 66. Springer Science+Business Media, LLC.

Gordon, J., D. Gillespie, J. Potter, A. Frantzis, M.P. Simmonds, R. Swift, and D. Thompson. 2003. A review of the effects of seismic surveys on marine mammals. Mar. Technol. Soc. J. 37, 16–34.

Hammond, P.S., C. Lacey, A. Gilles, S. Viquerat, P. Börjesson, H. Herr, K. Macleod, V. Ridoux, M.B. Santos, M. Scheidat, J. Teilmann, J. Vingada, and N. Øien. 2017. Estimates of cetacean abundance in European Atlantic waters in summer 2016 from the SCANS-III aerial and shipboard surveys. Available at https://synergy.st-andrews.ac.uk/scans3/files/2017/04/SCANS-III-design-based-estimates-2017-04-28-final.pdf.

Holst, M., V.D. Moulton, T. Gerwing, and R.E. Harris. 2018. Review of Marine Mammal Monitoring and Mitigation Data and Outcomes for 2-D Seismic Surveys in the Canadian Beaufort Sea, 2006–2012. LGL Report FA0150. Report by LGL Limited, St. John's, NL for Fisheries and Oceans Canada, Winnipeg, MB. 31 p.

JNCC (Joint Nature Conservation Committee). 2017. JNCC guidelines for minimising the risk of injury to marine mammals from geophysical surveys. http://jncc.defra.gov.uk/pdf/jncc_guidelines_seismicsurvey_aug2017.pdf.

Menze S., D.P. Zitterbart, I. van Opzeeland, and O. Boebel. 2017. The influence of sea ice, wind speed and marine mammals on Southern Ocean ambient sound. R. Soc. open sci. 4: 160370. http://dx.doi.org/10.1098/rsos.160370.

Moulton, V.D., and J.W. Lawson. 2002. Seals, 2001. p. 3-1 to 3-48 In: W.J. Richardson and J.W. Lawson (eds.), Marine mammal monitoring of WesternGeco's open-water seismic program in the Alaskan Beaufort Sea, 2001. LGL Rep. TA2564-4. Rep. from LGL Ltd., King City, Ont., for WesternGeco LLC, Anchorage, AK, BP Exploration (Alaska) Inc., Anchorage, AK, and Nat. Mar. Fish. Serv., Anchorage, AK, and Silver Spring, MD. 95 p. Available from the AFSC/MML Library, Seattle.

NOAA (National Oceanic & Atmospheric Administration). 2018. Takes of marine mammals incidental to specified activities; taking marine mammals incidental to geophysical surveys in the Atlantic Ocean. Federal Register / Vol. 83, No. 235 / Friday, December 7, 2018. https://www.federalregister.gov/documents/2018/12/07/2018-26460/takes-of-marine-mammals-incidental-to-specified-activities-taking-marine-mammals-incidental-to

Nowacek, D.P., C.W. Clark, D. Mann, P.J. Miller, H.C. Rosenbaum, J.S. Golden, M. Jasny, J. Kraska, and B.L. Southall. 2015. Marine seismic surveys and ocean noise: time for coordinated and prudent planning. Front. Ecol. Environ. 13, 378–386.Richardson, W.J., C.R. Greene, C.I. Malme, and D.H. Thomson. 1995. Marine mammals and noise. San Diego, CA. Academic Press.

Parsons, E.C.M., S.J. Dolman, M. Jasny, N.A. Rose, M.P. Simmonds, and A.J. Wright. 2009. A critique of the UK's JNCC seismic survey guidelines for minimising acoustic disturbance to marine mammals: Best practise? Mar. Pollut. Bull. 58, 643–651.

Richardson, W.J., Greene, C.R., Malme, C.I., Thomson, D.H. 1995. Marine mammals and noise. San Diego, CA. Academic Press. 576 p.

Romano, T.A., M.J. Keogh, C. Kelly, P. Feng, L. Berk, C.E. Schlundt, D.A. Carder, and J.J. Finneran. 2004. Anthropogenic sound and marine mammal health: measures of the nervous and immune systems before and after intense sound exposure. Can. J. Fish. Aquat. Sci. 61(7), 1124–1134.

RPS Energy Canada. 2014. Wildlife Observation Report BP Tangier 3D WATS Seismic Survey Halifax, Nova Scotia 17 May to 14 September 2014. 46 p. + appendices.

Smith, H.R., D.P. Zitterbart, M. Flau, T.F. Norris, E.L. Ferguson, C. Jones, O. Boebel, and V.D. Moulton. *submitted*. A field comparison of marine mammal detections via visual, acoustic and thermal (IR) imaging methods offshore Atlantic Canada.

Southall, B.L., A.E. Bowles, W.R. Ellison, J. Finneran, R. Gentry, C.R. Green, C.R. Kastak, D.R. Ketten, J.H. Miller, P.E. Nachtigall, W.J. Richardson, J.A. Thomas, and P.L. Tyack. 2007. Marine mammal noise exposure criteria. Aquat. Mamm. 33, 411-521, DOI 10.1578/AM.33.4.2007.411.

Southall, B.L., J.F. Finneran, C. Reichmuth, P.E. Nachtigall, D.R. Ketten, A.E. Bowles, W.T. Ellison, D.P. Nowacek, and P.L. Tyack. 2019. Marine mammal noise exposure criteria: updated scientific recommendations for residual hearing effects. Aquat. Mamm. 45, 125–232, DOI 10.1578/AM.45.2.2019.125.

Stone, C.J. 2015a. Marine mammal observations during seismic surveys from 1994-2010. JNCC report, No. 463a.

Stone, C.J. 2015b. Implementation of and considerations for revisions to the JNCC guidelines for seismic surveys. JNCC report, No. 463b.

Todd, V.L.G., I.B. Todd, J.C. Gardiner, and E.C.N. Morrin. 2015. Marine mammal observer and passive acoustic monitoring handbook. Pelagic Publishing, Exeter, UK. 395 p.

Verfuss, U.K., D. Gillespie, J. Gordon, T.A. Marques, B. Miller, R. Plunkett, J.A. Theriault, D.J. Tollit, D.P. Zitterbart, P. Hubert, and L. Thomas. 2018. Comparing methods suitable for monitoring marine mammals in low visibility conditions during seismic surveys. Mar. Pollut. Bull. 126, 1-18.

Weir C.R. and S.J. Dolman. 2007. Comparative review of the regional marine mammal mitigation guidelines implemented during industrial seismic surveys, and guidance towards a worldwide standard. J. Int. Wildl. Law Policy. 10, 1–27.

Yano, K.M., E.M. Oleson, J.L. Keating, L.T. Ballance, M.C. Hill, A.L. Bradford, A.N. Allen, T.W. Joyce, J.E. Moore, and A. Henry. 2018. Cetacean and seabird data collected during the Hawaiian Islands Cetacean and Ecosystem Assessment Survey (HICEAS), July–December 2017. NOAA Tech. Memo. NMFS-PIFSC-72, 110 p. https://doi.org/10.25923/7avn-gw82.

Zitterbart, D.P., H.R. Smith., M. Flau, S. Richter, E. Burkhardt, J. Beland, L. Bennet, A. Cammareri, A. Davis, M. Holst, E. Kniest, C. Lanfredi, H. Michel, M. Noad, K. Owen, and O. Boebel. *submitted*. Scaling the laws of thermal imaging-based whale detection.

SUPPLEMENTAL MATERIAL

SPECIFICATIONS OF THE PAM AND IR IMAGING SYSTEMS[5]

Passive acoustic monitoring (PAM) system

Towed hydrophone array

The towed hydrophone array, THA, consisted of four hydrophone elements, pre-amplifier circuit boards, and pressure sensors, housed in a mineral oil-filled flexible tube. Two high-frequency hydrophones (High Tech, Inc. HTI-99-UHF elements; effective frequency response from 2 Hz to 250 kHz), were spaced 0.5 m apart, and positioned mid-way between a pair of mid-frequency hydrophones (APC International 42-1021 elements; effective frequency response from ~1 to 100 kHz) that were spaced 3 m apart. Custom designed hydrophone pre-amplifiers and pressure sensors (Kellar 7SE) were integrated into the array. Pre-amplifiers for the high frequency hydrophones were designed and integrated by the supplier (HTI) with 38 dB of gain and a high-pass corner frequency of 500 Hz. Preamplifiers for the mid-frequency hydrophones had 40 dB of gain, a 250 Hz high pass filter, and a 35 kHz low pass filter. The tow-cable was weighted with lead rope that was attached a few meters aft of the array to sink it below the water's surface. The array was deployed 300 m behind the stern of the vessel using a large block and a hydraulic winch and drum system.

The pre-amplified analog signals from the hydrophones were passed from the array up a copper-wire cable into an acoustic processing system (APS) for signal conditioning, digitization and subsequent monitoring and recording (Rankin and Barlow, 2011). The APS was powered by a 12 V DC battery bank, that was independent of the vessel's power system, in order to avoid electrical noise which is typically introduced by the vessel's power. Additional measures taken to minimize noise in the acoustic signal included electronic shielding of all acoustic analog cables from the array to the APS and avoiding the use of high-power electronics and radio-wave emitting (e.g., VHF radio) devices in the acoustics lab on the vessel.

In the APS, analog signals were filtered to reduce low-frequency noise using an adjustable high-pass filter and amplifier (Magrec; up to 20 dB gain). Signals were then split and sent to independent monitoring systems for the experienced (ePAM) and inexperienced (iPAM) Operators. For ePAM Operators, signals were digitized using a SAIL DAQ sound card (SMRU Instrumentation) sampling at 500 kHz (i.e., high frequency signal) and an RME digital audio-interface (model FireFace UCX), sampling at 192 kHz (i.e., mid frequency signal). For iPAM Operators, only mid-frequency signals were digitized using an RME sound card (model BabyFace) sampling at 192 kHz; iPAM operators did not monitor the high frequency signal.

PAMGuard software was used to record, visualize, and analyze digitized signals in real-time (Gillespie et al. 2008; www.pamguard.org). In addition to bioacoustic data collection, *PAMGuard* also

[5] From Smith et al. (submitted).

simultaneously logged GPS data, and provided forms for operators to enter information, such as on/off effort, filter and amplifier settings, configurations used, notes about encounters, and other ancillary data.

Sonobuoys

In order to detect, monitor and localize calls of large baleen whales that went undetected on the THA due to vessel and flow noise, additional passive acoustic data were collected using a DIFAR sonobuoy system. The sonobuoy system was used to monitor low frequencies from ca. 20 Hz to 500 Hz, which would include calls from blue and fin whale, as well as other species such as North Atlantic right (*Eubalaena glacialis*) and sei whales (*Balaenoptera borealis*) and low-frequency components of humpback whale calls and songs. Sonobuoys were usually deployed when large baleen whales were sighted and when ePAM operators were available to assist and monitor. When possible, sonobuoys were deployed in pairs to allow detected calls to be localized.

The sonobuoy system consisted of specialized sonobuoy radio receivers (WinRadio model WR-G39WSBe), and omni-directional (Cushcraft Ringo-Ranger ARX2) vertical pole antennas that were mounted on the ships superstructure, to receive the VHF radio signals transmitted from deployed sonobuoys. Signals were digitized using an RME sound card (model BabyFace), sampling at 48 kHz. Acoustic data from the sonobuoys were recorded using the *PAMGuard* DIFAR module (Miller et al. 2016), which was operated on a separate, dedicated laptop.

Thermal (IR) imaging system

The rotating IR camera (FIRSTnavy sensor, Rheinmetall Defence Electronics GmbH, RDE) and actively stabilized gimbal were installed on the vessel by RDE. The gimbal compensated for up to 12° roll and 10° pitch. The IR camera had a viewing height of 7.8 m above the vessel's waterline, and a 234.5° field of view (FOV; from 239.1°, past 0°, to 123.6°, where the 0° reference is looking forward; Figure 3). The IR camera FOV was slightly obstructed by vessel superstructure ahead and to the sides of the vessel, similar to that of the MMOs. The FIRSTnavy sensor is cooled to 84 K using a Sterling cooler; was operated continuously while the survey was underway; and scanned 360° horizontal x 18° vertical at 5 revolutions per second, providing a 5-Hz video stream of the radiance field of view at horizontal and vertical resolutions of 0.05°/pixel and 0.03°/pixel, respectively. Data were transmitted to a workstation in a lab located aft of the bridge.

Data acquisition and processing were performed with custom developed software (*Tashtego*). *Tashtego* utilized a multi-step detection and classification approach that was developed with data collected during previous expeditions (Zitterbart et al. submitted.). *Tashtego* made automatic detections by tracking contrast in radiance in the IR video stream and applying a set of heuristic rules designed to reduce the number of non-marine mammals detected. Nevertheless, automatic detections include both true and false positives. An engineer oversaw the functioning of the IR system, and adjusted the detection threshold when the false positive rate increased such that the frequency of alerts was a hindrance to MMOs. IR data were acquired and recorded 24 hours per day for the duration of the survey. The IR system simultaneously logged GPS data and heading.

References

Gillespie, D., D. Mellinger, J. Gordon, D. McLaren, P. Redmond, R. McHugh, P. Trinder, X. Deng, and A. Thode. 2008. PAMGUARD: semiautomated, open source software for real-time acoustic detection and localisation of cetaceans. J. Acoust. Soc. Am. 30, 54–62.

Miller, B.S., S. Calderan, D. Gillespie, G. Weatherup, R. Leaper, K. Collins, and M.C. Double. 2016. Software for real-time localization of baleen whale calls using directional sonobuoys: A case study on Antarctic blue whales. J. Acoust. Soc. Am. 139, EL83–EL89.

Rankin, S. and J. Barlow. 2011. Acoustic recording system: a portable hardware system for shipboard passive acoustic monitoring of cetaceans using a towed hydrophone array. SWFSC Administrative Report, LJ-11-01.

Smith, H.R., D.P. Zitterbart, M. Flau, T.F. Norris, E.L. Ferguson, C. Jones, O. Boebel, and V.D. Moulton. *submitted*. A field comparison of marine mammal detections via visual, acoustic and thermal (IR) imaging methods offshore Atlantic Canada.

Zitterbart, D.P., H.R. Smith., M. Flau, S. Richter, E. Burkhardt, J. Beland, L. Bennet, A. Cammareri, A. Davis, M. Holst, E. Kniest, C. Lanfredi, H. Michel, M. Noad, K. Owen, and O. Boebel. *submitted*. Scaling the laws of thermal imaging-based whale detection.