canadian acoustics
acoustique canadienne

Journal of the Canadian Acoustical Association - Journal de l'Association Canadienne d'Acoustique

JUNE 2004	JUIN 2004
Volume 32 -- Number 2	Volume 32 -- Numéro 2

canadian acoustics

THE CANADIAN ACOUSTICAL ASSOCIATION
P.O. BOX 1351, STATION "F"
TORONTO, ONTARIO M4Y 2V9

CANADIAN ACOUSTICS publishes refereed articles and news items on all aspects of acoustics and vibration. Articles reporting new research or applications, as well as review or tutorial papers and shorter technical notes are welcomed, in English or in French. Submissions should be sent directly to the Editor-in-Chief. Complete instructions to authors concerning the required camera-ready copy are presented at the end of this issue.

CANADIAN ACOUSTICS is published four times a year - in March, June, September and December. The deadline for submission of material is the first day of the month preceeding the issue month. Copyright on articles is held by the author(s), who should be contacted regarding reproduction. Annual subscription: $20 (student); $60 (individual, institution); $250 (sustaining - see back cover). Back issues (when available) may be obtained from the CAA Secretary - price $10 including postage. Advertisement prices: $500 (centre spread); $250 (full page); $150 (half page); $100 (quarter page). Contact the Associate Editor (advertising) to place advertisements. Canadian Publication Mail Product Sales Agreement No. 0557188.

acoustique canadienne

L'ASSOCIATION CANADIENNE D'ACOUSTIQUE
C.P. 1351, SUCCURSALE "F"
TORONTO, ONTARIO M4Y 2V9

ACOUSTIQUE CANADIENNE publie des articles arbitrés et des informations sur tous les domaines de l'acoustique et des vibrations. On invite les auteurs à soumettre des manuscrits, rédigés en français ou en anglais, concernant des travaux inédits, des états de question ou des notes techniques. Les soumissions doivent être envoyées au rédacteur en chef. Les instructions pour la présentation des textes sont exposées à la fin de cette publication.

ACOUSTIQUE CANADIENNE est publiée quatre fois par année - en mars, juin, septembre et décembre. La date de tombée pour la soumission de matériel est fixée au premier jour du mois précédant la publication d'un numéro donné. Les droits d'auteur d'un article appartiennent à (aux) auteur(s). Toute demande de reproduction doit leur être acheminée. Abonnement annuel: $20 (étudiant); $60 (individuel, société); $150 (soutien - voir la couverture arrière). D'anciens numéros (non-épuisés) peuvent être obtenus du Secrétaire de l'ACA - prix: $10 (affranchissement inclus). Prix d'annonces publicitaires: $500 (page double); $250 (page pleine); $150 (demi page); $100 (quart de page). Contacter le rédacteur associé (publicité) afin de placer des annonces. Société canadienne des postes - Envois de publications canadiennes - Numéro de convention 0557188.

EDITOR-IN-CHIEF / RÉDACTEUR EN CHEF

Ramani Ramakrishnan
Department of Architectural Science
Ryerson University
350 Victoria Street
Toronto, Ontario M5B 2K3
Tel: (416) 979-5000; Ext: 6508
Fax: (416) 979-5353
E-mail: rramakri@ryerson.ca

EDITOR / RÉDACTEUR

Chantai Laroche
Programme d'audiologie et d'orthophonie
École des sciences de la réadaptation
Université d'Ottawa
451, chemin Smyth, pièce 3062
Ottawa, Ontario K1H 8M5
Tél: (613) 562-5800 # 3066; Fax: (613) 562-5428
E-mail: claroche@uottawa.ca

ASSOCIATE EDITORS / REDACTEURS ASSOCIES

Advertising / Publicité

Karen Fraser
Jade Acoustics Inc.
545 Rivermede Road, Suite 203
Concord, Ontario L4K 4H1
Tel: (905) 660-2444
Fax: (905) 660-4110
E-mail: karen@jadeacoustics.com

News / Informations

Steven Bilawchuk
aci Acoustical Consultants Inc.
Suite 107, 9920-63rd Avenue
Edmonton, Alberta T6E 0G9
Tel: (780) 414-6373
Fax: (780) 414-6376
E-mail: stevenb@aciacoustical.com

Proceedings of the 2003 Workshop on
Detection and Localisation of Marine Mammals
using Passive Acoustics

Les actes de l'Atelier 2003 sur
la détection et la localisation des mammifères
marins à l'aide du repérage acoustique passif

Dartmouth, Nova Scotia
19-21 November 2003

Dartmouth, Nouvelle-Écosse
19-21 novembre 2003

Defence R&D Canada – Atlantic

DALHOUSIE
University

Note: The graphic design for this page and the Cover Page was provided by Imaging Services, DRDC Atlantic

EDITORIAL / EDITORIAL

As the Associate Editor of Canadian Acoustics, I am pleased to present to our readers this special issue. In November 2003, Defense Research & Development Canada (Atlantic), in collaboration with Dalhousie University, hosted the first Workshop on detection and localization of marine mammals using passive acoustics. Francine Desharnais, one of our members, proposed to publish the proceedings of this workshop in Canadian Acoustics. As announced last March, the current issue is dedicated to this initiative. It is our hope that other coordinators of conferences or workshops will follow this example in the future. Francine was invited to write the editorial in order to describe the context in which the workshop was held.

Chantal Laroche, Associate Editor

The concern that acoustic signals can affect marine mammals has increased over the past few years, mainly within the context of sonars and seismic exploration. Whether it is in support of mitigation measures, or in the larger context of marine mammal studies and conservation research, recent years have seen a significant increase in research on marine mammal detection and localization techniques using passive acoustics. As many techniques are now maturing towards automation and implementation in real-time or near real-time systems, the time had come for a workshop on this subject.

The main objective of the workshop was to provide a medium for interested parties to compare their detection and localization algorithms with those of others, identify the advantages and limitations of the various techniques, as well as their relative accuracy and efficiency. For this purpose, two common datasets were made available to the participants, provided by DRDC Atlantic, Dalhousie University, and the Cornell Laboratory of Ornithology.

Using common datasets was a fruitful approach, focusing the presentations and the discussions. We received very positive feedback from the participants, who valued this methodology and the idea exchange that it triggered. The workshop was a resounding success, with over fifty people from eight countries participating, and sharing their state-of-the-art research techniques.

We are pleased to present to you through this special proceedings issue the papers that were submitted as a result of this workshop. They are organized following the workshop presentations: background papers, papers on detection and classification algorithms, and papers on localization. Garry

En tant que Rédactrice Associée d'Acoustique Canadienne, il me fait plaisir de présenter cette édition spéciale à nos lecteurs. En novembre 2003, Recherche et développement pour la défense Canada – Atlantique, en collaboration avec la Dalhousie University, a organisé un premier atelier sur la détection et la localisation de mammifères marins à l'aide de l'acoustique passive. Francine Desharnais, une de nos membres, a proposé de publier les actes de cet atelier dans Acoustique Canadienne. Tel qu'annoncé en mars dernier, cette édition est dédiée à cette initiative. Nous espérons que d'autres coordinateurs de conférences ou d'ateliers suivront cet exemple dans le futur. Francine a été invitée à écrire cet éditorial pour décrire le contexte dans lequel cet atelier a eu lieu.

Chantal Laroche, Rédactrice Associée

La crainte que les signaux acoustiques mettent les mammifères marins en péril s'est intensifiée au cours des dernières années, surtout dans le contexte des sonars ou de l'exploration sismique. Qu'elles cadrent dans les mesures d'atténuation ou dans le contexte plus large des études des mammifères marins ou de leur conservation, les recherches sur les techniques de détection ou de localisation de mammifères marins à l'aide de l'acoustique passive se sont considérablement accrues ces dernières années. Comme un grand nombre de ces techniques évolue vers l'automation, ou l'implantation dans des systèmes en temps réel ou en temps quasi réel, il était opportun d'organiser un forum sur ce sujet.

Le grand objectif de l'atelier consistait à donner aux intéressés un milieu où comparer entre eux leurs algorithmes de détection et de localisation et où distinguer les avantages et limites des différentes techniques, de même que leur précision et leur efficacité relatives. Pour ce faire, un ensemble commun de données a été mis à la disposition des participants. Ces données ont été assemblées par RDDC Atlantique, la Dalhousie University, et le Cornell Laboratory of Ornithology.

L'utilisation d'ensembles de données communs a été une approche fructueuse, et a permis de cibler les présentations et les discussions. Les participants ont réagi positivement à cette méthodologie, et ils ont apprécié l'échange d'idées que ce format a favorisé. L'atelier a connu un grand succès, rassemblant plus de cinquante participants de huit pays différents qui ont pu ainsi partager leurs techniques de recherche modernes.

Heard and Nicole Collison of our technical committee helped in processing the manuscripts and getting them peer-reviewed. We hope you will enjoy reading them.

Francine Desharnais, Guest Editor.

This workshop was co-sponsored by Defence R&D Canada Atlantic and Dalhousie University (Dr. Alex Hay), and endorsed by the Canadian Departments of National Defence and Fisheries and Oceans, World Wildlife Fund, Center for Coastal Studies, Canadian Whale Institute, and New England Aquarium.

Il nous fait plaisir de vous présenter dans cette édition spéciale les articles qui ont été soumis à la suite de l'atelier. Ils sont organisés suivant l'ordre des présentations de l'atelier: articles documentares, articles sur les algorithmes de détection et de classification, et articles sur la localisation. Garry Heard et Nicole Collison, de notre comité technique, ont aidé à traiter les manuscrits et les faire évaluer par des pairs. Nous espérons que vous apprécierez ces articles.

Francine Desharnais, Rédactrice Invitée.

L'atelier a été coparrainé par Recherche et développement pour la défense Canada – Atlantique et la Dalhousie University (Dr Alex Hay), et a eu l'appui des ministères canadiens de la Défense nationale et des Pêches et Océans, du Fonds mondial pour la nature, du Center for Coastal Studies, du Canadian Whale Institute et du New England Aquarium.

$\mathcal{D}$efence R&D Canada (DRDC) and Dalhousie University began working together on the Right Whale Project in July 1999 when they joined forces to investigate the feasibility of acoustically locating whales, with help from 415 Squadron of CF Base Greenwood, Nova Scotia. That trial gave encouraging results, and was the foundation for several follow-on exercises.

These trials provided an excellent opportunity for DRDC scientists to refine techniques to detect, classify and locate sources of sound. The techniques can be applied to a range of problems, with both civilian and military applications. Being able to localize marine mammals in near real-time is an important tool to support mitigation techniques for any kind of acoustic activities underwater.

The workshop helped identify the current and most promising techniques in this field and the robust exchange of ideas that occurred will provide the basis for future collaboration between universities, government research centers, and companies. Working with colleagues from different environments fosters excellent science and everyone benefits.

$\mathcal{R}$echerche et développement pour la défense Canada (RDDC) et la Dalhousie University ont entamé leur collaboration dans le cadre du projet des mysticètes en juillet 1999, quand ils ont joint leurs efforts pour enquêter sur la faisabilité de situer les baleines au moyen de l'acoustique, avec l'aide du 145e Escadron de la base des Forces canadiennes Greenwood, Nouvelle-Écosse. Cet essai a produit des résultats encourageants et a servi d'assise aux exercices subséquents.

L'essai a constitué pour les scientifiques de RDDC une excellente occasion de perfectionner leurs techniques de détection, de classification et de localisation de sources sonores. Ces techniques peuvent s'appliquer à une variété de problèmes et ont des applications tant civiles que militaires. La capacité de localiser les mammifères marins en temps quasi réel constitue un outil important à l'appui des techniques d'atténuation de tous les types d'activités sous-marines.

L'atelier a aidé à identifier les techniques actuelles et les techniques les plus prometteuses du domaine, et l'échange robuste d'idées qui s'est produit sera la base de coopération future entre universités, centres gouvernementaux de recherches et entreprises. Le travail entre collègues de milieux divers favorise l'excellence scientifique et tout le monde en bénéficie.

Dr Ross Graham
Director General, DRDC Atlantic
Director général, RDDC Atlantique

On behalf of the Faculty of Science of Dalhousie University, I extend congratulations to the organizers of the Workshop on the Detection and Localization of Marine Mammals Using Passive Acoustics. I had the opportunity to attend part of the workshop and wished that my schedule would have allowed me to take in all the talks.

As a pure mathematician, trained in abstract harmonic analysis, who has drifted into aspects of signal processing in recent years, I was most appreciative of the mathematical aspects of the presentations. As a dean of science, I was absolutely delighted with the theme of the workshop and the wonderful array of gifted participants who brought diverse skills to the issue of using passive acoustic methods to identify and locate marine mammals.

We, at Dalhousie University, are justifiably proud of our research strengths in the marine sciences and we were proud to contribute to this workshop. However, we are also proud to be part of a larger community of gifted marine scientists in the Halifax region including those at Defence R&D Canada – Atlantic who hosted this workshop jointly with Dalhousie. I want to thank Francine Desharnais, from DRDC, and Alex Hay, from our Department of Oceanography, for their superb organization of this important scientific event.

Keith F. Taylor, PhD
Dean of Science
Dalhousie University

2003 Workshop on Detection and Localization of Marine Mammals Using Passive Acoustics

Les actes de l'Atelier 2003 sur la detection et la localization des mammifères marins à l'aide du repérage acoustique passif

Table of Contents / Table des matières

Localization

Acousti-Mat

The Best Barrier Between Floors and Noise Complaints.

For Upgraded Sound Control
ACOUSTI-MAT® II

Ceramic Tile
1" (25 mm) approved Maxxon® Underlayment
Acousti-Mat®II 1/4" (6.4 mm)
Blown-in insulation
Wood Subfloor
Resilient Channel
5/8" (16 mm) Gypsum Board
Parallel Chord Truss

Wood Flooring
1" (25 mm) approved Maxxon® Underlayment
Acousti-Mat®II 1/4" (6.4 mm)
Cast-in-place concrete 8" (205 mm) thick
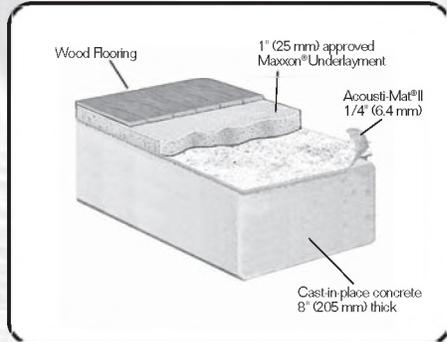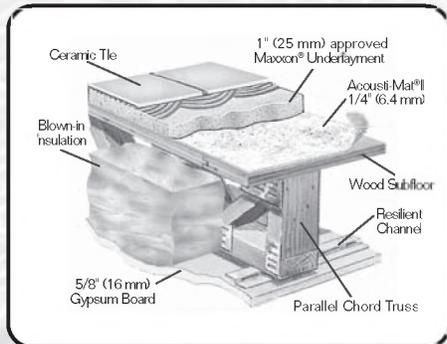
For Super Sound Control
ACOUSTI-MAT® 3

Wood
Ceramic Tile
1½" (38 mm) of approved Maxxon® Underlayment
Wire Mesh
Acousti-Mat® 3
Old Wood Subfloor
Open Beams
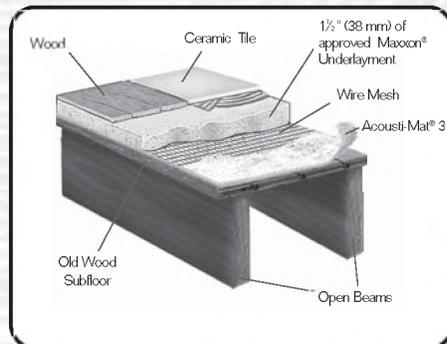
- **Renovation**
- **New Construction**
- **Concrete Construction**
- **Open Beam Construction**
- **Hard Surface Floor Goods Areas**

Acousti-Mat® II is the low-profile system that actually isolates sound waves, reducing transmission of impact and airborne sound up to 50 percent over wood frame, and up to 75 percent over concrete.

Acousti-Mat® 3 is the answer for "impossible" sound control challenges, such as open beam construction. A poured Maxxon® Underlayment covers either system option, increasing IIC and STC up to 17 rating points for wood frame, and up to 25 IIC for concrete construction.

Proven on over 23 million square feet. Specify Acousti-Mat for proven results.

For more information or a FREE GUIDE to acoustical construction, call 800-356-7887 or visit www.maxxon.ca

ACOUSTI-MAT®
Superior Sound Control Systems

OVERVIEW OF THE 2003 WORKSHOP ON DETECTION AND LOCALIZATION OF MARINE MAMMALS USING PASSIVE ACOUSTICS

Francine Desharnais[1], Alex E. Hay[2]

Workshop Organizers

[1]DRDC Atlantic, PO Box 1012, NS, Canada, B2Y 3Z7, francine.desharnais@drdc-rddc.gc.ca

[2]Department of Oceanography, Dalhousie University, Halifax, NS, Canada B3H 4J1

1. INTRODUCTION

The *2003 Workshop on Detection and Localization of Marine Mammals Using Passive Acoustics* was held in Dartmouth, NS, Canada, 19-21 November 2003.

The main objective of this workshop was to provide a forum at which interested parties could compare their detection and localization algorithms with those of others, identify the advantages and limitations of the various techniques, as well as their relative accuracy and efficiency. For this purpose, a common dataset was made available to the participants by DRDC Atlantic and Dalhousie University. After initial distribution, the Cornell Laboratory of Ornithology offered an additional dataset to expand the base for detection algorithms. These datasets are described in detail in these proceedings.

The workshop was divided into four sessions: background presentations, detection and classification, localization, and discussion periods. The background presentations provided examples of passive detection and localization for the purpose of species conservation or mitigation. The participants presented their algorithms during the detection and localization sessions. During the discussion periods, participants compared results obtained from the workshop datasets, different detection and localization technologies, and the possibilities for automation and future collaboration. This short paper recaps the techniques that were presented, as an introduction to the papers that were submitted in these proceedings. It also summarizes the discussions, and some of the highlights from the workshop.

2. BACKGROUND PAPERS

Angela d'Amico (Space and Naval Warfare Systems Command) had been invited to give the keynote address of this workshop but was unable to attend, and her presentation was given by Robert Gisiner (Office of Naval Research). Their presentation [1] reviewed the experience gained by the SACLANT Undersea Research Centre (now NATO URC) on visual and acoustic cetacean surveys in the Ligurian Sea, and discussed the benefits and limitations of both techniques.

Since the two datasets offered to the workshop participants were based on sounds from the endangered North Atlantic right whale (*Eubalaena Glacialis*), Douglas Gillespie (International Fund for Animal Welfare) put the right whale conservation effort into perspective by describing an acoustic detection system that is being developed for the purpose of managing the species. The paper was submitted to these proceedings by Moscrop *et al* [2]. Zimmer *et al.* [3] from the NATO URC discussed how data from various sources, in this case visual, acoustic and tag data, can be merged together to reconstruct sperm whale tracks in three dimensions. Finally, Vagle and Ford [4] discussed a passive acoustic system that is being developed for the purpose of detecting baleen and killer whales on the Canadian west coast.

3. DETECTION & CLASSIFICATION

3.1 Techniques

Nine papers were presented on the topics of detection and classification, and more algorithms were presented during the localization session. All but two algorithms were based on frequency/time analyses; the other two techniques were time based.

The frequency/time techniques generally work with an energy detector that exploits the frequency/time characteristics of the signal. Once a detection is made, the signal is parameterized using specific features. The signal is then classified by decisions based on these features.

The classification algorithms of Gillespie [5] and Mellinger [6] were right-whale specific. Gillespie [5] used an edge detector on the smoothed spectrogram of vocalizations, which are then parameterized using features such as start and stop frequency, signal duration, *etc*. Mellinger [6] compared two algorithms: neural networks and spectrogram correlation with a synthetic kernel. Their results for right whale vocalizations showed that the neural net technique worked best.

The technique of Matthews [7] is broader in classification as it is aimed at frequency-modulated vocalizations, which are broken into sequences of linear chirps, and parameterized by features such as chirp rate, start frequency, *etc*. Other techniques were aimed at odontocete echolocation signals such as Adam *et al.*'s wavelet-based algorithm [8]. Using wavelets is a way to adapt the time-frequency resolution to the signal to be detected. The technique was tested on sperm whale clicks, and is expected to be robust for signals with low signal-to-noise ratios.

Harland and Armstrong [9] presented a suite of

algorithms aimed at the general detection of mysticetes and odontocetes. Their normalized spectrograms are converted to binary spectrograms based on a selected threshold, and signal boundaries are defined using an 8-connectivity neighbourhood algorithm. The signals are parameterized with features such as spectral slope, minimum and maximum frequency, duration, *etc.* depending on the category of sound to be identified: low or high frequency echolocation calls, low or high frequency mysticete tonals, or odontocete tonals. Their algorithms can be tuned for specific species.

The technique of van IJsselmuide and Beerens [10], which is also aimed at general detection, uses normalized lofargrams from a broadband beamformer. A Page's test with a power-law integrator isolates data that are believed to be signal free, and defines the signal's start and finish times. Signal clusters are parameterized with a pattern recognition algorithm based on signal frequency, duration, *etc.*

The two time-based techniques were those of LaCour and Linford [11] and Johannson and White [12]. The LaCour and Linford technique is based on independent component analysis. The hypothesis is that whale sounds are non-Gaussian and statistically independent, and this is used as a detection statistic. Johansson and White's technique is based on parametric modeling, using AutoRegressive-Moving-Average (ARMA) models which are appropriate for narrowband signals in noise. The sample-by-sample processing can be implemented for real-time detection.

3.2 Detection results and algorithms review

Three of the workshop datasets were available to test algorithms with: a 20-min sample from the DRDC/ Dalhousie dataset, and the two extensive datasets from Cornell. The participants were requested to provide the following information:

- relative time of detection, classification of sound;
- basis of classification criteria;
- pros and cons of criteria.

Unfortunately, participants selected different subsets of these datasets, and answered the questions differently. Since the datasets contained mainly right whale sounds, some of the algorithms were tuned specifically for right whales while others were for general detection. Thus, the definition of a "detection" was not consistent amongst participants.

It was recommended that future workshops use a more stringent definition in terms of "probability of detection" and "probability of false alarms", such as a Receiver Operating Characteristic (ROC) curve. Douglas Gillespie (IFAW) suggested that two sets of data be provided to the participants: one with human browser information so that people can tune their detectors, and a second to serve as a blind test set with the "truth" only provided at the workshop.

It was mentioned during the discussion that it is difficult with marine mammals to establish the true number of properly-

identified vocalizations, so that plausible probabilities of detection and false alarms cannot be readily established. Human classifiers are relied upon to calibrate a training set, but there is variability even within human classifiers. The quality of a training set or uncertainty in classification will affect probabilities of detection and false alarms. It was suggested that perhaps more biologists are needed to tell the acousticians what the calls are and how useful they are.

3.3 General comments

The most rugged algorithms are species-specific: the more you know about the species you are trying to detect and about the local environment (including the species which generate false alarms), the better your algorithms can become.

Energy detectors need good signal-to-noise ratios, therefore noise reduction techniques or the use of a noise adaptive threshold, are important. Additional ways to simplify the signals, such as using binary spectra or defining calls with an edge detector, help classification and make the information easier to compress.

Noise removal, whether through adaptive noise removal techniques, equalization filters, *etc.*, is important and needs to be documented. Noise removal may be done before a detection to improve detection rate, or after the detection to reduce impact on the classification (i.e. some signal features could be removed as well with the noise removal techniques). There are potential problems depending on the type of signal (tonal-*vs.*-broadband), and the technique. This topic may be worth a second look at a future workshop.

Time-based techniques have a strong advantage in detecting overlapping signals and dealing with signals of variable duration, but they may need to work jointly with other algorithms to strengthen their classification capabilities.

4. LOCALIZATION

4.1 Techniques

Twelve papers were presented on the topic of localization, and the techniques used fall under the general headings of hyperbolic fixing, optimization, model-based approaches, and bearing triangulation.

Hyperbolic fixing is based on the intersections of constant arrival time difference hyperbolae for the receiver pairs in an array. Simard *et al.* [13], Laurinolli and Hay [14], Munger [15] and Wahlberg [16] all use hyperbolic fixing, but employ different techniques to estimate the time differences. Simard *et al.* use both a filtered waveform cross-correlation and a spectrogram cross-coincidence (overlapping pixels on a binary spectrum). Munger uses cross-correlation with a synthetic kernel. Laurinolli and Hay use spectrogram cross-correlation. Wahlberg uses cross-correlation in the time domain.

Optimization techniques home in on a position by

minimizing the overall error based on pre-defined criteria. Simons et al. [17] use hyperbolic fixing to obtain a first estimate, followed by an iterative process to optimize the solution of the linearized relative travel time equations. Desharnais et al. [18] use an optimization technique based on a downhill simplex algorithm. The full sound speed profile was used for the 2000 workshop dataset, but a constant sound speed was required to resolve the 2002 dataset.

Several talks described model-based approaches to localization. Morrissey et al. [19] used the Marine Mammal Monitoring on Navy Ranges (M3R) toolset for passive detection, localization, and tracking of marine mammals, which has the potential to use shallow water algorithms such as matched-field tracking or shallow-water path-based tracking algorithms. They used a direct path assumption to solve the workshop dataset.

Wiggins et al. [20] use a Pekeris-type normal mode model to determine range from the mode-dependent group velocities. The method provides both source range and depth estimates from a single sensor. Tiemann and Porter [21] use a ray-tracing model (Bellhop) with Gaussian beam-spreading to include indirect paths in the location estimates. Like the hyperbolic technique, locations are determined from pair-wise differences in arrival times among the array elements. Localization estimates are 3D and include a maximum likelihood score. Laplanche et al. [22] localize the depth of sperm whale clicks using sea surface- and bottom-reflected signals detected on a single hydrophone and ray-tracing to construct a virtual line array.

Bearings from DIFAR sensors are used by both Greene et al. [23] and Mcdonald [24] to determine 2D positions. The technique has the advantage of not depending on a constant sound speed approximation, and is not affected by multipath. The same can be said for techniques that use bearings from other types of sensors, such as towed array beamforming. Zimmer et al. described such data in their presentation [3].

4.2 Results

The localization results obtained by participants for using the workshop DRDC/Dalhousie datasets are shown in Fig. 1 (2002) and Fig. 2 (2000). Most authors used a constant sound speed assumption, as listed in Table 1.

The localizations plotted in Figs. 1 and 2 do not consider the errors, or differences due to the detection and localization algorithms. Comparisons should therefore be made carefully. Nevertheless, it is good news that the localizations obtained by the different groups using the 2002 data are mainly within 1.5 km of each other, for the sounds positioned within or near the OBH array, which spans over 14 km. For the two whales positioned approximately 35 km south of the central OBH, the localizations spread over 7 and 4 km, or approximately 12-20% of the range. Though the nearest positions are Laurinolli's and are based on the slowest sound speed, the farthest positions are not those based on the highest sound speed. Since no direct path exists between these two southern locations and the individual OBHs, all results for these two
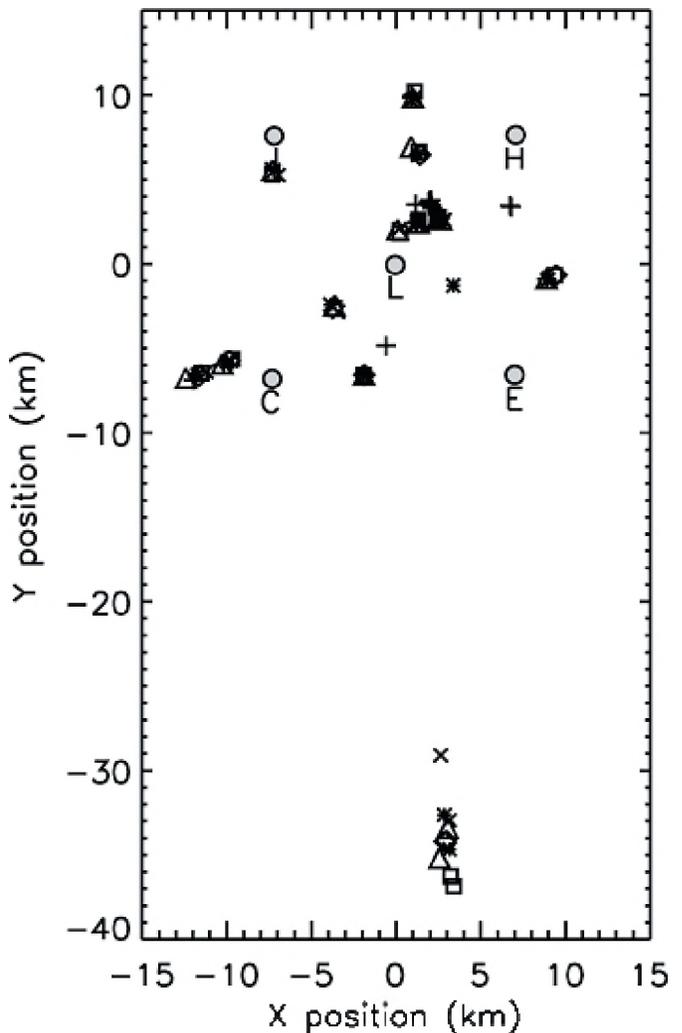
Fig. 1. 2002 results: Laurinolli and Hay (x); Simard et al., spectrum coincidence(); Simard et al., cross-correlation(*); Morrissey et al. (+); Simons et al. (◊); Desharnais et al. (Δ). Shaded circles indicate the OBH positions.

whale positions could be overestimated. Whichever speed is closest to an average group velocity (likely lower than the average sound speed) for these sounds should lead to the most accurate answer for the two farthest sources.

Table 1. Sound speed used by authors

Authors	Sound speed [m/s]	
	Yr2002	Yr2000
Laurinolli and Hay	1485.	–
Simard et al.	1491.	1491.
Morrissey et al.	1500.	1500.
Simons et al.	1492.	–
Desharnais et al.	1499.	full sound speed profile

The year 2000 calibration dataset consisted of right whale playbacks transmitted with a projector lowered into the water
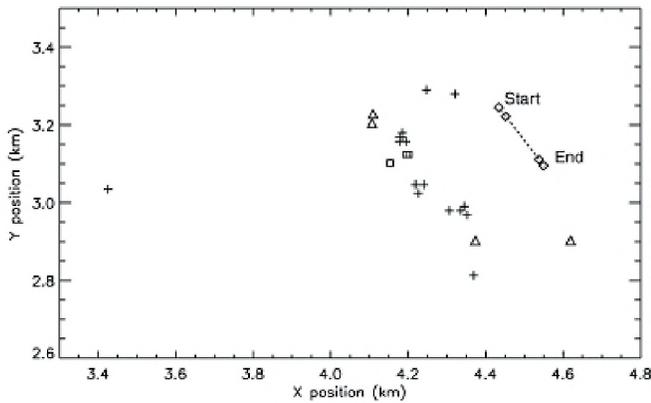
**Fig. 2. 2000 results: Simard et al., spectrum coincidence ();
Morrissey et al., (+); Desharnais et al., (△); location of the
RHIB (◊).**

from a rigid-hull inflatable boat (RHIB). These transmissions were made by Susan Parks (Woods Hole Oceanographic Institute), from a sound file provided by Scott Kraus, of the New England Aquarium. Unfortunately, the playback tapes were not available when the dataset was prepared for the workshop. As a consequence, the vocalizations that were picked from the playback recordings on the OBHs were not confirmed playback sounds. Fig. 2 shows that most authors localized the sounds 250 to 300 m from the known RHIB positions. It is possible that the sounds selected for the calibration dataset were actually right whale vocalizations, as opposed to sounds from the playbacks. However, the OBH positions in this case were spread over 8 km, and the localizations were roughly in the middle of the pattern. A 250 to 300-m error is consistent with the differences observed between authors for the 2002 dataset. Also, the localizations appear to track the RHIB drift. This may indicate that the sounds were truly from the playback recordings, and that the error represents the accuracy of the localization techniques in this environment.

4.3 Errors

Direct comparison of the localization accuracy of the different algorithms and techniques is not attempted here, in part because the different papers use different measures of error, and we have no independently verified locations.

Within the context of the workshop data set, unambiguously-coded signals from a source at known locations within and around the array would have provided independent measures of absolute accuracy and precision, but were not available. For the hyperbolic method, the simplest measure of error is the statistical spread among the hyperbolae intersections (*i.e.*, the precision of the estimate). Simons *et al.* determine 95% confidence ellipses, representing the precision envelope, from the covariance matrix of the time difference equations linearized about a first-guess position. In addition, Simons *et al.* carry out Monte Carlo simulations

to estimate the relative contribution to the location error from uncertainties in sound speed, arrival time difference, and hydrophone position, and conclude for the 2002 Bay of Fundy data set that location error was primarily determined by uncertainty in the relative arrival times. Wahlberg uses the Yr2000 data set to investigate the location error obtained with both linear and non-linear error propagation methods. Input variables are allowed to vary within their specified error range, and the probable location determined within the overlap area of clouds of points for different sub-arrays, with a specified likelihood. Wahlberg concludes that the non-linear method gives more reasonable error estimates.

Among the model-based techniques, Tiemann and Porter construct an ambiguity surface by using ray-tracing methods to allow for uncertainties in arrival time. The source location is the maximum likelihood position on the ambiguity surface, and both the magnitude and shape of the likelihood peak represent measures of the location error. Laplanche *et al.* estimate depth error analytically, by assuming that the input variables (arrival time, water depth, sound speed) are Gaussian-distributed random variables.

Error is also related to array design and to the environment. Localization at ranges greater than the array aperture has obvious value but, as Figure 1 indicates, comes at the price of larger errors. The spread among the points at the lower right in Figure 1 is due mainly to the small angle of intersection between hyperbolae at long range. However, Chapman [25] points out that for a hydrophone mounted near the seabed, the direct and bottom-reflected paths arrive at very similar times but with different phases. Thus, interference between the two paths can occur, and sound following an indirect path may have a larger amplitude depending upon range, water depth, and the sound speed profile. In this case the direct path assumption would be flawed, and would lead to a localization bias. Ray-tracing models (*e.g.*, Tiemann and Porter), which take the sound speed profile and bathymetry into account and include both the direct and indirect paths, provide one approach to extending the localization range of an array. When dispersive effects are apparent in the received sounds, normal-mode models (*e.g.*, Wiggins *et al.*) provide another.

5. DISCUSSION TOPICS

The workshop hosted four discussion periods, and it is not the intention of this paper to summarize all of the points raised. Those relevant to detection, classification and localization are included in the summaries above. The participants also exchanged views on automation and on collaboration opportunities, including sharing of algorithms, data, and equipment.

5.1 Automation

Automating detection and classification is a case-specific compromise between probability of detection and false alarm

rate. This trade-off affects the choice of algorithm, as well as the amount of supporting environmental information that is required as input into the system. The automation process is also different depending on whether localization is required or only a presence/absence decision. Success will depend largely on the species to detect, and the knowledge available for the local environment.

Contextual information (such as source bearing) can be imbedded in a mature system to further improve its performance. This lack of contextual information in an automated system has been identified by one of the participants as the biggest impediment to full automation. Multiple target tracking is also required, but the computational cost may be too high.

Data reduction may be required for transmission from a remote location, or for localization with a sparse array of sensors with limited communication abilities. But how do you characterize your signal detected so that you can transmit limited information for future localization, while preserving enough information to identify the same signal recorded on other sensors? This ability relies heavily on the quality and variety of the training sets used for the development of the automatic system.

5.2 Future collaboration: Data/algorithm/equipment sharing and development.

There are a few repositories of marine mammal sound data, such as the Macauley Library of Natural Sounds at the Cornell Laboratory of Ornithology. This or other web sites could have the ability to provide shareable algorithms also, and this should be encouraged.

The calibration dataset was in our view one of the factors which generated widespread response to the call for participation in the workshop. However, the dataset had to be assembled in a hurry, and did have flaws: *e.g.* the original playback tape was not available. The need for a high quality calibration dataset remains, such as controlled data from an acoustic range.

6. CONCLUSIONS

Advertisement for this workshop was done mainly through word-of-mouth and email forwarding. Yet, it attracted over fifty participants from eight countries. This by itself demonstrates how active this research community is, and how relevant these specialist meetings are.

We touched the tip of the iceberg. Many topics, techniques and algorithms were not discussed during this first meeting, and participants felt that a follow-on workshop would be welcomed. Olivier Adam of LiiA - iSnS, a laboratory of the Université Paris 12 (www.liia-paris12.net) is presently gauging interest for a second Workshop, which would be organized jointly with the Centres d'Études Biologiques de Chizé (CEBC), a laboratory of the Centre National de la Recherche Scientifique (www.cebc.cnrs.fr). This workshop could be hosted in Monaco, October 2005.

Meanwhile, the datasets that were made available for the 2003 Workshop are still available for researchers who want to benchmark their algorithms to those of others. We hope that this first experience will continue to be built on.

REFERENCES

[1] A. D'Amico and R. Gisiner, Experience gained on visual and acoustic cetacean surveys in the Ligurian Sea. Presented at the Workshop, November 2003.

[2] A. Moscrop, J. Matthews, D. Gillespie, and R. Leaper, Development of passive acoustic monitoring systems for Northern right whales, Can. Acoust. Vol. 32, #2, June 2004.

[3] W.M.X. Zimmer, M.P. Johnson, and P.L. Tyack, 3-D reconstruction of sperm whale tracks during foraging dives using visual, acoustic and tag data. Presented at the Workshop, November 2003.

[4] S. Vagle, J.K.B. Ford, N. Erickson, N. Hall-Patch, and G. Kamitakahara, Acoustic recording systems for baleen whales and killer whales on the West Coast of Canada, Can. Acoust. Vol. 32, #2, June 2004.

[5] D. Gillespie, Detection and classification of right whale calls using an 'edge' detector operating on a smoothed spectrogram, Can. Acoust. Vol. 32, #2, June 2004.

[6] D.K. Mellinger, A comparison of methods for detecting right whale calls, Can. Acoust. Vol. 32, #2, June 2004.

[7] J. Matthews, Detection of frequency-modulated calls using a chirp model, Can. Acoust. Vol. 32, #2, June 2004.

[8] O. Adam, M. Lopatka, and J. Motsch, The 'fanning out' shape, in the acoustic signature of sperm whales. Presented at the Workshop, November 2003.

[9] E J Harland and M S Armstrong, The real-time detection of the calls of cetacean species, Can. Acoust. Vol. 32, #2, June 2004.

[10] S.P. van IJsselmuide and S.P. Beerens, Detection and classification of marine mammals using an LFAS system, Can. Acoust. Vol. 32, #2, June 2004.

[11] B.R. La Cour and M.A. Linford, Detection and classification of North Atlantic right whales in the Bay of Fundy using independent component analysis, Can. Acoust. Vol. 32, #2, June 2004.

[12] A.T. Johansson and P.R. White, Detection and characterization of marine mammal calls by parametric modelling, Can. Acoust. Vol. 32, #2, June 2004.

[13] Y. Simard, M. Bahoura and N. Roy, Acoustic detection and localization of whales in Bay of Fundy and St. Lawrence estuary critical habitats, Can. Acoust. Vol. 32, #2, June 2004.

[14] M.H. Laurinolli and A.E. Hay, Localisation of right whale sounds in the workshop Bay of Fundy dataset by spectrogram cross-correlation and hyperbolic fixing, Can. Acoust. Vol. 32, #2, June 2004.

[15] L.M. Munger, North Pacific right whales: Long-term acoustic detection and localization in the Bering Sea. Presented at the Workshop, November 2003.

[16] M. Wahlberg, Comparing a linear with a non-linear method for acoustic localization, Can. Acoust. Vol. 32, #2, June 2004.

[17] D.G. Simons, C. van Moll and M. Snellen, A two-stage method for determining the position and corresponding

precision of marine mammal sounds, Can. Acoust. Vol. 32, #2, June 2004.

[18] F. Desharnais, M. Côté, C.J. Calnan, G.R. Ebbeson, D.J. Thomson, N.E.B. Collison and C.A. Gillard, Right whale localization using a downhill simplex inversion scheme, Can. Acoust. Vol. 32, #2, June 2004.

[19] R. Morrissey, D. Moretti, S.M. Jarvis, N. DiMarzio, and D. Hummels, Marine Mammal Detection & Localization on Widely Spaced, Bottom Mounted Sensors (M3R toolkit), Can. Acoust. Vol. 32, #2, June 2004.

[20] S.M. Wiggins, M.A. McDonald, L.M. Munger, S.E. Moore, and J.A. Hildebrand, Waveguide propagation allows range estimates for North Pacific right whales in the Bering Sea, Can. Acoust. Vol. 32, #2, June 2004.

[21] C.O. Tiemann and M.B. Porter, Automated model-based passive acoustic tracking of marine mammals. Presented at the Workshop, November 2003.

[22] C. Laplanche, O. Adam, and J. Motsch, Accuracy in the localization of sperm whales resident in the Strait of Gibraltar using one hydrophone, Can. Acoust. Vol. 32, #2, June 2004.

[23] C.R. Greene, Jr, M.W. McLennan, R.G. Norman, and W.J. Richardson, Bowhead whale call localization using DIFAR techniques. Presented at the Workshop, November 2003.

[24] M.A. McDonald, DIFAR hydrophone usage in whale research, Can. Acoust. Vol. 32, #2, June 2004.

[25] D.M.F. Chapman, You can't get there from here: shallow water sound propagation and whale localization, Can. Acoust. Vol. 32, #2, June 2004.

Conference Sponsors

Workshop participants, 20 November 2003.

Conference Endorsors

©1986 WWF   ® WWF Registered Trademark

© Copyright1986 du WWF-Fonds Mondial pour la Nature
® Marque déposée par le WWF

CANADIAN WHALE INSTITUTE

Fisheries and Oceans
Canada

Pêches et Océans
Canada

DEVELOPMENT OF PASSIVE ACOUSTIC MONITORING SYSTEMS FOR NORTHERN RIGHT WHALES

MOSCROP, A., MATTHEWS, J., GILLESPIE, D., AND LEAPER, R.

INTERNATIONAL FUND FOR ANIMAL WELFARE, 87-90 ALBERT EMBANKMENT, LONDON, SE1 7UD, UK

ABSTRACT

Both species of northern right whale (North Atlantic, *Eubalaena glacialis* and North Pacific, *Eubalaena japonica*) are critically endangered. The overall distribution of these small, migratory populations is not well known, especially outside of summer. Passive acoustic monitoring is a tool that could provide information on locations of whales. Better distributional information will inform management efforts to reduce anthropogenic mortalities caused by both ship strikes and fisheries interactions. Recent research on passive acoustic monitoring is summarised, focusing on developments relevant to detection and classification of right whale calls. Some outstanding research requirements are outlined, including the need for the development of models to investigate the potential for risk reduction from acoustic data. Buoys capable of fully automatic whale vocalisation detection, classification and transmission to shore are currently under development.

SOMMAIRE

Les deux espèces de baleines franches nordiques (Atlantique Nord, Eubalaena glacialis et Pacifique Nord, Eubalaena japonica) sont gravement menacées d'extinction. La distribution générale de ces petites populations migratoires demeure à ce jour méconnue, plus spécifiquement en dehors de la période estivale. Le monitorage acoustique passif est un outil pouvant donner de l'information sur l'emplacement des baleines. Une meilleure indication de la distribution de ces dernières pourrait fournir des renseignements sur les efforts mis en place pour réduire les mortalités de cause anthropogénique causées à la fois par les collisions avec les bateaux et les interactions avec l'industrie de la pêche. Les recherches les plus récentes sur le monitorage acoustique passif sont ici résumées, en mettant l'accent sur les développements pertinents à la détection et la classification des vocalisations de baleines franches. Quelques lacunes tirées des conclusions de ces recherches sont soulignées, incluant le besoin de développer des modèles pour investiguer le potentiel des données acoustiques dans la réduction du risque. Des balises capables de détection, de classification et de transmission vers la côte sont présentement en développement.

1. INTRODUCTION

The purpose of this paper is to (a) provide very brief summaries of the conservation status of northern right whales (populations and threats) (b) outline potential application of passive acoustics to this problem (c) outline recent developments in passive acoustic research and development relevant to management (d) describe some outstanding research topics.

2. STATUS OF NORTHERN RIGHT WHALES

2.1 Population status

Both species of northern right whale are critically endangered. The North Atlantic right whale (Eubalaena glacialis), now regularly found only off the eastern US-Canadian seaboard, numbers about 300-350 (IWC 2001). Previous abundance

estimates of the North Pacific right whale (Eubalaena japonica) are not reliable but the numbers are certainly very low. The eastern North Pacific population in particular was severely impacted by illegal whaling during the twentieth century (Brownell, 2001).

Population projections for E. glacialis have been made by Caswell et al. (1999) and Fujiwara and Caswell (2001), based on a reliable long-term photo-identification record maintained by the New England Aquarium. These analyses suggest that current levels of mortality will lead the population to extinction within 100-400 years. Collisions with ships and entanglement in fishing gear have been identified as major sources of mortality (Knowlton and Kraus 2001). Anthropogenic mortalities are a large proportion (40%) of known mortalities. Falling survival rates over the last two decades have brought this small population into decline. According to Fujiwara and Caswell (2001), 'preventing the deaths of only two female right whales per year would

increase the population growth rate to replacement level'.

2.2 Management systems

On the eastern US seaboard, of 45 reliably documented deaths of right whales, 16 were confirmed to be due to ship strikes between 1970 and 1999 (Knowlton and Kraus 2001), and unreported deaths are also thought to occur (a proportion likely to be a result of collisions).

A Mandatory Ship Reporting scheme was introduced in 1999 that requires ships over 300 tonnes to report in to the US coastguard as they enter critical right whale habitats. The mariners then receive the most up-to-date information on right whale sightings in the area as well as on the characteristics of right whales, and are cautioned to keep a sharp look out for, and avoid all right whales. This system currently relies on rather patchy sightings information from research vessels, aerial surveillance and opportunistic sightings reported by mariners, fishermen, etc. As far as we are aware, there are no equivalent risk reduction or management schemes in place in the Pacific.

In the Bay of Fundy, Canada, a long-term data set documenting right whale distribution enabled the Government of Canada to propose changes to the shipping lanes away from areas of high right whale use. The Maritime Safety Committee of International Maritime Organisation approved and adopted this proposal, which came into effect in July 2003.

In the US, a series of comprehensive and region-specific recommended measures to reduce ship strikes (including speed reductions, alternative routing etc) have been developed in consultation with the industry, scientists and management authorities (Russell and Knowlton 2001). The proposed measures are currently under consideration by the relevant authorities in the US Government, with a view to incorporating them into national legislation.

The efficacy of any management scheme is reliant on the quality and quantity of data to inform it, and on effective mitigation measures being in place to deal with the conflicts between whales and human activity. Acoustic monitoring could compliment or supplement existing surveillance systems.

3. CONSERVATION APPLICATIONS OF PASSIVE ACOUSTICS

Passive acoustic monitoring is being considered as a method of detecting and locating right whales. It offers several advantages, including the ability to monitor autonomously for long periods in inhospitable conditions or in poor visibility (including at night). There is also potential for automation of the detection/classification process. Automation is a particularly useful feature when dealing with rare or widely dispersed populations such as northern right whales (Gillespie and Leaper, 2001).

However, the successful application of acoustic monitoring is dependent on a number of factors, including: sufficiently high vocalisation rates; appropriate detection ranges (source levels, propagation characteristics); detection algorithms robust to noise and capable of finding variable signals; sufficient understanding of detector efficiency and false alarm rates; and availability of technological expertise and equipment. A greater understanding of these issues has been obtained in recent years.

A workshop devoted to the subject of passive acoustics in management of right whales (Gillespie and Leaper, 2001) noted the following potential applications:

1. Detecting the presence of right whales in areas where there is little or no dedicated surveillance or other data, including poorly surveyed areas known to be used at least occasionally.
2. Assessing the predictability of right whale distribution patterns in known high use areas, including use of these areas in non-peak seasons.
3. Verifying aerial surveys.
4. Detecting a threshold number of whales in an identified high risk area that would 'trigger' some management action
5. Monitoring levels of ship traffic
6. Real-time detection systems for dynamic management

4. RIGHT WHALE ACOUSTICS

4.1 Summary

Descriptions of northern right whale acoustics (e.g. McDonald and Moore 2002, Vanderlaan et al. 2003) are more recent and more limited than the extensive study of the southern right whale (E. australis) repertoire made by Clark (1982, 1983). Tonal or pulsive sounds are typically in the 50-600 Hz range (but may reach over 1000Hz) and usually about 0.5-1 second duration. These calls are usually frequency-modulated: for example, McDonald and Moore (2002) categorised E. japonica calls as 'up', 'down-up', 'down', 'constant' and 'unclassified'. Another distinctive type of sound is the 'gunshot', described from E. australis by Clark (1982, 1983) and frequently heard from E. glacialis (e.g. Matthews et al. 2001, Laurinolli et al. 2003). As their name suggests, these are brief, broadband transients.

Various technologies are available and are being or have been in use to research and monitor right whales, including: towed arrays, acoustic tags, bottom-mounted recorders, sonobuoys and shore-cabled systems. Accounts of their use can be found in e.g. Clark et al. (2000), Matthews et al. (2001), McDonald and Moore (2002), Laurinolli et al. (2003), Nowacek et al. (2004) and ONR (1997).
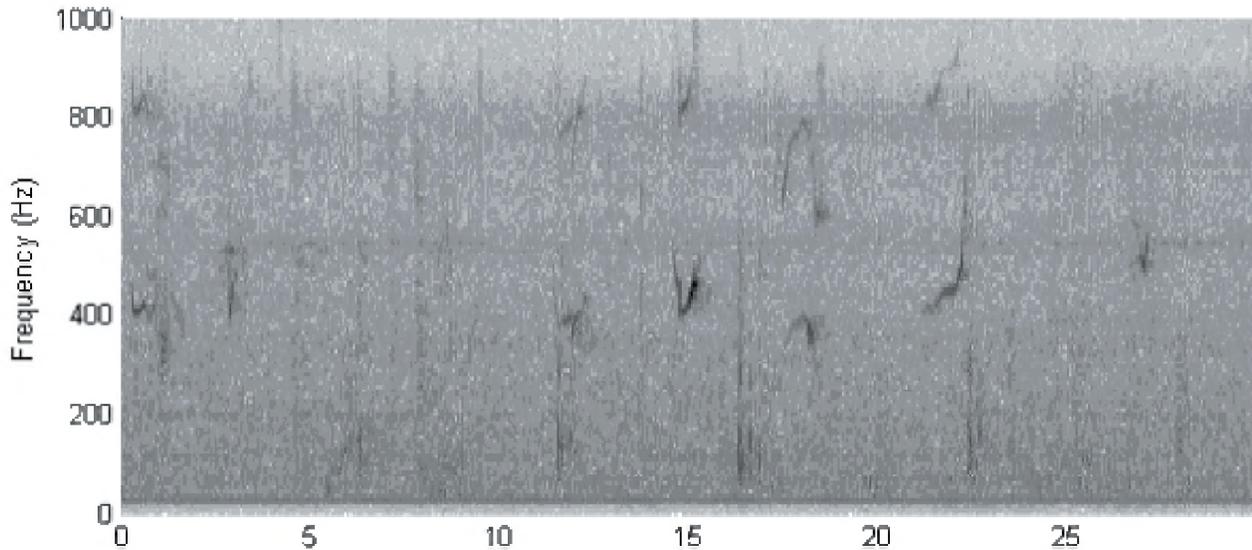
Figure 1. Spectrogram of a 30 second sequence of calls from North Atlantic right whales in the Bay of Fundy.

4.2 Recent developments relevant to passive acoustic monitoring

Characterisation of repertoire

In the early 1980s Clark (1982, 1983) found that southern right whales make common use of upsweeping calls (50-200 Hz, 0.5-1.5 seconds), and established that they are used to help animals maintain contact at a distance. Recent studies have now shown this type of call to be commonly produced by northern right whales too. The sample of E. glacialis calls illustrated in ONR (1997) from the Florida critical habitat were upsweeps. McDonald and Moore (2002) found upsweeps to be the predominant type (85%) in a sample of 511 from E. japonica. Gillespie (2004) and Matthews (2004) supply further evidence of their common occurrence in E. glacialis. In a study in the Bay of Fundy, Laurinolli et al. (2003) reported that 38 of 255 tonal sounds (15%) were low-frequency upsweeps. The relatively low proportion in this area may be due to increased use of other sounds with social or sexual functions (Parks, 2003). Although sounds corresponding to contact calls were not found in the study by Vanderlaan et al. (2003), they state that this is 'likely because we experienced high noise levels at these lower frequencies'.

Establishing the widespread use of upsweeps by northern right whales is a significant step forward for automatic detection systems. These calls are variable (see e.g. Gillespie 2004, Fig. 1) but also more-or-less invariant in shape. This means a good conceptual understanding of a common type of sound is now available when designing detectors and classifiers. These systems must nevertheless be flexible enough to respond to a degree of variation.

Vocalisation rates and patterns

A study of vocalisation rates of E. glacialis by Matthews et al. (2001) in the Bay of Fundy, Great South Channel and Cape Cod Bay, found the waiting time between vocalisations varied broadly from less than a minute to a few hours. In the Bay of Fundy, Vanderlaan et al. (2003) report waiting times of 2-700 seconds. According to McDonald and Moore (2002) after call bouts E. japonica in the Bering Sea 'commonly … became silent for an hour or more, with some animals not calling for periods of at least four hours'.

Several studies have now demonstrated that in general terms vocalisations are clustered in time. Matthews et al. (2001) reported clustering in most recordings made of E. glacialis, and Vanderlaan et al. (2003) also observed this in the Bay of Fundy. McDonald and Moore (2002) reported clustering of sounds in E. japonica.

Detection/Classification

Of significant interest is the discovery of excellent propagation characteristics in the Bering Sea, which allowed Wiggins et al. (2004) to detect right whales at long ranges of up to 50 km using bottom-mounted recording units. Unfortunately, the propagation and ambient noise conditions in many parts of the range of northern right whales are not so favourable.

Several techniques for automatic detection and/or classification were discussed in Gillespie and Leaper (2001): energy detectors, matched filtering, spectrogram correlation, neural networks, statistical pattern recognition and time sequence detection. Several studies have applied these techniques to cetacean vocalisations in general (e.g. Potter et al. 1994, Mellinger and Clark 2000). Some detectors have now been examined in more detail with respect to right whales in this

volume. Mellinger (2004) examines spectrogram correlation and neural networks. Gillespie (2004) examines statistical pattern recognition using image analysis of a spectrogram. Matthews (2004) looks at a parametric detection method for frequency-modulated tones. La Cour and Linford (2004) describe a method using independent component analysis.

Localisation

Passive acoustic data can provide highly accurate information on source location, providing a sufficient number of detectors are available and properly configured (Spiesberger and Fristrup, 1990). A number of estimation methods are available and under active investigation (this volume). Localisation is a valuable tool in applications such as behavioural, source-level estimation or ground-truthing studies. For example, Clark et al. (2000) placed seafloor recorders in groups of three, in the Great South Channel in 2000 and Cape Cod Bay in 2001, which allowed sounds to be localised and the results compared with aerial survey sightings.

However, accurate localisation may be unnecessary in many management applications. Any detection is also a crude localisation of an animal to somewhere within the detection range. Large-scale monitoring programmes do not necessarily require resolution at a finer scale than this. On the other hand, localisation may be useful for management applications in restricted spaces. Laurinolli et al. (2003) looked at the feasibility of passive acoustic localisation, particularly 'to clarify the frequency and numbers of right whale incursions into the [Bay of Fundy] shipping lane ...'.

An experimental real-time detection system

A real time acoustic monitoring system is currently being developed by Cornell University, Woods Hole Oceanographic Institute, and the International Fund for Animal Welfare. The goal is to implement a system of moored buoys to automatically detect and classify right whale sounds, and report this data on a daily basis. Each buoy will transmit right whale sound detections by cell phone or satellite to computers at Cornell University. Prototype buoys will be on trial off Cape Cod, USA, in 2004 and it is hoped that similar systems may be able to provide real time information on the presence and distribution of right whales in high-risk areas in the future.

5. SOME SUGGESTED TOPICS FOR FUTURE RESEARCH

5.1 Risk reduction modelling

There is a need for modeling of potential management systems, using data from passive acoustics, to assess the potential for risk reduction. Models should incorporate data on ranges of detection, reliability of detection, vocalisation rates, whale distribution and movement patterns as well as the potential hazards and possible mitigation measures. These models are likely to be area specific.

The temporal and spatial scales for any mitigation measure should be considered. The total time taken to receive vocalisations from whales, process the acoustic data, make a management decision and issue instructions to mariners needs to take place within the time frame for which the surveillance data are valid.

The research outlined in sections 5.2-5.4 will provide valuable information for this.

5.2 Detection

It is crucial to assess detection/classification systems for both their efficiency (probability of correctly detecting a sound) and their false alarm rates (probability of incorrectly detecting a sound). Any management system based on passive acoustics will require some expression of confidence in the detections/ classifications made. In the context of a monitoring system for right whales, it is likely that the cost of false detections will be high, since this may lead to management actions as described in section 2.

To assess the detection performance, some means of ground-truthing is required. Matthews (2004) and Gillespie (2004) attempt to assess the performance of their detectors using the following approach. They compare the numbers of calls detected from recordings when right whales were known to be present (as verified by visual survey) with recordings when whales were thought absent (using visual surveys and human listening).

This approach allows the performance of detection algorithms to be quantitatively compared. The efficiency of detection systems cannot be determined absolutely because the actual occurrences of whale sounds are not fully known. However, the relative efficiency of detectors (including human operators) can be tested. The false detection rate can be determined absolutely. These tests are, of course, conditional on the testing environment.

More extensive investigations of the performance of detection algorithms are needed. Ground-truthing using visual observations is likely to be an important part of this process.

5.3 Classification

Current classification methods for northern right whales search for upsweeping calls in about the 50-400 Hz range. Detector/classifier efficiency could be increased significantly if the desired signals could be expanded to include more of the right whale repertoire. Particularly in social or sexual contexts, right whales can be highly vocal and will use

relatively complex, higher-frequency calls in abundance (Fig. 1; Parks, 2003). However, expanding the call range of classifiers also increases the chances of confusion with humpback whales (Megaptera novaeangliae), which are highly vocal, and produce many sounds similar in frequency and duration to those of right whales. In addition to their well-known 'song', humpback whales produce other potentially confusing sounds in feeding or social contexts (Cerchio and Dahlheim 2001, Thompson et al. 1986).

Humpback whales do not restrict their singing to low-latitude wintering grounds. Singing has been documented in spring in the same regions where part of the North Atlantic right whale population regularly gathers (Mattila et al. 1987). They have also been reported singing in Alaskan waters occasionally in late summer (McSweeney et al. 1989).

5.4 Vocalisation rates

Vocalisations are a prerequisite for passive acoustic detection. Some areas where further information is desirable are listed below.

(i) No information appears to be available on vocalisation rates from right whales on migration.
(ii) The Great South Channel is an area with relatively heavy shipping traffic but rather sparse data on vocalisation rates.
(iii) Matthews et al. (2001) found higher vocalisation rates from right whales at night in the Bay of Fundy, and ONR (1997) reported higher rates between 1700 and 0500 in the southeast US. No published information is available on diurnal rates in the Great South Channel/Cape Cod, or in the North Pacific.
(iv) Information on vocalisations rates in the southeast US, the only known calving ground for E. glacialis, is poor. The only study we are aware of (ONR 1997) reported vocalisation rate results from a shore-cable system. The overall rate appeared to be relatively low (690 calls in about 300 hours or ~2.3 calls per hour), although the number of whales present is not stated. The tapes examined were 'most likely to yield vocalisations … based on … visual sightings in the vicinity of the fixed array'. It is not clear (a) whether the recordings were examined exactly when whales were within detection range (i.e. excluding periods when whales were out of range), and (b) whether tapes were selected on the basis of higher vocalisation rates (which would bias the estimated rate upwards). A fuller account of the acoustic and visual data would assist interpretation.

The statistics of the waiting times between vocalisations are crucial to understanding the efficacy of a management system based on passive acoustic detections. Detection probabilities will be more sensitive to the waiting times between clusters than the waiting times between vocalisations. The waiting time between clusters of calls relative to the period for which whales are within detection range will be a major factor affecting detection probability.

5.5 Extension of acoustic monitoring programme

It would be of interest to extend the acoustic monitoring programme to lesser-studied areas where animals may gather or pass through. Some areas e.g. Roseway Basin and Browns Bank off Nova Scotia, are known summering grounds for E. glacialis, but use of these areas varies considerably between years. Outside of summer, the whereabouts of a major part of the E. glacialis population is not known. The distribution of E. japonica outside of summer is also not well understood and no wintering grounds have been confirmed (Brownell 2001).

The US eastern seaboard forms migratory habitat for North Atlantic right whales en route between summer feeding grounds, and winter nursery and calving grounds. Shipping activity is intense in parts, and ship strikes are known to have occurred there. Passive acoustics could play a role in providing information on the whales' movements in this migratory coastal zone, including how close to shore the whales migrate, and whether they spend extended periods in certain areas.

6. CONCLUSION

The potential for passive acoustic techniques to contribute towards right whale conservation has been widely acknowledged, including by the Scientific Committee of the International Whaling Commission (IWC, 2002). In recent years there has been a substantial increase in applied studies of northern right whale acoustics, moving towards this goal. Further work is still required to determine how passive acoustic information and techniques might best be applied to practical management scenarios for risk reduction. Nevertheless, with continued research efforts, passive acoustic techniques could play an important role in reducing right whale mortality within a few years.

REFERENCES

Brownell, R. L. 2001. Conservation status of North Pacific right whales. J. Cetacean Res. Manage. Special Issue 2: 269-286.

Caswell, H., Fujiwara, M. and Brault, S. 1999. Declining survival probability threatens the North Atlantic right whale. Proc. Natl. Acad. Sci 96: 3308-3313.

Cerchio, S., and Dahlheim, M. 2001. Variation in feeding vocalizations of humpback whales Megaptera novaeangliae from Southeast Alaska. Bioacoustics 11: 277-295.

Clark, C. 1982. The acoustic repertoire of the southern right whale, a quantitative analysis. Animal Behaviour 30: 1060-1071.

Clark, C. 1983. Acoustic communication and behavior of the southern right whale. In: R.S. Payne (ed.), Behavior and Communication of Whales. Westview Press: Boulder, CO. pp. 163-198

Clark, C.W., Gillespie, D., Moscrop, A., Fowler, T., Calupca, T., and Fowler, M. 2000. Acoustic sampling for right whale vocalizations in the Great South Channel using sea-floor pop-up recorders. Report of the Right Whale Consortium, October 26-27, 2000, Boston, Massachusetts.

Fujiwara, M., and Caswell, H. 2001. Demography of the endangered northern right whale. Nature 414: 537-543.

Gillespie, D. 2004. Detection and classification of right whale calls using an 'edge' detector operating on a smoothed spectrogram, Canadian Acoustics Vol. 32 #2, June 2004.

Gillespie, D. and Leaper, R. 2001. Report of the Workshop on Right Whale Acoustics: Practical Applications in Conservation. Paper SC/53/BRG2 presented to Scientific Committee of International Whaling Commission, London, 2001. Available from secretariat of IWC, Cambridge, UK.

IWC 2001. Report of the workshop on status and trends of western north Atlantic right whales. J. Cetacean Res. Manage. Special Issue 2: 61-87.

IWC 2002. Report of the Scientific Committee. J. Cetacean Res. Manage. 4 (Supplement) p.45

Knowlton, A., and Kraus, S. 2001. Mortality and serious injury of northern right whales (Eubalaena glacialis) in the western North Atlantic. J. Cetacean. Res. Manage. Special Issue 2: 193-208.

La Cour, B. and Linford, M. 2004. Detection and classification of north Atlantic right whales in the Bay of Fundy using independent component analysis Canadian Acoustics Vol. 32 #2, June 2004.

Laurinolli, M., Hay, A., Desharnais, F. and Taggart, C. 2003. Localization of north Atlantic right whale sounds in the Bay of Fundy using a sonobuoy array. Mar. Mamm. Sci. 19:708-723.

Matthews, J. 2004. Detection of frequency modulated calls using a chirp model, Canadian Acoustics Vol. 32 #2, June 2004.

Matthews, J., Brown, S. Gillespie, D., Johnson, M., McLanaghan, R., Moscrop, A., Nowacek, D., Leaper, R., Lewis, T. and Tyack, P. 2001. Vocalisation rates of the North Atlantic right whale. J. Cetacean. Res. Manage. 3(3):271-282

Mattila, D., Guinee, L. and Mayo, C. 1987. Humpback whale songs on a North Atlantic feeding ground. J. Mammal. 68:880-883.

McDonald, M. A., and Moore, S. 2002. Calls recorded from North Pacific right whales (Eubalaena japonica) in the eastern Bering Sea. J. Cetacean. Res. Manage. 4:261-266.

McSweeney, D., Chu, K., Dolphin, W. and Guinee, L. 1989. North Pacific humpback songs: a comparison of south eastern Alaskan feeding ground songs with Hawaiian wintering ground songs. Mar. Mamm. Sci. 5:139-148.

Mellinger, D.K. and Clark, C.W. 2000. Recognizing transient low-frequency whale sounds by spectrogram correlation. J. Acoust. Soc. Am. 107: 3518-3529.

Mellinger, D 2004. Comparison of optimised methods for detecting right whale calls, Canadian Acoustics Vol. 32 #2, June 2004.

Nowacek, D.P., Johnson, M.P., and Tyack, P.L. 2004. North Atlantic right whales (Eubalaena glacialis) ignore ships but respond to alerting stimuli. Proc. R. Soc. Lond. B. 271:227-231

ONR 1997 Northern Right Whale Monitoring Project: Final Report. Office of Naval Research.

Parks, S., 2003. Response of north Atlantic right whales (Eubalaena glacialis) to playback of calls recorded from surface active groups in both the north and south Atlantic. Mar. Mamm. Sci. 19(3):563-580.

Potter, J., Mellinger, D. and Clark, C. 1994. Marine mammal call discrimination using artificial neural networks. J. Acoust. Soc. Am. 96: 1255-1282.

Russell, B.A. and Knowlton, A.R. 2001. Recommended Measures to Reduce Ship Strikes of North Atlantic Right Whales. Report to NMFS North east and South east Implementation Teams, August 2001.

Spiesberger, J. and Fristrup, K. 1990. Passive localisation of calling animals and sensing of their acoustic environment using acoustic tomography. American Naturalist 135: 107-153.

Thompson, P., Cummings, W. and Ha. S. 1986. Sounds, source levels, and associated behavior of humpback whales (Megaptera novaeangliae), Southeast Alaska (USA). J. Acous. Soc. Am. 80: 735-740.

Vanderlaan, A., Hay, A. and Taggart, C. 2003. Characterization of North Atlantic right-whale (Eubalaena glacialis) sounds in the Bay of Fundy. IEEE J. Oceanic Engineering 28: 164-173.

Wiggins, S., McDonald, M., Munger, L., Hildebrand, J. and Moore, S. 2004. Waveguide propagation allows range estimates for North Pacific right whales in the Bering Sea, Canadian Acoustics Vol. 32 #2, June 2004.

Acoustic Recording Systems For Baleen Whales And Killer Whales On The West Coast Of Canada

Svein Vagle[1], John K.B. Ford[2], Neil Erickson[1], Nick Hall-Patch[1], and Grace Kamitakahara[1]

1) Institute of Ocean Sciences, 9860 West Saanich Road, Sidney, BC, V8L 4B2
2) Pacific Biological Station, 3190 Hammond Bay Rd, Nanaimo, BC, V9T 6N7

ABSTRACT

The threat to the survival of several whale species and the introduction of the Species at Risk Act (SARA) has highlighted the need for better knowledge about the biology and ecology of marine mammals in Canadian waters. The North Pacific right whale (*Eubalaena japonica*), once plentiful across much of the North Pacific Ocean, is now rarely seen in coastal British Columbian waters, and the number of killer whales (*Orcinus orca*) in southern British Columbia has been steadily decreasing in recent years. The recovery plan for these species is based on the gathering of baseline data on occurrence, distribution, abundance and habitat, and one significant component of this data collection is based on the deployment of multiple passive acoustical recording systems off the coast of British Columbia. In addition to the development and use of a simple but effective two-hydrophone array, two different autonomous passive acoustical instruments have been developed, one deployable at shore sites and the other for offshore locations. To limit data storage and power requirements, both of these systems have been equipped with killer whale recognition hardware to record only when the probability of killer whales in the area is relatively high. In addition the offshore units have been designed as hybrid recorders, sampling at 1000 Hz for the larger baleen whales and 20 kHz when killer whales are present. Both of these instruments have been designed for deployment periods as long as 12 months and are presently deployed in locations along the BC coast. The analysis of the large data sets from these instruments is a challenge and we are currently investigating the use of neural network algorithms to perform not only species recognition but also, with regards to the killer whale population, clan or group identification. The goal is to adapt these algorithms directly into the self-contained instruments.

RÉSUMÉ

La menace à la survie de plusieurs espèces de baleines et la présentation de la Loi sur les espèces en péril (LEP) a fait ressortir le besoin d'approfondir nos connaissances au sujet de la biologie et de l'écologie des mammifères marins dans les eaux canadiennes. La baleine franche du Pacifique nord (*Eubalaena japonica*), pourtant abondante autrefois dans une grande partie de l'océan Pacifique Nord, est maintenant rare dans les eaux côtières de Colombie-Britannique, et le nombre d'épaulards (*Orcinus orca*) dans le sud de la Colombie-Britannique a diminué de façon régulière ces dernières années. Le plan de rétablissement de ces espèces est fondé sur la collecte de données de référence concernant la présence, la distribution, l'abondance et l'habitat, et un élément significatif de cette collecte de données se fonde sur le déploiement de multiples systèmes d'enregistrement acoustique passif au large de la côte de la Colombie-Britannique. En plus du développement et de l'utilisation d'un réseau de deux hydrophones, qui est simple mais efficace, deux appareils acoustiques passifs autonomes ont été développés; un appareil pouvant être déployé sur le rivage et l'autre pouvant être déployé au large. Les deux systèmes sont munis de matériel de reconnaissance qui enregistre seulement s'il existe une forte probabilité qu'il y a des épaulards dans la région afin de limiter les exigences en matière de stockage de données et d'alimentation en énergie. En outre, les diapositifs utilisés au large ont été conçus en tant qu'enregistreurs hybrides; le taux d'échantillonnage est de 1000 Hz pour les grandes baleines à fanons et de 20 kHz pour les épaulards. Les deux appareils ont été conçus en vue de les déployer pour un période maximale de 12 mois et ils sont actuellement déployés le long de la côte de la C.-B.. L'analyse des grands ensembles de données provenant de ces appareils est un défi et nous étudions actuellement la possibilité d'utiliser des algorithmes de réseau neurologique pour exécuter non seulement l'identification d'espèce mais, également, l'identification du clan ou du groupe en ce qui concerne la population des épaulards. L'objectif est d'incorporer ces algorithmes directement aux appareils autonomes.

1 INTRODUCTION

A large number of marine mammal species are living in or migrating through the offshore waters of the west coast of Canada, but the understanding of their life cycle and migratory patterns is currently very limited. In addition to the killer whale and the North Pacific right whale, which will be specifically dealt with in this paper, the marine mammal species on the west coast that we are interested in monitoring include fin whales *(Balaenoptera physalus)*, blue whales *(Balaenoptera musculus)*, humpback whales *(Megaptera novaeangliae)*, gray whales *(Eschrichtius robustus)*, and sperm whales *(Physeter macrocephalus)*. All of these species use acoustic energy for communication and some for echolocation. The frequency band used by the different animals of interest are these: fin whale 14-750 Hz, blue whale 10-390 Hz, humpback whale 30-8000 Hz, gray whale 20-2000 Hz, sperm whale 100-30000 Hz, killer whale 1200-25000 Hz, and North Pacific right whale 20-900 Hz (National Research Council, 2000).

1.1 Killer whale (*Orcinus orca*)

Researchers classify northeastern Pacific killer whales into three distinct populations, or 'ecotypes': "residents" feed exclusively on fish and squid and are very vocal, "transients" prey exclusively on marine mammals and seabirds and are very quiet (Barrett-Lennard et al., 1996; Ford et al., 2000)), and "offshores", which are very poorly known.

The resident killer whales in the northeastern Pacific are the most frequently encountered whales in this region and have been studied for more than 30 years. Large groups of 10-50 animals can often be seen in coastal waters in the summer months, and these groups have consistent and stable memberships. The population assessment of resident killer whales based on photo-identification of individuals using natural markings has identified the presence of complex matrilineal societies(Biggs et al., 1987; Ford et al., 2000). In the late 1970s, Ford began studies of the group specific vocalizations, or dialects, of the resident communities (Ford, 1989,1991). He found that pods have stable repertoires of stereotyped calls and that these repertoires differ among resident pods and matrilines. It has been suggested that these dialects may play a role in defining the group identity and outbreeding. The resident killer whales predominantly vocalize in the 1-10 kHz range. Figure 1 shows examples of spectrograms of vocalizations from two clans, one from the northern community of resident whales and one from the southern community, clearly showing the different dialects of the two groups.

Even though killer whales are protected, the Committee on the Status of Endangered Wildlife of Canada (COSEWIC) has declared the northern community of resident killer whales as threatened and the southern community as endangered. Since 1995 the number of animals in the southern community as declined from about 100 to approximately 80 (Figure 2).
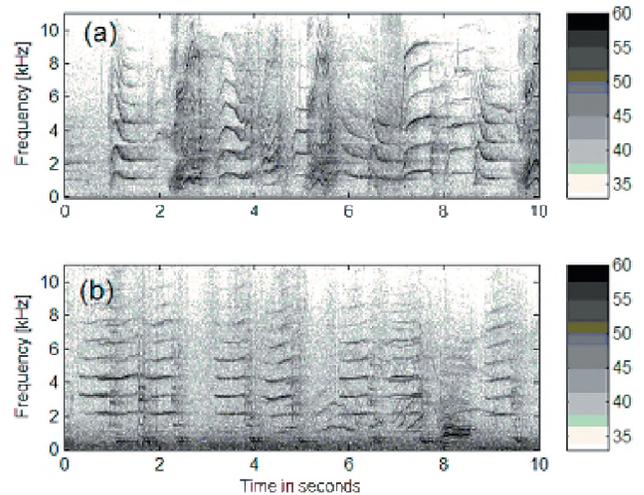
Figure 1. (a) Spectrograms from a northern resident community pod, and (b) a southern resident community pod show the distinct differences between the vocalizations of members of the two killer whale communities. The greyscale represents sound pressure levels in decibels relative to an arbitrary reference.

The reasons for the decline are not known, but factors including vessel disturbance, underwater noise, declining fish stocks, and man-made contaminants have been suggested (Ross et al., 2000). Before any recovery strategy can be successful, we need to learn more about the biology and ecology of these animals. For example, one critical knowledge gap is to understand what they do in the winter months. As can be seen in Figure 3, the majority of the sightings take place in the summer and fall months from May to November. The whales leave the summer core areas in the winter months and their distribution during this period is unknown. Food might be relatively scarce during this period, especially for the salmon-feeding resident communities, and studies have shown that most mortalities take place during this period.

It is therefore crucial for any recovery strategy to study the winter behaviour and habitat of the animals. Vessel- and shore-based visual surveys are very difficult in the winter months. Frequent foul weather periods make it hard to see the animals at the ocean surface, and the number of hours with daylight is limited. Long-term satellite tracking using radio transmitters attached to the whales is also not an option in this region given the animals' endangered or threatened status and the potential risk of harmful effects from surgical tag attachment. Because of the relatively high vocalization rate of the resident killer whales and because of their distinct dialects, these mammals are well suited for monitoring by passive acoustic devices.

1.2 North Pacific right whale (*Eubalaena japonica*)

The North Pacific right whale is a large robust baleen whale that once was plentiful across much of the North Pacific
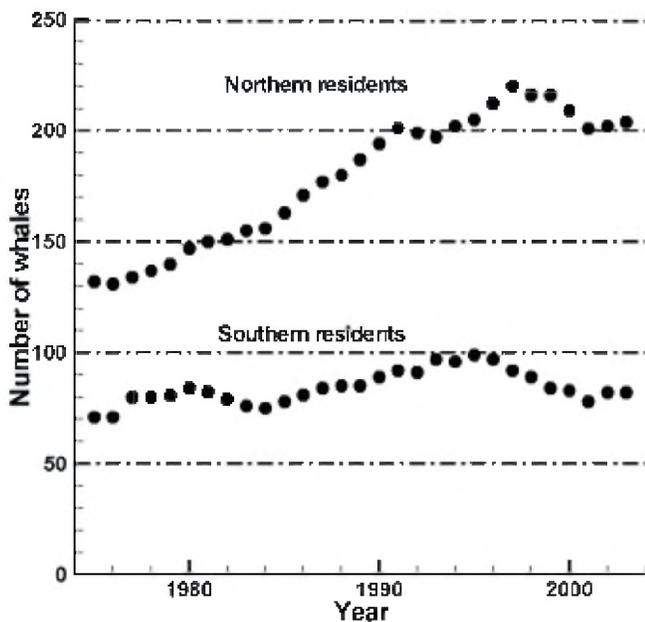
Figure 2. The sizes of the resident killer whale communities from 1974 to 2003 (Ford et al., 2000; Ford, unpublished data)

Ocean but now is rarely seen in coastal British Columbia waters (Rosenbaum et al., 2000). These whales were decimated by whaling during the 1800s and, despite having received international protection since 1935; their numbers have not appeared to recover. Basic aspects of the biology and ecology of the species remain unknown. It is unclear how many individual animals exist and the location of the calving grounds is still a mystery (Brownell et al., 2001). The scarcity of recent sightings suggests that the population may number less than 100, at least in the eastern North Pacific. Causes for the continued low or declining population are not known. However, climate-driven regime shifts are manifested faster at lower trophic levels in the marine echosystems (Benson and Trites, 2002), and an increase in surface water temperature could result in declining zooplankton populations (Roemmich and McGowan, 1995). Right whales feed exclusively on zooplankton, and preferably on large calanoid copepods. They have a narrow range of acceptable
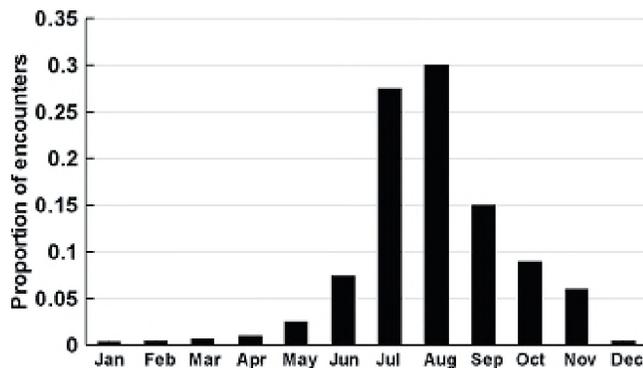
Figure 3. Proportion of northern resident killer whale encounters by month (Ford, unpublished data)

prey species and require prey concentrations of exceptionally high densities, which again are dependent on factors such as nutrient levels, currents and temperature. Kenney (2001) suggested this might make the right whale, which is a low trophic grazer, more sensitive than other cetaceans to impacts from global climate change.

The right whales produce the "up" call, which is a frequency-modulated upsweep in the 50-200 Hz range (McDonald & Moore, 2002). This call is one of the most common calls and is therefore a useful indicator of the presence of right whales for passive acoustic tracking devices (see other papers in this issue).

1.3 Recovery Strategy

Fisheries and Oceans Canada (DFO) has outlined a new a recovery strategy aimed at restoring the North Pacific right whale to Canadian waters and maintaining the long-term viability of the population. To reach this goal, the recovery objectives include; 1) to gather baseline data on occurrence, distribution, abundance and habitat to support recovery efforts and 2) to conduct long-term monitoring of the status of these animals and evaluate the effectiveness of mitigation strategies.

We have developed a three-pronged approach to the use of passive acoustic systems to aid in reaching these goals. In *situ* use of towed arrays can be very useful as a supportive tool in the gathering of baseline data, and self-contained recording instruments, which also can be used to collect baseline data in the earlier stages of the recovery, are well suited for long-term monitoring of the mammal population.

2 BASELINE DATA GATHERING

2.1 Towed arrays

The information from towed hydrophone arrays can add valuable information to visual marine mammal surveys from different vessels. With the use of more than one hydrophone it is possible to obtain directional information relative to the ship heading. Two spaced hydrophones embedded in a single cable and being towed behind a ship will have left-right ambiguity, i.e., it is impossible to know whether the vocalizing animal(s) is on the left or the right side of the vessel. This ambiguity can be resolved by adding a third hydrophone.

An array that can be towed at or near the cruising speed of the vessel allows for large area coverage compared to single hydrophones lowered from stationary platforms. A well-built system can also be operated 24 hours a day and during a range of sea states.

The design criteria we used for our towed hydrophone system are these: 1) the system has to be light weight to accommodate mounting on a variety of different small and

large vessels; 2) it has to be easy to use so that the system can be loaded onto a vessel of opportunity with no specially trained operator; 3) it has to be relatively inexpensive to acquire and maintain; and 4) its design must accommodate future implementation of artificial intelligence algorithms for species recognition and operator alarms.

When using ship-based systems, there is in effect no electrical power limitation and no limit to onboard real-time signal processing, something that in many respects simplifies the design when compared to the self-contained battery powered instrumentation discussed below.

Our operational system, as shown in Figure 4 and Figure 5(a), consists of 200 m of armoured cable with a drogue at the far end. The cable is wound on a small electrically operated winch giving a towing length of approximately 170m. The two hydrophones (Biomon model BM212) have a separation of 10 m and an acoustic bandwidth of 1 - 41 kHz with a maximum received response of −151 dB re 1V/μPa. The built-in 1 kHz high-pass cut-off frequency was included to limit the possibility that low-frequency ship noise and flow noise might saturate the hydrophone preamplifier. The signals from the hydrophones are passed through two SRS SR-650 filter units, which act as signal amplifiers (with up to 90 dB gain) and as band-pass filters. The SR-650 units are controllable via RS232 serial protocol from a laptop computer. Typically we have been filtering the signals between 3 and 10 kHz to reduce ship noise while preserving killer whale vocalizations. The amplified and filtered signals are passed to the 'line in' connectors on the laptop and recorded to disc and/or used to estimate source direction and vocalization identification.

Raw data are written to hard drive in VOC format using the built-in sound acquisition system at a rate of 22050 Hz at 16 bits per channel. The VOC format, which is compatible with numerous commercially available playback programs, incorporates different block types allowing additional information such as time stamping, GPS position data and user comments to be recorded. System specifications are listed in Table I.
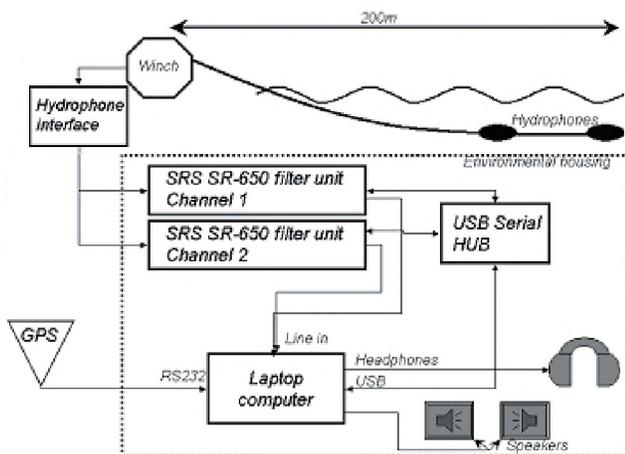
Figure 4: Schematic diagram of towed hydrophone array system.

The software generates a continuous scrolling spectrogram display, and a cross correlation algorithm using a user-defined frequency range generates a user-friendly graphical display showing, in real time, the direction of the sound sources relative to the towing vessel, thus aiding in locating them. Future developments of this system include the incorporation of real time species and group specific vocalization recognition algorithms. The plan is to incorporate these into alarm routines that can inform ship operators and scientist when mammal vocalizations have been detected by the system.

The disadvantages of these types of systems include the cost of ship time and the acoustical noise from the towing vessel and the limited time that any given area can be surveyed. Ship noise entering the hydrophone arrays generally limits their use to higher frequencies and generally limits the monitoring capabilities to the smaller toothed whales. However, towed arrays have been used in sperm whale surveys for many years (e.g., Barlow & Taylor, 1998).

3   LONGER TERM MONITORING

3.1 Shore based monitoring system

For longer term passive acoustic monitoring of killer whales in coastal waters - for example, to determine their location during the winter months - we designed and built a system that can be placed on shore in remote areas with a hydrophone deployed in nearby waters. Such a system must be rugged to withstand winter weather and possible animal attacks as well as having the capability of recording for extended periods up to a year.

The main advantages of our present system are: 1) the relative ease of deployment by two people in a small boat; 2) the ease at which data can be recovered and the instrumentation serviced; 3) the economics of the electrical power supply. (it is easy to add a number of car batteries to a site onshore).

In addition we have implemented solar panels for charging batteries and have been experimenting with the use of wind power for winter deployments in northern waters where the hours of daylight are severely restricted. The system we have designed and constructed, called "Orcabox", is shown in Figure 5(b) and the schematic diagram of the main components is shown in Figure 6.

The underwater part of the system consists of a broadband hydrophone (50-20,000 Hz) deployed at the end of an armoured cable (<200 m). For these types of systems the cable going through the tidal zone is the most vulnerable for longer-term deployments. Several approaches have been attempted to address this problem using different types of conduits, with varying degree of success. The chosen approach will be quite site dependent. A cable with a thick polyurethane coating supplied by "Specialty wiring and Cable" in Calgary, AB has worked very well. However, we have also had good experiences with using old CTD cables

for this purpose. The outer steel jacket of a good CTD cable is quite resistant to the chafing in the tidal zone.

The dry end of the armoured cable is connected to a 0.53 m 0.44 m by 0.22 m waterproof case mounted with lead acid batteries in a lockable solid aluminium housing on dry land, away from the wave zone. Solar panels can be positioned at appropriate sites in the vicinity of the aluminium housing (Figure 5(b)). Inside the waterproof case, the hydrophone signal is passed through a preamplifier and an 80 dB automatic gain control (AGC) circuit before going into an analog acoustic pattern recognition circuit (Figure 6). This circuit is designed to recognize killer whale calls and consists of 3 parallel filters; the first two being peaking filters with a Q of about 10 centered on 4 kHz and 14 kHz, and the final one is a 50-450 Hz band-pass filter. The third filter is used to differentiate boat noise from killer whale calls, as there is little energy in that range in whale calls.

The output sine waves from the filters are squared and counted by a microcontroller. The algorithm then looks for a certain number of positive going transitions in the output of these squaring circuits for a given period of time.

When the user-defined conditions are met, this circuit powers up the PC104 computer and the hard drive; then records for a specified period (typically 3-10 minutes) at 20 kHz using 8 bit words. The hard drive in this system is easily replaceable for easy transfer of data.

As before we are using the VOC format to allow for time stamping of the data. (The specifications for this system are also listed in Table I.)

Figure 7 shows the results from a nearly month-long test deployment off Hanson Island in Johnstone Strait, BC, during the summer of 2002. This area is known for frequent killer whale sightings during the summer months. Manual inspection of each recorded sound file confirms this, with many days when killer whale vocalizations could be heard for more than 15 minutes per hour (Fig. 7(a)). A more surprising result is the number of false triggers associated with boat noise (Fig. 7(b)), suggesting that these boats generate significant noise in the frequency bands used by the killer whales and therefore used by our trigger circuit. Jet boat noise is especially difficult to discern from actual killer whale vocalizations because of the significant harmonics in the exact same frequency bands as the ones used by the whales. However, preliminary comparisons with nearby listening stations suggest that the monitoring system is capable of recording the presence of killer whales in the area, and in the more remote locations of the coast where we plan to locate these devices the number of all sorts of vessels is very limited, especially during the winter months.

The first such system is presently deployed off Langara Island (Table II) at the northern tip of the Queen Charlotte Islands to monitor for the presence of wintering resident killer whales.

	Two hydrophone towed array	Shore base monitoring system "Orcabox"	Underwater deployable monitoring system (PATC)
Frequency range	1-22.05 kHz	0.1-10 kHz (programmable)	0.05-1 kHz and 0.05-10.5 kHz (programmable)
Size	Electronics: 0.75 by 0.6 by 0.6 m. Winch: 1 by 0.9 by 1.2 m	1.5 by 0.5 by 0.5 m Solar panels: 1.3 by 0.6m	0.21 m diam. by 1.7 m long. Deployment depth to 1000m
Power consumption	Electronics: 90W Laptop: 70W Winch: 1500W	Idle mode: 1.5 W Recording: 7.5 W	Average power: 1.8 W
Disk capacity	Gbytes (computer hard disk)	20 Gbytes	24 Gbytes (expandable)
Deployment periods	Days, weeks. Duration of survey. Data backed up on CDs.	<12 months (batteries charged by solar cells. Disk space limited)	<12 months (depends on size of battery pack and number of hard disks installed)
Species recognition	Under development	Killer whale recognition circuit.	Killer whale recognition circuit (high freq. sampling). Low frequency sampling presently on timer.

Table I. System specifications for the three passive monitoring systems discussed in the text; a two hydrophone towed array, a shore based monitoring system called the "Orcabox", and a long- term underwater deployable monitoring system called "Passive Acoustic Tracking of Cetaceans" (PATC).
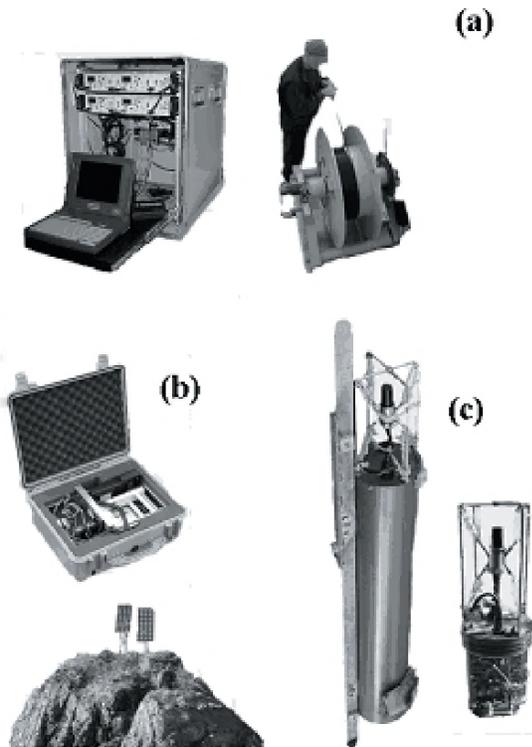
Figure 5: Photographs of the three different passive mammal-tracking systems discussed in the text. (a) The towed array consists of an armoured cable with two hydrophones connected to amplification, filtering units and a laptop computer where the signals are processed and recorded with GPS positions and time information. (b) A shore based battery powered system that uses solar panels to charge batteries. Shown are also two solar panels mounted on a local rock. (c) A self-contained battery operated recording package designed for offshore, submerged deployments, typically on deep moorings. The hydrophone and electronics are shown with and without the pressure housing which also contains the battery pack.

Figure 6. Schematic diagram showing the main components of the shore based acoustic monitoring system, or "Orcabox".
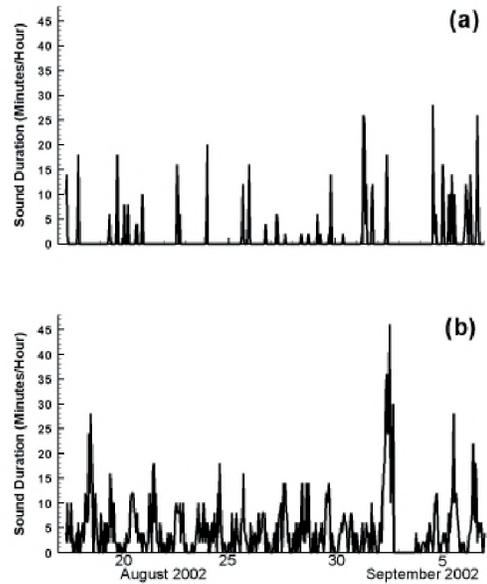
Figure 7. Results from Hanson Island, Johnstone Strait, showing the (a) the number of minutes per hour that killer whales were present, and (b) the significant time with boat noise.

3.2 Underwater deployable system

For offshore long-term monitoring or in areas where the water depth is more than a few hundred meters, it is necessary to use stand-alone self-contained instruments. A number of systems with a range of capabilities have been developed for a variety of applications by different groups over the years, including; "L-CHEAPO" (Worcester et al., 1995), "Haruphone" (Fox et al., 2001), "Pop-up" (Clark et al., 2002), "EARS Buoy" (Wright, 2003), and "ARP" (Munger et al., 2004; Wiggins et al., 2004).

To limit data storage and power requirements we wanted such a system to incorporate some limited artificial intelligence to minimize operation when no mammal signals are present. The passive acoustic tracking of cetaceans (PATC) instrument designed and built for this purpose is shown in Figure 5(c). A block diagram of the most significant components in this system is shown in Figure 8.

This instrument incorporates low-power components and is designed for deployment periods lasting up to 12 months and incorporates timing and vocalization recognition schemes as outlined below. The aluminium pressure housing has been designed for depths up to 400m, while the broad band (10-20000 Hz) hydrophones presently used (VEMCO) are certified only to 100 m. However, this hydrophone can easily be swapped for another hydrophone with better depth rating if so desired. The signal from the hydrophone is passed through the same amplifier and AGC circuit as used in the "Orcabox". From there on the design differs significantly from the previously discussed system.
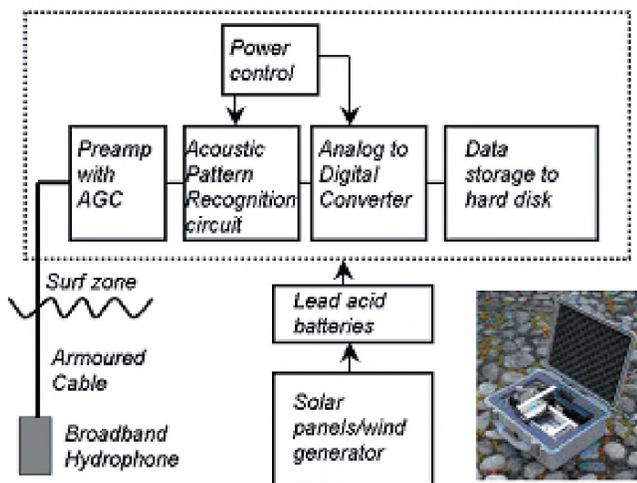
Due to the requirement of the PATC instruments to record the low-frequency signals from the larger baleen whales such as the North Pacific right whale in addition to any killer whale calls, the system was designed as a hybrid system; recording at a lower 1000 Hz sampling rate (50-500 Hz bandwidth) during most of the time, while recording at 21000 Hz (50-10500 Hz bandwidth) in the presence of killer whales. When a reliable automated scheme to detect the larger whale calls becomes available, the design will be modified to automatically decide on the sampling scheme to use based on the type of animal call. Research is underway to develop such algorithms.

The signal from the AGC circuit is sent both to an analog acoustic pattern recognition circuit, equivalent to the one described earlier, and via a 500 Hz low-pass filter (LPF) directly to the analog to digital converter (A/D), which has programmable sampling rate. Most of the time the signal will be digitized at 1000 Hz (16 bits) for time periods controlled by a low-power CF1 microcontroller. However, if during these recording periods the acoustic pattern recognition circuit detects killer whale calls, the instrument automatically switches to a 10 kHz LPF and increases the sampling rate to 20000 Hz for a pre-programmed period before returning to the low-frequency sampling scheme. All data are time-stamped and stored on one of several 8Gbyte hard drives.

The frequency and length of the 1000 Hz sampling periods are determined prior to instrument deployment and are based on the duration of the deployment, the size of the battery pack and the available disk space. We have presently deployed two instruments with one hour recording periods followed by one hour of low-power sleep mode. Sixteen Gbytes have been reserved for the low-frequency sampling and 8 Gbytes are allocated for high frequency killer whale recordings. This arrangement is intentionally made quite flexible, making it easy to add or remove hard discs and added battery packs for different requirements.
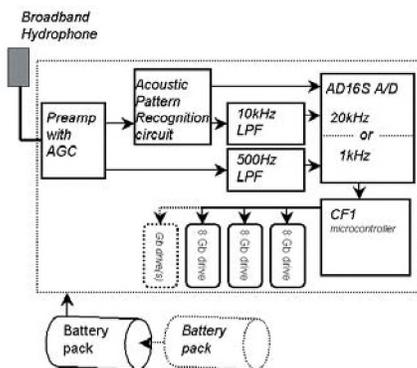
Figure 8. Schematic diagram showing the main components of the self-contained PATC instrument designed for offshore large baleen and killer whale monitoring.

With any automated recording scheme where the data recording is based on the fulfilment of certain predetermined conditions such as the acoustic pattern recognition circuit used here, it is important that the probability of false triggers is kept to a minimum. There will always be a possibility that a nearby noise source, such as parts of the mooring itself, will trigger the recording circuitry continuously or periodically and therefore fill up the disc space and drain the batteries prematurely. To eliminate this possibility we designed a recording scheme as outlined in Figure 9.

During a typical six month deployment the CF1 microcontroller will continuously repeat a 2 hour cycle: recording time stamped 500 Hz signal to hard disk for 1 hour; then going into a low power ($6\mu A$) mode for 1 hour. If during the recording periods the acoustic pattern recognition circuit detects killer whales in the area, the system will switch to the faster 20 kHz sampling and record a 10-minute section of data at this higher rate. However, before deployment the system has been allocated a fixed number of these recording periods per day and if the quota has been used up on a given day the system will not record. However, if the quota from one day was not used, this number of recording periods will be added to the quota for the next day to allow for more frequent sampling during some limited periods when killer whales might be in the area. If no killer whales are detected over a predetermined period (typically one week) the system has been programmed to record ambient noise to fill up part of the quota. These data will be used to investigate the ambient noise conditions in the instrument location and can also be used for environmental monitoring such as wind speed, rainfall, and overall shipping noise levels. This information might be useful in interpreting the vocalization data.

Even though the PATC instruments can easily be modified for bottom-mounted deployments, they are specifically designed to be deployed on moorings in deep water. In water depths of more than a couple of hundred meters there are several advantages to placing these monitoring devices higher up in the water column. The cost of components such as pressure housings, connectors and hydrophones increases significantly if the deployment depth is more than about 400m. Also, the sound speed profile in the water column is often such that there is a surface duct in the upper ocean in which the sound from vocalizing whales can travel much greater distances (Urick, 1983). This duct might be seasonal or only occur at certain times of the day or in certain locations. However, any instrument deployed below such a surface duct would be limited to vocalizations from a relatively small area above the instrument. The instruments presently deployed on the west coast of Canada have been placed at a depth of 60 m in water depths of 400 m and 2110 m respectively (Table II).

Clearly the deployment location and depth will depend on the objectives of the monitoring program. If the goal is to detect the presence of any marine mammals in an area as large as possible, which is the goal for us with respect to North Pacific right whales, the hydrophone should be placed in the sound channel. However, if the objective is to monitor a certain area for the presence or absence of mammals, deeper instruments might be preferable.

In the search for the North Pacific right whale, the optimum instrument locations have to be determined based
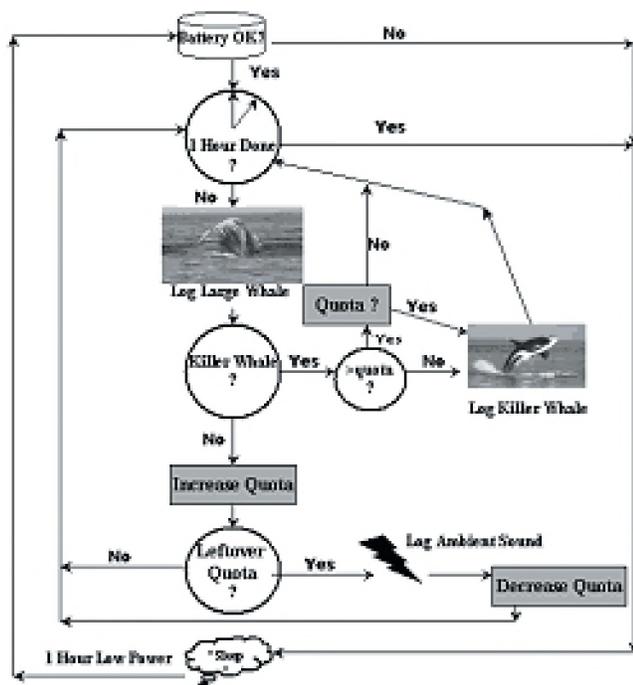
Figure 9. The recording scheme used in recent PATC deployments off the west coast of BC.

on historic sighting data. However, because these are few and far between, we based the initial instrument locations on survey data from other large baleen whales over the last number of years, assuming that the right whale will use the same feeding grounds if they are present.

Table II. Deployment information for the three instruments deployed during winter 2003/2004.

	"Orcabox"	PATC1	PATC2
Latitude	54 14.1N	50 57.5N	48 45.4N
Longitude	132 58.0W	129 59.6W	126 22.1W
Water depth	15m	2110m	400m
Hydrophone depth	15m	60m	60m

4 DATA INTERPRETATION

With the large data sets acquired by these systems comes the challenge of extracting species and in the case of killer whale vocalizations, matriline dialects in an efficient way. A quantitative measure of acoustic similarity is important to allow for species identification, or in particular cases determination of social groups or individual animals. Several approaches are presently being investigated as ways of detecting particular whale calls in sound recordings using both time-based and spectrogram-based techniques (see other

papers in this issue). However, due to our initial interest in identifying different killer whale dialects and some familiarity with the neural network approach (Deecke and Ford, 1999) we chose this as a starting point for our evaluation of a suitable technique. We are presently just starting this work so the results presented below are by no means a complete investigation into the use of neural networks for species and dialect recognition.

4.1 Neural networks

Artificial neural networks were developed by modeling biological systems of information processing (Dasgupta, 1991; Hinton, 1992), and an artificial neuron is a processing element that takes an input vector, multiplies each element according to a series of weights, adds a bias term (possibly zero), and processes the result according to a transfer function (often a sigmoidal function) to produce a scalar output.

Neural networks are capable of performing well on noisy data, and on data where the underlying signal is unknown or difficult to differentiate from background data. However, the performance of the neural net relies on the availability of a large training set. A small training set can result in the net being unable to generalize to a larger data set. A lack of optimization techniques also requires testing many different sizes of nets, with different transfer functions, to obtain an optimal result.

The spectrograms used for training the neural networks are generated using the MATLAB signal processing toolbox, operating on the minimum number of data points that allow for visual differentiation of vocalization from noise (256 points sampled at 22050 Hz for the killer whale vocalizations), resulting in a 128 point complex vector.

Preliminary results suggest that fairly simple neural networks can be used to detect killer whale calls in the presence of boat noise as well as noise from breaking surface waves and that these results are in agreement with the results by Deecke and Ford (1999). However, further studies are required before real-time algorithms can be implemented into self-contained instruments.

5 CONCLUDING REMARKS

Recent concerns about the status of several whale species in Canadian waters is the driving force behind the work to establish an *in situ* and more long-term passive acoustic monitoring program on the west coast of Canada. Early results with the combined use of a towed array as well as land and offshore-based systems are encouraging. The use of passive acoustic monitoring is a non-intrusive approach to collecting the baseline data required for any successful recovery strategy.

The use of modern low-power electronics combined with powerful microcontrollers and readily available data storage are making long-term monitoring feasible. However,

the collection of ever-larger data sets is highlighting the need for improved software to deal with the data. It is no longer feasible to hire a large number of graduate students to sit and listen through the data in real time to recognize a particular whale species, dialect or individual animal.

A number of different approaches are presently being investigated (e.g., Harland and Armstrong, 2004; Johansson and White, 2004; van Ijsselmuide and Beerens, 2004)see other articles in this issue) ranging from parametric modeling and neural networks to different spectrogram correlation techniques. We have been looking into using neural network routines on spectrograms as a tool to extract killer whale vocalizations in large audio files, and for this specific use we find that this approach is showing some promise. However, it remains to be seen whether this approach will work for the more demanding task of distinguishing between all the different clan dialects among the resident killer whales on the west coast. We also intend to investigate the ability of this approach to detect the distinct, but much lower frequency, North Pacific right whale calls.

When suitable vocalization recognition algorithms have been developed, the natural next step will be to implement these into smart instruments to be deployed in the field. This will result in reduced demands on data storage and power and reduce the amount of post-deployment processing that will be required. We envision a time when these instruments will send back species and clan information directly to the scientists via satellite phone. There will clearly always be need for storage of some raw data for post-processing, but it should be possible to significantly reduce this in the future.

The instruments discussed here record only information about sound pressure level as a function of time. With a single pressure hydrophone there is limited information about the actual location of any vocalizing mammal. A next step for us will be to incorporate a second vertical hydrophone to the systems. A vertical array will make it possible, in combination with some sound propagation modeling, to estimate the range to low-frequency large mammal vocalizations (Laplanche et al., 2004)see other articles in this issue). This should be a straightforward addition to the PATC instruments because they are already set up for multi-channel inputs. Direction to the calling whales can be obtained by the use of a horizontal array of acoustic monitoring devices and relative travel time (see other articles in this issue). Another approach would be to make use of Directional Fixing And Ranging (DIFAR) hydrophones (McDonald, 2004). With a DIFAR hydrophone, three signals will have to be monitored and digitized in the monitoring device in addition to the compass heading of the hydrophone. The three channels would be two orthogonal components of particle velocity and a signal from a separate pressure hydrophone, which is required to resolve directional ambiguity.

6 ACKNOWLEDGEMENTS

7 REFERENCES

Barlow, J., and B.L. Taylor. 1998. Preliminary abundance of sperm whales in the northeastern temperate Pacific estimated from a combined visual and acoustic survey. Document number SC/50/CAWS20, Int. Whaling Commission, Cambridge.

Barrett-Lennard, L.G., J.K.B. Ford, and K.A. Heise. 1996. The mixed blessing of echolocation: differences in sonar use by fish-eating and mammal-eating killer whales. *Anim. Behav.* 51:553-565.

Benson A.J. and A.W. Trites. 2002. Ecological effects of regime shifts in the Bering Sea and eastern North Pacific Ocean. *Fish and fisheries.* 3:95-113.

Bigg, M.A., G.M. Ellis, J.K.B. Ford, and K.C. Balcomb, 1987, *Killer Whales: A Study of their Identification, Genealogoy and Natural History in British Columbia and Washington State.* Nanaimo, BC: Phantom Press.

Brownell R.L., P.J. Clapham, T. Miyashita, and T. Kasuya. 2001. Conservation status of North Pacific right whales. *J. Cetacean Res. Manage.* (Special issue) 2:269-286.

Clark, C.W., F. Borsani, and G. Notarbartolo-di-Sciara. 2002. Vocal activity of fin whales, Balaenoptera physalus, in the Ligurian Sea. *Mar Mamm. Sci.* 18:286-295.

Dasgupta, C. 1991. Neural networks: An overview. *J. Indian Inst. Sci.* 71: 491-502.

Deecke, V.B. and J.K.B. Ford. 1999. Quantifying complex patterns of bioacoustic variation: Use of neural network to compare killer whale (*Orcinus orca*) dialects. *J. Acoust. Soc. Am.* 105: 2499-2507.

Ford, J.K.B. 1989. Acoustic behaviour of resident killer whales (*Orcinus orca*) off Vancouver Island, British Columbia. *Can. J. Zool.* 67:725-745.

Ford, J.K.B. 1991. Vocal traditions among resident killer whales (*Orcinus orca*) in coastal waters of British Columbia. *Can. J. Zool.* 69:1454-1483.

Ford, J.K.B., G.M.Ellis, and K.C. Balcomb, 2000. Killer Whales: The Natural History and Genealogy of Orcinus orca in British Columbia and Washington, Second Edition, UBC Press, Vancouver and U. of Washington Press, Seattle, 102pp.

Fox, C.G., H. Matsumoto, and T.K.A. Lau. 2001. Monitoring Pacific Ocean seismicity from an autonomous hydrophone array. *J. Geophys. Res.* 106(B3):4183-4206.

Harland, E.J., and M. Armstrong, 2004, The real-time automated detection of calls of all cetacean species. Can. Acoust. Vol. 32, #2, June 2004.

Hinton, G.E. 1992. How neural networks learn from experience. *Sci. Am.* 268: 145-151.

Johansson, A.T., and P.R. White, 2004, A system for detection and identification of cetaceans by parametric modeling of whistle signals. Can. Acoust. Vol. 32, #2, June 2004.

Kenney, R.D. 2001. North Atlantic, North Pacific, and Southern right whales. pp. 806-813 in W.F. Perrin, B. Wursig, J.G.M. Thewissen, W.F. Perrin, B.G. Wursig (eds.), The Encyclopedia of Marine Mammals. 1414pp.

McDonald, M.A, 2004, Directional Fixing and Ranging (DIFAR) Hydrophone applications in whale research. Can. Acoust. Vol. 32, #2, June 2004.

McDonald, M.A., and S.E. Moore. 2002. Calls recorded from North Pacific right whales (Eubalaena japonica) in the eastern Bering Sea. *J. Cetacean Res. Manage.* 4(3):261-266.

Munger, L., M. McDonald, S. Moore, S. Wiggins, D.K. Mellinger, and J. Hildebrand, 2004, North Pacific right whales: long-term acoustic detection and localization in the Bering Sea. Can. Acoust. Vol. 32, #2, June 2004.

National Research Council, 2000, Marine Mammals and Low-Frequency Sound, National Academy Press, Washington, D.C., USA.

Roemmich D. and J. McGowan. 1995. Climate warming and the decline of zooplankton in the California current. *Science.* 267: 1324-1326.

Ross. P.S., G.M. Ellis, M.G. Ikonomou, L.G. Barrett-Lennard, and R.F. Addison, 2000. High PCB concentrations in free-ranging Pacific killer whales, *Orcinus orca*: effects of age, sex and dietary preference. *Mar Poll. Bulletin.* 40:504-515.

Rosenbaum, H.C., M.G. Egan, P.J. Clapham, R.L. Brownell, M.W. Brown. 2000. Utility of North Atlantic right whale museum specimens in assessing changes in genetic diversity. *Conserv. Biol.* 14:1837-1842.

Van Ijsselmuide, S.P., and S.P. Beerens, 2004, Detection and classification of MARINE MAMMALS using an LFAS system. Can. Acoust. Vol. 32, #2, June 2004.

Wiggins, S.M., M.A. McDonald, L.A. Munger, J.A. Hildebrand, and S.E. Moore, 2004, Waveguide propagation allows range estimates for North Pacific right whales in the Bering Sea, Can. Acoust. Vol. 32, #2, June 2004.

Worcester, P.F., K.R. Hardy, D. Horwitt, and D.A. Peckham. 1995. A deep ocean data recovery module. Proc. IEEE Oceans '95, pp. 1225-1231.

Wright, A.J. 2003. The Effects of a Tropical Storm on the Use of Clicks by Sperm Whales in the Northern Gulf of Mexico. Masters Thesis, Univ. Wales, Bangor.

A DESCRIPTION OF THE WORKSHOP DATASETS

Francine Desharnais[1], Marjo H. Laurinolli[2], Douglas J. Schillinger[3], Alex E. Hay[3]
[1]DRDC Atlantic, PO Box 1012, NS, Canada, B2Y 3Z7, francine.desharnais@drdc-rddc.gc.ca
[2]Department of Fisheries and Oceans, Bedford Institute of Oceanography, PO Box 1006, Dartmouth, NS B2Y 4A2
[3]Department of Oceanography, Dalhousie University, Halifax, NS, B3H 4J1

ABSTRACT

An acoustic dataset was provided to the participants of the *2003 Workshop on detection and localization of marine mammals using passive acoustics* (Dartmouth, NS, Canada, 19-21 November 2003). This document contains environmental and other technical information regarding supporting the acoustic dataset.

RÉSUMÉ

Un ensemble de données a été mis à la disposition des participants de l'*Atelier 2003 sur la détection et la localisation des mammifères marins à l'aide du repérage acoustique passif* (Dartmouth, NÉ, Canada, 19-21 novembre 2003). Ce document contient de l'information sur l'environnement, et autres détails techniques reliés aux données acoustiques.

1. INTRODUCTION

In support of the *2003 Workshop on detection and localization of marine mammals using passive acoustics* (Dartmouth, NS, Canada, 19-21 November 2003), an initial dataset was provided for the participants to test their algorithms. This DRDC/Dalhousie dataset was originally limited to North Atlantic right whale sounds recorded in the Bay of Fundy during 2000 and 2002. After initial distribution of this dataset, an additional dataset was made available by the Cornell Laboratory of Ornithology. This second dataset was made of similar recordings of right whales off the US east coast.

This document contains a brief description of the datasets as given to the participants. It also contains environmental information for the Bay of Fundy to support the data analysis efforts. The complete DRDC/Dalhousie package can be downloaded from the workshop web site [1]. The Cornell dataset is available on request (see Sec. 4).

2. DRDC/DALHOUSIE DATASET

2.1 The Location

This dataset was collected in the Bay of Fundy, Canada, in September 2002, using 5 Ocean Bottom Hydrophones (OBHs). Figure 1 shows the OBH locations, as well as water depth contours in 50-m increments. The OBH pattern is approximately 14.5 km wide – this distance was selected based on previous localization experiments in the Bay.

2.2 The Recording Devices

The OBHs (Fig. 2) are self-recording devices that are moored to the seabed. Each OBH system includes an omnidirectional hydrophone (OAS model E-2SD), a canister with electronics to amplify and record the hydrophone signal, and an acoustic release. The units recorded files of 19.58 minutes in length, with a variable spacing of approximately 10 seconds between files during which no data were recorded. The sampling frequency is programmed before each deployment, and affects the overall recording duration. The hydrophone is 0.9 m from the seabed when deployed.

Table 1 includes the location of the ship and water depth where the OBHs were deployed. The OBHs were deployed on 2 separate days, during different phases of the tide. Because the tidal currents in this area are strong, up to 1 m/s during the deployment, and the OBH descent rate is of similar magnitude, the actual position of the OBH on the bottom could differ from the ship's position by up to 100 m.

Table 1. Location of the ship during the deployment of the five OBHs in September 2002.

OBH	Deployment position		Water depth (m)
	Latitude (N)	Longitude (W)	
C	44.60202	66.49737	210
E	44.60197	66.31650	134
L	44.66320	66.40397	183
H	44.72943	66.31638	123
J	44.72910	66.49693	170
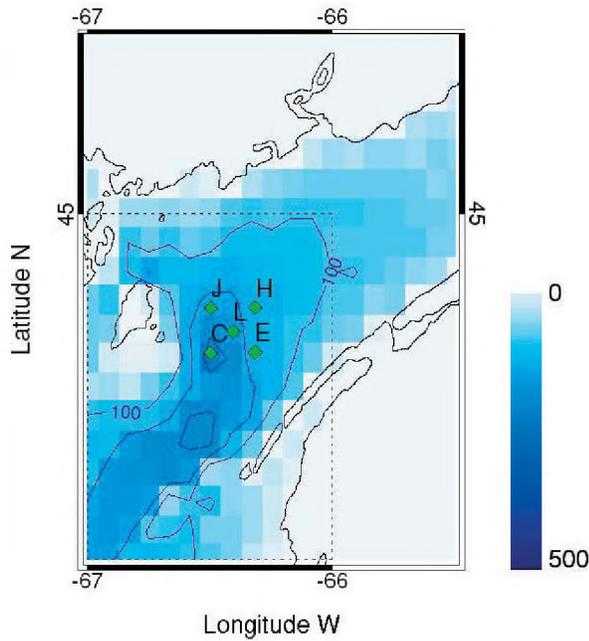
Figure 1. Locations of the five OBHs in the Bay of Fundy.

Figure 2. An OBH ready for deployment from the research ship.

To better refine their positions, a hydrophone was deployed from the ship at several locations around each OBH, while the ship was drifting. The output from this hydrophone was synchronized to the GPS clock. Each OBH is equipped with a pinger that sends out a series of acoustic signals at known times relative to the OBH clock, which is also synchronized to the GPS time base prior to deployment. The arrival times of these signals at the hydrophone give the travel time between the OBH and the hydrophone. Travel times recorded at each OBH (order of 40 per OBH) were reduced to 4 mean values at 4 mean locations. The exact OBH location is determined from the intersections of the four corresponding circles. The results for OBH C are shown in Figure 3. The OBH localization experiment led to the updated positions listed in Table 2, along with

estimates of position uncertainty (the standard deviation of the intersections among the equal travel time circles, Fig. 3).
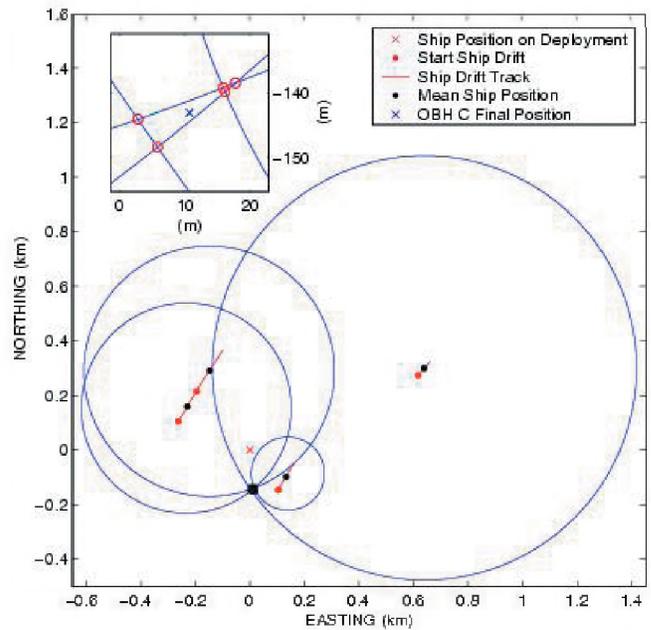
Figure 3. Results of localization experiment for OBH C. Red lines indicate the ship drift tracks, solid red circles the drift start, and solid black circles the mean ship position during the drift. The red 'x' indicates the ship's position when the OBH was launched. The solid black square is expanded in the inset and shows the intersections (open red circles) among the equal travel time circles (in blue) about the mean ship positions, and the best estimate position of the OBH on the seabed (blue x). Note the 150 m distance (approximate) between this final position and the position at launch.

Table 2. OBH updated positions and uncertainty.

OBH	Updated positions		Uncertainty	
	Latitude (°N)	Longitude (°W)	Northing (m)	Easting (m)
C	44.60073	66.49723	2.15	6.06
E	44.60237	66.31591	5.13	4.16
L	44.66203	66.40453	3.14	2.08
H	44.73051	66.31556	12.55	11.47
J	44.73038	66.49619	0.42	9.72

Each OBH has an accurate temperature-stabilized clock. Clock drift was estimated by measuring the offset relative to GPS time before the deployment and after the recovery. Table 3 lists the clock offset for each OBH deployed in September 2002, as measured before and after the deployment. A negative number means the OBH lags the GPS time signal. Linear drift is assumed over the deployment duration. The acoustic files provided within this data set have not been corrected for clock drift.

Table 3. Clock offsets for each OBH in the 2002 experiment.

OBH	Synchronization time	Clock offset [ms]
	Before deployment	
C	9 Sep 20:16	1.888
E	9 Sep 16:53	0.0366
L	10 Sep 15:55	-2.042
H	9 Sep 21:14	4.139
J	9 Sep 18:46	-1.985
	After recovery	
C	18 Sep 11:48	-3.142
E	18 Sep 12:57	1.809
L	18 Sep 16:24	0.337
H	18 Sep 14:02	-9.437
J	18 Sep 11:34	34.262

Table 4. XBT and CTD records close in time to the data samples included in the data set.

Data Type	Filename	Date [2002]	Time (UTC)	Lat (°N)	Long (°W)	Water Depth (m)
XBT	T7_00004	Sept 11	19:28	44 40.379	66 26.967	194
XBT	T7_00005	Sept 11	23:34	44 38.601	66 29.634	202
XBT	T7_00007	Sept 12	11:41	44 38.611	66 31.354	195
XBT	T7_00008	Sept 12	18:42	44 40.566	66 27.038	192
CTD	q269ctd3	Sept 11	19:48	44 40.411	66 27.263	194
CTD	q269ctd4	Sept 12	12:33	44 36.020	66 29.875	215.1
CTD	q269ctd5	Sept 12	18:41	44 40.479	66 27.384	192.7
CTD	q269ctd6	Sept 13	11:53	44 39.770	66 29.020	197.5
CTD	q269ctd7	Sept 13	12:14	44 40.000	66 28.063	193.8
CTD	q269ct18	Sept 14	9:17	44 40.484	66 26.115	190.1

To maximize the OBHs bottom time, a sampling frequency of 1200 Hz was selected in 2002, with a low-pass filter of 800 Hz. A filter roll-off frequency above the Nyquist frequency was selected to maximize localization opportunities for sounds in the upper end of the frequency range. Calibration files are included with the data set, but users should keep in mind the energy fold over. The amplifiers in all electronics packages were bench-calibrated, while the hydrophones were calibrated at the DRDC Atlantic calibration barge. The OBHs have a constant sensitivity over 50 to 700 Hz, and the sensitivity is independent of the instrument or the hydrophone, except for OBH-J, which recorded levels approximately 20 dB lower than expected. The data were digitized using a 12-bit A/D converter with ±5 V range. The levels can be calibrated using Equation (1):

$$L = P + 20 \cdot LOG_{10}[5/(2^{12}-1)] - G - S \qquad (1)$$

where L is the calibrated level in dB re 1 μPa^2, P is the initial level in dB re 1 V^2, G is the gain [59 dB], and S is the hydrophone sensitivity [-187 dB re 1 V/μPa].

2.3 The Environment

The bathymetry for the environment was extracted from the DBDB-V database (version 4.2) from the Naval Oceanographic Office, accessed through their web site [Idbms.navo.navy.mil/dbdbv/dbvquery.html]. The data were approved for release, and have a 5-min resolution in both latitude and longitude. The data set includes the file *dbdbv.txt*, which contains the depth in m over the 5-min grid from W to E, N to S. The file *dbdbv_coord.txt* contains the same information with the coordinates for each point.

Several expendable bathythermograph (XBTs) and conductivity-temperature-depth (CTD) profiles were taken during the experiment. Table 4 lists the profiles nearest in time to each data sample. Figure 4 shows the selected collection of temperature profiles. These data are included as ASCII files in the data set.

The sediment in the area is composed mainly of a layer of variable thickness LaHave clay over a thick layer of Scotian Shelf drift [2]. The postglacial silty sandy clay is loosely compacted, and is generally characterized by a low compressional sound speed. The physical properties of the LaHave clay found on the Scotian Shelf around Nova Scotia are typically those listed in Table 5 [3]. The Scotian Shelf drift is glacial till, a cohesive, poorly sorted sediment, generally containing angular fragments in the pebble/cobble/boulder range. It is predominantly sandy, but contains abundant silt and clay [3]. Its typical properties are also listed in Table 5; the compressional sound speed, attenuation and density were taken from [3], shear speed and attenuation were estimated from [4].

The thickness of the upper sediment layer was extracted from sub-bottom profiler data for the area. The file *sediment_thickness.dat* contains our estimates of sediment thickness as a function of latitude and longitude along specific tracks. Note that the thickness data in this file are for the upper clay sediment layer only. Figure 5 shows the sediment thickness data. Note that the layer thickness is expected to increase north of the OBHs. A poorly-defined reflector led to estimates at end "3" that are lower than expected. It is believed that gas within the surficial sediment layer led to the lower (and possibly inaccurate) estimates over this small area, though the data were insufficient to confirm this.

Table 5. Seabed parameters for upper sediments.

	LaHave clay	Scotian Shelf drift
Compressional sound speed (km/s)	1.261-1.49	1.745-1.92
Comp. attenuation (dB/m-kHz)	0.023, 0.056	0.0065
Density (g/cm^3)	1.5-1.54	2.1
Shear speed (km/s)	0.0	0.4-0.5
Shear attenuation (dB/m-kHz)	0.0	10.0
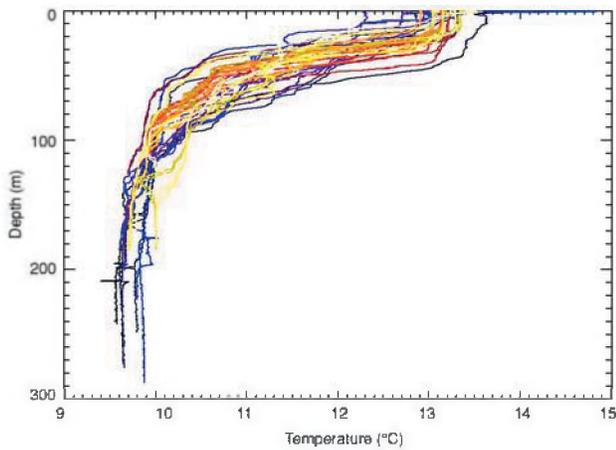
Figure 4. Temperature profiles from XBT and CTD records.
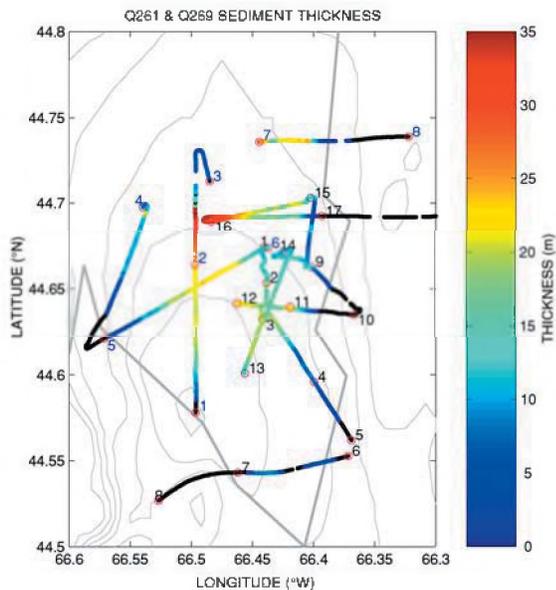
Figure 5. Upper sediment layer thickness.

2.4 The Acoustic Data

The raw OBH data were converted to audio files of WAV format using the READ_WAV function of IDL (Interactive Data Language, Research Systems Inc). The READ_WAV function does not normalize the data, which are integers within the file. Various sound types – vocalizations and "gunshot" sounds – were selected, and isolated into five 30-sec files (one for each OBH). The 30-sec period was selected so that the sound recorded on OBS-L was approximately in the middle of the file. For a given sound, all five files start at precisely the same time, except for the clock drift which has not been accounted for. Table 6 includes a list of all filenames, with a brief description of the sound types and the file start times. The filenames in Table 6 are to be completed with the letter describing each

OBH; for example, filename 'S013-1' indicates that there are 5 files named: 'X-S013-1.wav', where 'X' will be either C, E, L, H or J, depending on which OBH the file was recorded on.

For the purpose of testing detection algorithms, an 18-min long file (L-138.wav) from OBS-L is included. This file was selected as it was rich in sound occurrences and sound types. An equivalent file for the other OBHs can be made available on request.

Table 6. List of sounds in the data set.

	Filename	File start date [2002]	File start time	Sound type
1	S013-1	11 Sep	17:23:04	Gunshot
2	S035-2	12 Sep	0:34:23	Gunshot
3	S070-3	12 Sep	12:00:44	Gunshot
4	S093-4	12 Sep	19:24:09	Gunshot
5	S110-5	13 Sep	1:03:06	Gunshot
6	S092-7	12 Sep	19:08:57	Low-frequency call
7	S093-9	12 Sep	19:32:15	Low-frequency call
8	S131-10	13 Sep	7:47:15	Low-frequency call
9	S131-11	13 Sep	7:47:37	Low-frequency call
10	S131-12	13 Sep	8:02:16	Low-frequency call
11	S131-13	13 Sep	8:03:04	Low-frequency call
12	S134-6	13 Sep	8:50:56	Low-frequency call
13	S143-8	13 Sep	11:52:50	Low-frequency call
14	S209-14	14 Sep	9:29:52	Mid-frequency call
15	S210-15	14 Sep	9:34:30	Mid-frequency call
16	L-138	13 Sep	10:02:32	Multiple calls

2.5 The Calibration Dataset

A calibration dataset was included for the workshop, though it was taken from a different trial than the acoustic dataset described above. This trial occurred during August 2000, and only four OBHs were available for that trial. The positions of the four OBHs are listed in Table 7. No OBH localization experiments were done in 2000. The sampling frequency was 5000 Hz, with a low-pass filter of 1000 Hz.

Table 7. Location of the ship during the deployment of the four OBHs in August 2000.

OBH	Deployment position		Water depth (m)
	Latitude (N)	Longitude (W)	
B	44.7118	66.3494	131
C	44.6714	66.3753	165
D	44.6664	66.4331	190
E	44.7065	66.4083	166

The dataset was believed to be composed of transmissions of right whale calls played back with a

projector lowered into the water from a rigid-hull inflatable boat (RHIB). These transmissions were made by Susan Parks (Woods Hole Oceanographic Institute), from a sound file provided by Scott Kraus, of the New England Aquarium. The source was at 20-m depth. Table 8 has the GPS position of the RHIB at two times close to the transmissions. The source level was approximately 155-160 dB re 1 μPa.

Table 8. RHIB position during transmissions.

Time	Latitude (N)	Longitude (W)
17:19:08	44°41.744	66°22.635
17:28:06	44°41.677	66°22.566

The files from 2000 were 70-sec long. The selected files are listed in Table 9, along with the time of selected calls that were suggested for benchmarking purposes. The times listed in Table 9 are the arrival times (relative to the beginning of the file) of the calls on OBH-D. The filenames in Table 9 are to be completed with the letter describing each OBH; for example, filename "S282" indicates that there will be 4 files named: 'X-S282.wav', where 'X' will be either B, C, D or E, depending on which OBH the file was recorded on. The OBH clock drift was negligible in 2000, as the deployment times were short.

Table 9. List of sounds in the calibration data set.

	File name	File start date [2000]	File start time	Relative call time
1	S282	27 Aug	17:20:02	4
2	S282	27 Aug	17:21:44	62
3	S288	27 Aug	17:29:51	63
4	S289	27 Aug	17:31:02	53

It should be noted that the calls selected for the calibration datasets were believed to be from the playback transmissions. This assumption came from the initial localization exercise on these sounds, as the original playback tape was not available when the dataset was created. We cannot rule out that some of the calls from the calibration dataset files may be actual right whale vocalizations, possibly made in response to the playbacks.

The sound speed profile for the 2000 calibration data should be extracted from the XBTs or CTD of Table 10.

Table 10. XBT and CTD records close in time to the 2000 data.

Data Type	File name	Date [2000]	Time (UTC)	Lat (°N)	Long (°W)	Water Depth (m)
XBT	Q253_T 7_00006	Aug 27	14:39	44 42.46	66 20.96	134
XBT	Q253_T 7_00007	Aug 27	18:53	44 41.37	66 20.68	115
CTD	Q253018	Aug 27	06:38	44 40.82	66 24.21	176

3. THE CORNELL DATASET

The Cornell dataset is composed of two sub datasets of approximately 4 hours each: CCB 2001 and GSC 2000. These datasets are copyrighted by the Cornell Laboratory of Ornithology (all rights reserved).

The CCB 2001 data were collected during 2001 in Cape Cod Bay, Massachusetts. The original recordings were made with three retrievable bottom-mounted hydrophones (pop-ups), operating continuously from 8 March, 2001 to 2 April, 2001. Four hours of data were made available for the workshop.

The GSC 2000 data were recorded during 2000 in the Great South Channel (off Cape Cod Bay); these recordings were made for the International Fund for Animal Welfare. Six retrievable bottom-mounted hydrophones were used in 2000. The original recordings were made from 2300 h on 14 May to 2400 h on 14 June, 2000. Again, four hours of data were made available for the workshop.

The Cornell datasets consist of a collection of 5-min long sound files in the AIFF format. The data from all sensors were combined in these files; each sensor is on a separate channel, for a total of 3 channels in 2001, or 6 channels in 2000. The naming convention for the data files is: YYLLDD_ HHMMSS.aif, where YY is the year, LL the month, DD the date, HH the hour, MM the minute and SS the second; this identifies the start time of the data in each file (all times are Greenwich Mean Time). The sampling rate was 2000 Hz for both years.

The surface deployment locations of the hydrophones were based on shipboard GPS readings, as listed in Table 11.

Table 11. Hydrophone positions in 2000 and 2001.

Hydro-phone	2001 Lat [°N]	2001 Long [°W]	2000 Lat [°N]	2000 Long [°W]
1	41.931	70.166	41.844	69.301
2	41.957	70.166	41.858	69.269
3	41.947	70.193	41.830	69.269
4			41.935	69.094
5			41.910	69.095
6			41.923	69.062

4. CONTACTS

The DRDC/Dalhousie dataset is available on the Workshop web site, or by contacting:

Francine Desharnais
DRDC Atlantic
PO Box 1012
Dartmouth,NS,Canada
B2Y 3Z7

The Cornell dataset is available by contacting:

Bioacoustics Research Program
Christopher W. Clark, director
Cornell University Lab of Ornithology
159 Sapsucker Woods Rd.
Ithaca, NY 14850

REFERENCES

[1] The workshop data set can be downloaded from the web site:www.atlantic.drdc-rddc.gc.ca/newsevents/ workshop_01_e.shtml

[2] Surficial geology – Eastern Gulf of Maine and Bay of Fundy. Map 4011-G published by the Canadian Hydrographic Service, 1977.

[3] J.C. Osler, "A geo-acoustic and oceanographic description of several shallow water experimental sites on the Scotian Shelf", DREA Technical Memorandum 94/216, 1994.

[4] E.L. Hamilton, "Acoustic properties of sediments", Acoustics and Ocean Bottom, A. Lara-Sáenz, C. Ranz-Guerra and C. Carbó-Fité, eds., C.S.I.C., Madrid, pp. 3-58, 1987.

ACKNOWLEDGEMENTS

Detection And Classification Of Right Whale Calls Using An 'Edge' Detector Operating On A Smoothed Spectrogram

Douglas Gillespie

Song of the Whale Research Team,
International Fund for Animal Welfare
87-90 Albert Embankment
London, SE1 7UD
UK.

ABSTRACT

A detector has been developed which can reliably detect right whale calls and distinguish them from those of other marine mammals and industrial noise. Detection is a two stage process. In the first, the spectrogram is smoothed by convolving it with a Gaussian kernel and the 'outlines' of sounds are extracted using an edge detection algorithm. This allows a number of parameters to be measured for each sound, including duration, bandwidth and details of the frequency contour such as the positions of maximum and minimum frequency. In the second stage, these parameters are used in a classification function in order to determine which sounds are from right whales. The classifier has been tuned by comparing data from a period when large numbers of right whales were known to be in the vicinity of bottom mounted recorders with data collected on days when it was believed, based on ship and aerial surveys, that no right whales were present. Overall, the detection system is capable of picking out a high proportion of right whale calls logged by a human operator, while at the same time working at a false alarm rate of only one or two calls per day, even in the presence of background noise from humpback whales and seismic exploration. Although it is impossible to reduce the false alarm rate for individual calls to zero whilst still maintaining adequate efficiency, by requiring the detection of several calls within a set waiting time, it is possible to reduce false alarm rate to a negligible level.

SOMMAIRE

Un détecteur capable de déceler avec efficacité les vocalisations de baleines franches et de distinguer ces dernières des autres mammifères marins et du bruit industriel, a été développé. La détection se fait en deux étapes. En premier lieu, le spectrogramme est lissé par convolution avec un noyau de distribution gaussienne et le « contour » du son est extrait en utilisant un algorithme de détection d'arête. Pour chacun des sons, ceci permet de mesurer un certain nombre de paramètres incluant la durée, la largeur de bande et les détails sur les contours de fréquence, telle la position des fréquences maximale et minimale. Dans un second temps, ces paramètres sont utilisés lors de la fonction de classification, dans le but de déterminer quels sons proviennent des baleines franches. Le classificateur est optimisé en comparant les données découlant d'une période où il était établi qu'un large nombre de baleines franches était à proximité des appareils d'enregistrement ancrés sur le fond marin, avec des données recueillies les jours où il était plausible, basé sur des sondages maritimes et aériens, qu'aucune baleine franche n'était présente. Ce système de détection est capable de choisir une large proportion de vocalisations de baleines franches consignées par un opérateur, tout en opérant avec un taux de fausse alarme d'une ou deux vocalisations par jour et cela même en présence de bruit de fond provenant des baleines à bosses et de l'exploration sismique. Il est impossible de réduire à zéro le taux de fausse alarme pour chacune des vocalisations tout en maintenant une efficacité adéquate. Cependant, en imposant la détection de plusieurs vocalisations à l'intérieur d'un temps d'attente pré déterminé, il est possible de réduire le taux de fausse alarme à un niveau négligeable.

1 INTRODUCTION

The problems facing the North Atlantic right whale (*Eubalaena glacialis*) are well documented (IWC 2001). Although once widely distributed throughout the North Atlantic, only a remnant population of approximately 300 individuals survives. The known habitats of the North Atlantic right whale are all along the Eastern seaboard of the United States and Canada, the feeding and breeding grounds and the migratory routes between them coincide with major ship routes and important fishing grounds. It is believed that the North Atlantic right whale will become extinct within approximately 200 years (Caswell *et. al.*, 1999; Fujiwara and Caswell 2001) unless steps can be taken to reduce anthropogenic mortality due to collisions with ships and entanglement in fishing gear. Despite considerable efforts to better manage these activities in order to protect right whales, efforts are hampered by the lack of reliable, up to date, surveillance data in the areas where right whales are most at risk. Current survey methods rely primarily on the use of light aircraft which are expensive to run, require large numbers of personnel and cannot operate effectively during inclement weather. Passive acoustic monitoring has been proposed as a tool that could provide the information required for effective management action (IFAW, 2001).

Right whales make a variety of vocalizations (Clark, 1983). The work described in this paper is concerned solely with detection of one of the most commonly heard sounds from right whales, the frequency modulated (fm) up-sweep or 'contact call'. These are typically about a second long and sweep upwards in frequency between approximately 100 and 200 Hz. There is however considerable variability between individual upsweep calls, examples of which are shown in **Figure 1**. Vocalization rates of North Atlantic right whales are highly variable and individuals have been known to remain silent for several hours (Matthews *et. al.*, 2001).

Detection and accurate classification of right whale sounds was complicated by the wide variety of different sources of background noise present in the study area. As well as ships, either on passage or engaged in fishing activities, sounds from distant seismic exploration and other species of marine mammal were regularly heard on recordings made in areas frequented by right whales. The frequency range of many of these background noises overlaps that of right whale sounds. The most similar sounds to those of right whales encountered in this study, and the ones causing the greatest problem in classification, were found to be those of humpback whales (*Megaptera novaeangliae*). The problem was exacerbated by the fact that humpbacks are more numerous than right whales and also appear to vocalise more often and to be louder than right whales, so for every right whale vocalisation detected, it was necessary to avoid potential false detections from many thousands of humpback calls.
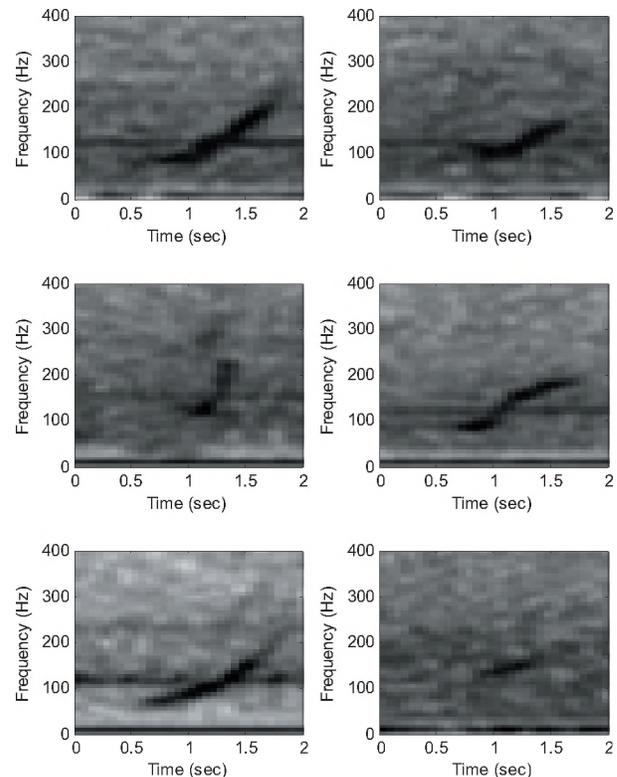
Figure 1. Example spectrograms of right whale upsweep calls selected by a human operator.

A good detection algorithm should be sensitive to up-sweeping sounds of varying sweep rate, but at the same time detect differences between the time-frequency contours of sounds from right whales and those from other sources. For the algorithm to be useful, it must work at a known efficiency and false alarm rate. Clearly it is desirable to maximise the former and minimise the latter. The coincidence in frequency of background noise sources means that simple energy detectors would have a high false alarm rate. Conversely, the considerable variation in sweep rate of the right whale calls is likely to cause detectors based on correlation techniques using a fixed template in either time or frequency to have low efficiency.

The algorithm described here used a detector which was sensitive to any type of sound rising above a predetermined threshold. The output of the detector was *edges* of the sound in time and frequency. From those edges, and the contour of maximum amplitude between them, a number of parameters were measured which were used in a statistical classifier to correctly identify right whale sounds.

2 METHODS

2.1 Data collection

Data were collected using bottom mounted recorders ('Pop-Ups') developed by Cornell University. They include a hydrophone, a microprocessor, a computer hard drive and a
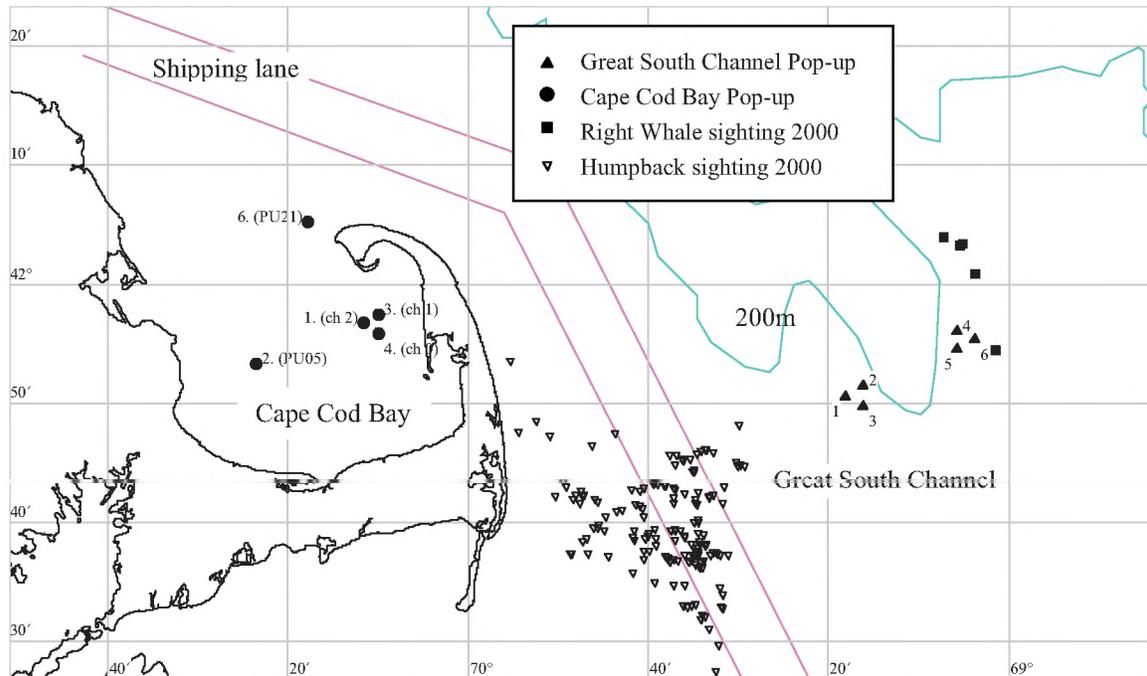
Figure 2. The locations of pop-up recorders in the Great South Channel 2000 and Cape Cod Bay 2001. Right Whale and humpback sightings form aerial surveys in 2000 are also shown. For clarity, sightings from 2001 in Cape Cod Bay have not been shown.

release mechanism. The units used in the study were capable of recording continuously to the hard disk for approximately 30 days. Each pop-up was moored two metres from the sea bottom to which it was attached using a disposable anchor (biodegradable sacks of gravel and sand). At the end of the recording period, an acoustically transmitted command from the surface caused the units to separate from their anchors and return to the surface. Data used in this study were collected using six pop-ups deployed in two groups of three at approximately 200m depth in the Great South Channel between 13 May and 12 June 2000 and three units deployed in a triangular configuration at the eastern side of Cape Cod Bay at 30m between 8 March and 2 April 2001 (**Figure 2**). Once on shore, data from the pop-ups were combined into six and three channel sound files for 2000 and 2001 data respectively. Synchronisation between channels was achieved by dropping light bulbs in 2000 (Marshall, 1993) and by playing FM sweeps from an underwater speaker in the vicinity of the 2001 pop-ups. All recordings were made at a sampling rate of 2000 Hz.

All recordings were browsed by a human operator viewing a spectrogram of the multi channel sound files and also listening on headphones whenever necessary. In 2000, the operator logged right whale calls on all six channels, but in 2001 only the loudest or the first occurrence of each sound was logged when it was observed on more than one channel.

Aerial surveys were flown over the Great South Channel in 2000 by NOAA Fisheries and over Cape Cod Bay in 2001 by the Center for Coastal Studies. The primary motivation behind these surveys was to provide information for ship and fisheries management and to collect identification

photos of individuals for long term population monitoring. The surveys were not concentrated on the precise locations of the pop-ups and could only take place during relatively calm conditions. The surveys did not therefore provide constant or even regular coverage over the pop-ups but could provide a general overview of the presence / absence of right whales during a pop-up deployment.

The 2000 aerial surveys found few right whales in the vicinity of the pop-ups. The only occasion on which right whales were spotted within ~10km of the pop-ups occurred on the morning of 26 May, 2000. The surveys did however spot a considerable number of right whales between 38 and 140 km NW of the pop-ups and also found large numbers of humpback whales ~20km to the SW, close to the shipping lanes to the East of Cape Cod. The Cape Cod Bay surveys indicted that large numbers of right whales were present in the Bay throughout the 2001 deployment period.

The detector and classifier were tuned using two days of data from each location, having no right whale calls and several thousand right whale calls respectively. Once tuning had been completed, the detector / classifier was used to analyse the entire data set.
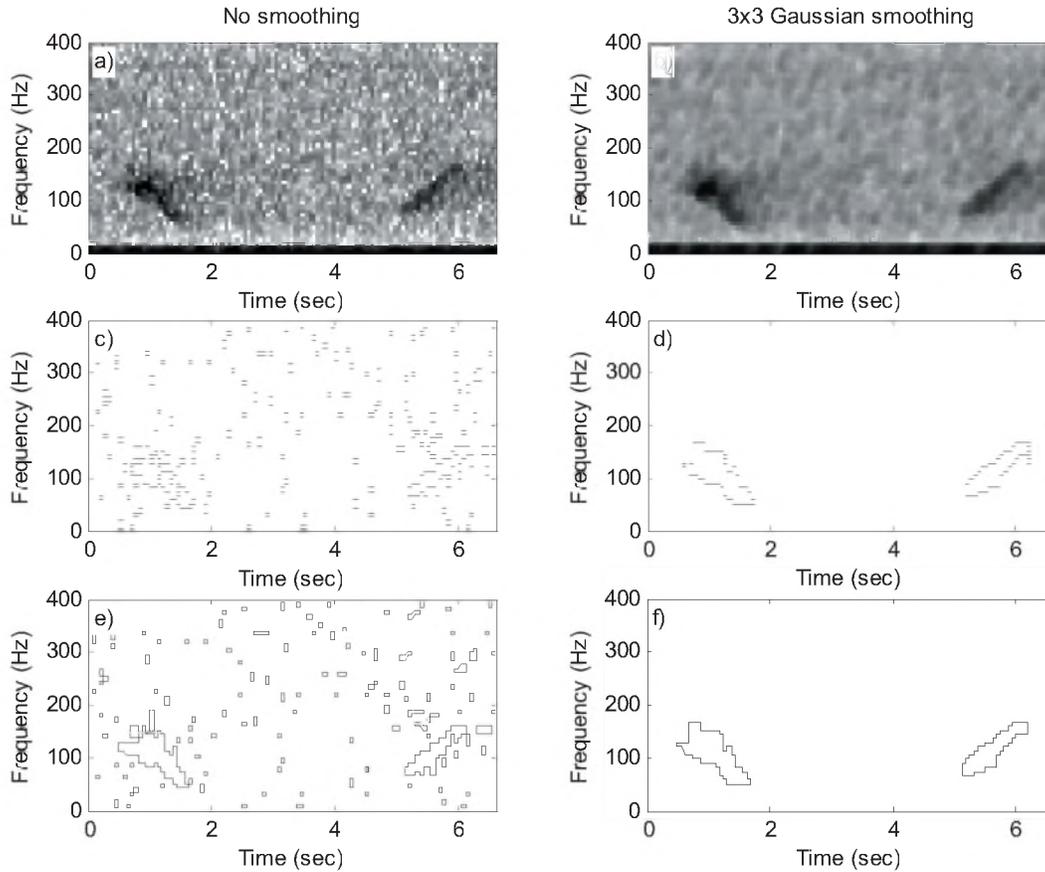
Figure 3. Sound Detection: a) Spectrogram containing two calls, one of which is a right whale upsweep; b) Spectrogram after Gaussian smoothing; c) Edge detection without Gaussian smoothing; d) Edge detection with Gaussian smoothing; e) sound outlines without Gaussian smoothing; f) sound outlines with Gaussian smoothing.

2.2   Sound Detection

The detection stage of the algorithm is not optimised to be any more sensitive to right whale sounds than any other type of sound. This lack of optimisation is important since, if the detector were optimised to only detect up-sweeping signals, it is possible that it may select up-sweeping parts of sounds having more complicated time-frequency contours and thereby create false detections.

Sounds were detected by searching for 'edges' in a spectrogram matrix and linking edges together to form the outlines of sounds. A number of edge detection algorithms of the type used in image analysis were tested, a simple threshold detector, which was found to have the best overall performance, in terms of efficiency verses false alarm rate, is described here.

Spectrogram smoothing

The power spectrogram S of the data was first calculated with a frame length of 256 samples (128ms) giving a frequency resolution of 7.8Hz. A Hanning window function

was used and successive frames overlapped by 131 samples to give a frame separation in time of 1/16 second. These values were chosen so that the spectrogram had approximately equal resolution in time and frequency, i.e. a typical right whale upsweep spanned 13 bins in frequency and 16 in time.

A common technique used in image edge detection is smoothing of the image matrix by convolving it with a Gaussian kernel (Embree and Kimble, 1991, Sonka *et. al.*, 1999). This has the beneficial effect of preventing edges breaking up into many parts, but also has the detrimental effect of reducing the resolution of the image if the smoothing kernel is too large. In this study a 3×3 smoothing kernel was used to compute the smoothed spectrogram $S' = S * G$, where

$$G = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}.$$

Edge detection

The edges of sounds were detected using a simple threshold detector where the signal at any point in the spectrogram $S'_{(t,f)}$ is compared with a background measurement $B_{(t,f)}$.

$B_{(t,f)}$ was continuously updated and computed independently for each frequency using

$$B_{(t,f)} = B_{(t-1,f)} + \left(\frac{S'_{(t,f)} - B_{(t-1,f)}}{\alpha} \right),$$

where α is the time constant for the background update, thereby allowing the detector to respond to changes in noise level such as would be caused by a passing ship.

Regions of the spectrogram were over threshold if

$$\frac{S'_{(t,f)}}{B_{(t-1,f)}} > Th,$$

where the threshold Th was set to 4 (6 dB).

Since the background measurement $B_{(t,f)}$ would tend to increase in the presence of any sound rising above the mean noise level, two different values of α were used in the background calculation – a high value (160), giving a long time constant (~10s) when the signal at any given frequency was above threshold and a lower value (16, giving a time constant of 1s) when the signal was below threshold.

Table 1. Data used in classifier training.

	Right Whale Sample	Non-Right Whale Sample
Location and Date	Cape Cod Bay, 16-17 March 2001	Great South Channel, 16-17 May 2000
Total Number of Detections	44,672	177,080
Total Number of Candidate upsweeps	6,294	19,098
Human Operator detected upsweeps	2077	0
Human upsweeps also found by detector	1879	0

Regions of the spectrogram in successive time slices which are adjacent or overlapping in frequency were then joined together to form a 'sound'. The edges of the sound were then the frequency at which the smoothed spectrogram rose above threshold and the frequency at which it fell back below threshold in each time frame. Rules were built into the joining process which allowed gaps of a single time frame containing no data above threshold within a sound. This helped prevent sounds which rose and fell above threshold from breaking into several parts.

The complete sound detection process is shown in **Figure 3** where the benefit of Gaussian smoothing (subplots b,d,f) compared to the raw spectrogram (subplots a,c,e) is clear.

2.3   Sound Parameterisation

Once a sound was detected, it was described by a relatively small number of parameters as listed below and shown in **Figure 4**.

1. Duration
2. Start Frequency[1]
3. Minimum Frequency[1]
4. Sweep Frequency (Maximum Frequency minus Minimum Frequency)[1]
5. Position of minimum frequency
6. Position of maximum frequency
7. Maximum instantaneous bandwidth (between the lower and upper frequency bounds of the sound outline)
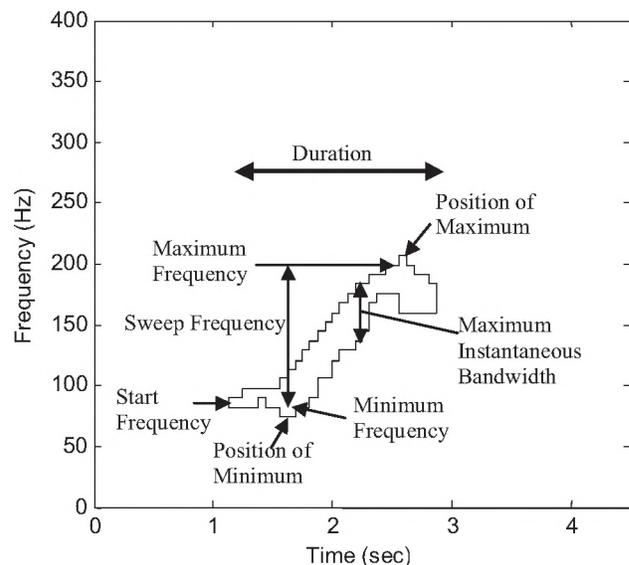
Figure 4. Parameters extracted from each sound for use in classification.

---

[1] Frequencies were taken as the frequency of maximum amplitude within each spectrogram time frame.

Canadian Acoustics / Acoustique canadienne
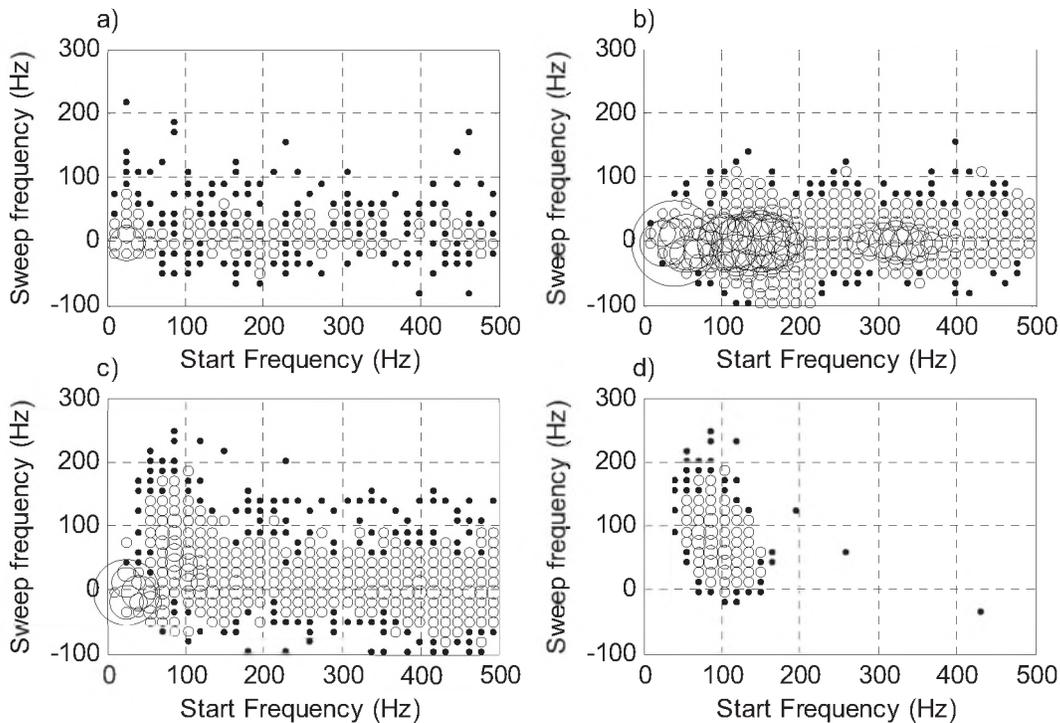
Figure 5. Sound start and sweep frequency distributions. a) A quiet day in the Great South Channel; b) a day in the Great South Channel when humpback whales are present; c) a day in Cape Cod Bay when both right whales and humpback whales are present; d) human operator selected calls from Cape Cod Bay. The area of each circle is proportional to the number of calls.

2.4   Sound Classification

Since there is a possibility that the human operator had missed genuine right whale calls and since the operator had only marked calls on a single channel, data from Cape Cod Bay, when many right whales were present, were not used to measure false alarm rate. Instead, the Cape Cod Bay data was only used to measure detection and classification efficiency and false alarm rate was measured using data from the Great South Channel from a period when no right whales were seen and none detected on the spectrograms by the operator. Two days of data from each location were used for classifier tuning.

The classifier was realised firstly by selecting right whale-like sounds using the selection criteria listed in Table 2. Final separation of right whale and non right whale sounds was then carried out with a multivariate discriminant analysis function utilising four of the parameters (start frequency, sweep frequency, duration and maximum instantaneous bandwidth) measured from the sounds which had passed the initial selection. Right whale sounds were then selected by choosing an appropriate cut on the first canonical variable resulting from the discriminant analysis.

The primary motivation for developing this algorithm is for use in dynamic ship management systems to mitigate against ship strikes. For such a management system to be

effective, the false alarm rate must be extremely low if it is to be accepted both by the industry and national / international regulators. Classifier tuning was carried out with this in mind by tuning it to detect the highest possible number of true right whale sounds for a maximum false alarm rate of $1 - 2$ calls per pop-up per day.

3   RESULTS

Table 1 shows the total numbers of sounds detected from the Great South Channel and Cape Cod Bay data on the two days used to tune the detector and classifier. Of 2077 right whale upsweeps detected by a human operator, the detector found 1897 (90%). Of these 1897 calls, the measured parameters showed that only 1753 (84%) swept up in frequency by at least 7 Hz (more than 1 frequency bin). It is believed that this is due to errors in the detection and parameterisation process occurring at low signal to noise ratio rather than errors on the part of the operator. 19,098 upsweeps were detected on the two days of Great South Channel data.

The numbers of detected sounds in the Great South Channel was considerably higher than that in Cape Cod Bay. This is primarily due to the presence of humpback whales, but airgun arrays used in seismic surveys are also audible on the Great South Channel recordings. **Figure 5** shows distributions of two of the parameters (start frequency and

Table 2. Selection criteria applied to calls before the multivariate discriminant analysis

	Loose	Medium	Tight
Minimum Duration	>=0.5 s	>=0.5s	>= 0.5 s
Maximum Duration			< 2 s
Sweep Frequency	>= 7 Hz	>= 23 Hz	>= 54 Hz
Start Frequency		50 ñ 160 Hz	50 ñ 160 Hz

sweep frequency) describing detected sounds for the different data sets. The upsweeping right whale calls are clearly visible on the distributions in subplots c and d. However, it is also clear from **Figure 5** that there is an overlap in the distribution of right whale call parameters and those of humpback whales.

Figure 6 shows plots of combined right whale detection and classification efficiency against the number of false alarms from non-right whale sounds for varying cuts on the canonical variable from the discriminant analysis. If no pre-selection of calls was made, the classifier performance was poor, particularly at low false alarm rates. If only sounds which started at between 50 and 160Hz, and swept through at least 23 Hz were selected, detector performance improved. It was found that the best detector performance at a false alarm rate of 1-2 calls per pop-up per day could be obtained by making the 'tight' pre-selections listed in Table 2. In this case, the algorithm correctly detected and classified approximately 60% of human detected calls.

Figure 7 shows the number of calls classified as right whale every 4 hours in the Great South Channel in 2000 using the detector operating point shown in **Figure 6**. Significant numbers of right whale calls were only detected between 0400 and 0800 UTC (0000 to 0400 local time) on 26 May. Obviously, it was impossible that the aerial surveys would have spotted them at that time, but right whales were seen at the locations shown in **Figure 2** later that morning, close to the three pop-ups on which the calls were detected.

4 DISCUSSION

The critical parameters describing any detection system are its efficiency and false alarm rate. False alarm rate was measured using data from a period when no right whales appeared to be present. If false alarm rate had been measured using data when right whales were present, counting calls found by the detector and not by humans as false alarms, it is possible that an artificially high false alarm rate would have resulted from the operator having

missed genuine calls. Since the number of calls produced can never be know precisely, efficiency can only be measured relative to that of a human. In spite of this, measurements of efficiency and false alarm rate made in this way can still be used to compare and optimise detectors.

The detector and classifier described here are capable of finding right whale sounds with a reasonable efficiency (~60%) while at the same time achieving a false alarm rate of 1-2 calls per pop-up per day, even in the presence of many tens of thousands of sounds from humpback whales and seismic exploration which are in the same frequency band as the right whale calls.

To obtain a reasonable detection efficiency it is not possible to reduce false alarm rate to zero. For many applications, such as management of shipping, detection of whales or
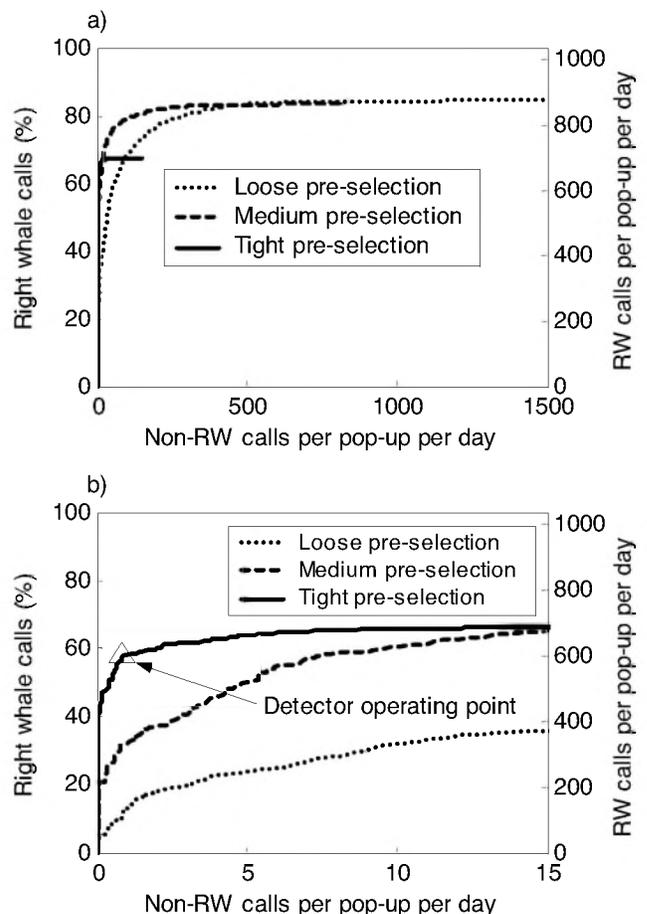
Figure 6. Efficiency / false alarm rate plots for different pre-selections of the calls (see table 2). Each plot shows the percentage and the number of detected right whale calls plotted against the number of false detections a) is scaled to show the full range of numbers of false detections, b) is scaled to show only the region of the curves in a) which are of interest, i.e. false alarm rates of only a few calls per pop-up per day.

groups of whales is of more interest than the individual calls. It is therefore possible to reduce false alarm rate to a negligible level by requiring a minimum number of sounds within a given waiting period. If false alarms are randomly distributed in time, then the number of calls which can be expected within a given waiting time is described by the Poisson distribution. As an example, if the false alarm rate were 10 calls per 24 hour period, then the probability of receiving 10 or more false calls in a one hour period is approximately 10^{-12}.

Even though human observers are more efficient than the automatic system, the automatic system has the advantage over the human that it is more objective and will not be affected by inter or intra observer variability. Although not impossible, the tasks of manually analysing many months of data from bottom mounted recorders is an onerous one and the bigger the dataset, the more likely it is to require more than one observer to analyse it. On the other hand, the adaptability of human observers may make them less likely

to become confused by an unexpected sound which was not present in the data used to tune the detector and classifier.

The current classification system analyses each sound in isolation. It does not use other available information such as the rate of call production or the presence of other types of sound. If this 'contextual' information were used, it should be possible to make the classifier adapt, using stricter criteria when sources such as humpback whales are known to be present and less strict criteria on quiet days.

The current classification system relies on a multivariate discriminant analysis. Such analysis assumes that the parameter distributions are Gaussian and is only optimal if this is the case. A Neural Network using the same parameters (start frequency, sweep frequency, etc.) describing the sounds as its input may give better performance. A preliminary investigation showed this not to be the case, although further studies are planned for the
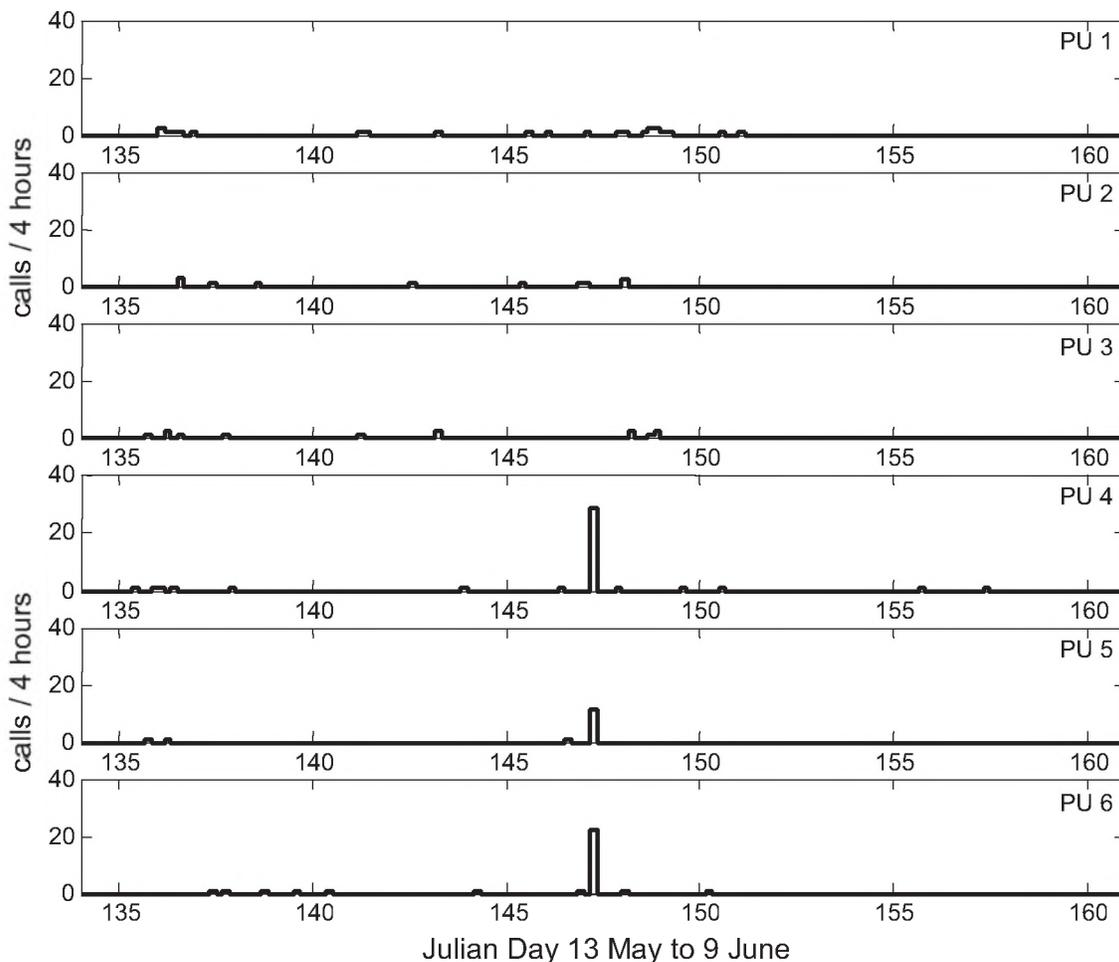
Figure 7. 4 hour call counts from pop-ups 1 to 6 deployed in the Great South Channel in 2000.

future.

Although the classifier described here has been tuned to detect only a certain type of right whale sound, the detector is designed to pick out and measure parameters of any sounds rising above threshold (section 2.2; **Figure 3**). Developing additional classifiers for other types of right whale sound, or sounds from other species, should be a relatively straight forward task if sufficient training data can be obtained. When using the detection algorithm with other sounds, careful consideration should be given to the FFT length and time frame overlap in order to optimise the quality of data provided to the classifier. Classifiers for more complex sound types, which may have a number of inflexions in their time frequency contour may also require the extraction of a different set of descriptive parameters for each sound.

5 ACKNOWLEDGEMENTS

6 REFERENCES

Caswell, H., Fujiwara, M. and Brault, S. 1999. Declining survival probability threatens the North Atlantic right whale. *Proc. Natl. Acad. Sci* 96: 3308-3313.

Clark, C.W. 1983. Acoustic communication and behavior of the southern right whale. In: R.S. Payne (ed.), *Behavior and Communication of Whales. Westview Press*: Boulder, CO. pp. 163-198.

Embree, P., and Kimble, B., 1991. C Language Algorithms for Digital Signal Processing. *Prentice Hall*, NJ. pp386-388

Fujiwara, M. and Caswell, H. 2001. Demography of the endangered North Atlantic right whale. *Nature* 414: 537-541.

IFAW, 2001. Report of the Workshop on Right Whale Acoustics: Practical Applications in Conservation. (eds. Gillespie, D. and Leaper, R.) *International Fund for Animal Welfare*, Yarmouth Port, MA 02675, USA. 23 Pages.

IWC. 2001. Report of the workshop on status and trends of western north Atlantic right whales. *J Cet. Res. Manage Special Issue* 2: 61-87.

William J. Marshall, Jr. 1993, Acoustic signatures of imploding underwater light bulbs, JASA 94:3, p. 1845

Matthews, J., Brown, S., Gillespie, D., Johnson, M., McLanaghan, R., Moscrop, A., Nowacek, D., Leaper, R., Lewis, T., and Tyack, P.. 2001. Vocalisation rates of the North Atlantic Right Whale. . *J. Cet. Res. Manage* 3(3):271-282.

Sonka, M, Hlavac, V., and Boyle, R., 1999. Image Processing, Analysis and Machine Vision. 2nd ed. PWS Publishing, CA. p445.

Detection and Classification of North Atlantic Right Whales in the Bay of Fundy using Independent Component Analysis

Brian R. La Cour and Michael A. Linford

Applied Research Laboratories, The University of Texas at Austin

P.O. Box 8029, Austin, TX 78713, E-mail: `blacour@arlut.utexas.edu`

ABSTRACT

A novel method of detection and classification for marine mammals is presented which uses techniques from independent component analysis to solve the blind source separation problem for North Atlantic right whales (*Eubalaena glacialis*). Using the fundamentally non-Gaussian nature of marine mammal vocalizations and data collected on multiple hydrophones, we are able to separate right whale source spectra, up to an unknown scale, from ambient noise. This technique assumes that the array data is a linear combination of non-Gaussian source signals but does not require specific knowledge of the array geometry. A detection algorithm which separates right whale vocalizations from ambient background using a Kolmogorov-Smirnov test statistic is presented and tested on data collected in the Bay of Fundy. The performance of the detector was found to be such that it was possible to achieve a probability of detection of about three-fourths with a false alarm probability of about one-third. Independent component analysis was found to provide little improvement over standard principle component analysis, which was used as preprocessing step.

RÉSUMÉ

Une nouvelle méthode de détection et de classification pour les mammifères marins est présentée. Elle utilise des techniques d'analyse par composantes indépendantes pour résoudre des problèmes de séparation aveugle de sources pour des baleines franches de l'Atlantique Nord (*Eubalaena glacialis*). En se basant sur la nature non gaussienne des vocalisations des mammifères marins et sur les données recueillies par un ensemble d'hydrophones, nous avons été capables de séparer les spectres de baleines franches, jusqu'à une échelle inconnue, du bruit ambiant. Cette technique suppose que l'ensemble des données est une combinaison linéaire des signaux sources non gaussiens, mais ne requiert pas de connaissance particulière sur la géométrie de l'ensemble des hydrophones. Un algorithme de détection permettant de séparer les vocalisations de baleines franches du bruit ambiant en utilisant un test statistique Kolmogorov-Smirnov est présenté et testé sur des données recueillies dans la Baie de Fundy. La performance du détecteur était telle qu'il a été possible de réaliser une probabilité de détection d'environ trois quarts, avec une probabilité de fausse alarme d'environ un tiers. L'analyse par composantes indépendantes n'a donné que des améliorations mineures comparé á l'analyse par composantes principales standard qui a été utilisée comme étape de pré-traitement.

1 INTRODUCTION

Passive acoustic detection of cetaceans has become an area of great interest in recent years due in part to the need to mitigate any possible impact due to shipping and naval training exercises on local populations. Visual observations, the traditional method of detection, are limited in several ways. They are restricted to daylight observations, require human observers, are limited in detection range, and can detect only surfacing animals. Passive acoustic detection overcomes these particular limitations, but brings also a new set of challenges. Marine mammals in the observation area must vocalize to be detected, and those vocalizations must be detectable and distinguishable from the multitude of competing background sound sources. In addition, it may be important to distinguish between different types of cetaceans based on their calls. A key advantage to using passive acoustics is the potential to perform 24-hour, real-time automated detection. Since, for purposes of impact mitigation, important time and distance scales for detection and localization are on the order of hours and miles, respectively, the demands placed upon an automated system are significantly reduced from that of more traditional anti-submarine applications.

In this paper we discuss a method for performing passive acoustic automated detection of marine mammals based upon the use of independent component analysis (ICA). ICA is a statistical analysis tool used to solve the blind source separation problem, wherein simultaneous recordings of multiple sound sources are

used to separate out the individual sound sources. The approach, then, is to apply ICA to segments of the recorded acoustic time series and separate out marine mammal calls from the ambient background. The separated signals may then be used to perform classification in an automated or human-directed classifier.

The algorithm presented here is applied to and tested on data gathered in Canada's Bay of Fundy that contains several calls from North Atlantic right whales (*Eubalaena glacialis*).[1] The focus on this species is motivated by its rapid decline in recent years [1], with over a third of deaths now attributable to ship collisions [2]. The data was collected in September of 2002, a time when right whales come to the bay in great numbers to feed themselves and their newborn calves, and was recorded on several bottom-mounted hydrophones.

The organization of this paper is as follows. In Sec. 2 we provide a brief introduction to Independent Component Analysis, with particular emphasis on its application to marine mammal detection. Sec. 3 contains a description of the detection algorithm used, which is based on the observed non-Gaussianity in the statistics of right whale calls. The application to right whale data from the Bay of Fundy is analyzed in Sec. 4, while a discussion of results and conclusions are given in Sec. 5.

2 INDEPENDENT COMPONENT ANALYSIS

2.1 Description

Independent component analysis (ICA) is a method for solving the blind source separation problem [3]. The basic model assumed by ICA is one in which the data, $\mathbf{x}(t) = [x_1(t), \ldots, x_M(t)]^T$, on M receivers at time t is a linear combination of P sources, $\mathbf{s}(t) = [s_1(t), \ldots, s_P(t)]^T$, plus ambient noise, $\mathbf{n}(t)$. The sources are assumed to be realizations of mutually independent, stationary stochastic processes in which at most one of the sources has a Gaussian marginal distribution [4]. Denoting by $\mathbf{A}$ the linear mixing matrix, we therefore have the following data model.

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

The goal of ICA is to find a suitable unmixing matrix, $\mathbf{W}$, such that $\mathbf{y} = \mathbf{W}\mathbf{x}$ is an approximation to $\mathbf{s}$. Due to the nature of the problem and the type of solution offered by ICA, there is a fundamental ambiguity such that $\mathbf{y}$ may differ from $\mathbf{s}$ by an arbitrary permutation and scaling of its rows.

---

[1]This data was provided courtesy of Defense Research and Development Canada—Atlantic and Dalhousie University.

What allows ICA to perform this inversion is the assumed non-Gaussianity of the sources. This may be understood as follows. For a given choice of $\mathbf{W}$, each element of $\mathbf{y}$ will be a linear combination of the sources. In general, this will result in a $\mathbf{y}$ which appears more Gaussian, due to the central limit theorem [5]. If $\mathbf{W}$ is suitably chosen, however, then each element of $\mathbf{y}$ will correspond to only one source, up to an arbitrary scaling, and thus will appear non-Gaussian. In this manner, non-Gaussianity may be used as a metric of independence, and the independent components estimated by $\mathbf{y}$ are maximally independent linear combinations of the original data.

It may be noted that ICA resembles Principle Component Analysis (PCA), which produces a linear combination of the data that is uncorrelated. In PCA the unmixing matrix $\mathbf{V}$ is given in terms of the eigenvalue decomposition of the data covariance matrix $\mathbf{C}$. If $\mathbf{U}$ is the matrix of eigenvectors of $\mathbf{C}$ and $\mathbf{D} = \mathbf{U}^H\mathbf{C}\mathbf{U}$ is the diagonal matrix of corresponding eigenvalues, then $\mathbf{V} = \mathbf{D}^{-1/2}\mathbf{U}^H$ is the PCA unmixing matrix. In applying ICA, PCA is often used as a preprocessing step. When this is done, the ICA algorithm is applied to the transformed variable, $\mathbf{z} = \mathbf{V}\mathbf{x}$, to yield the unmixing matrix $\mathbf{Q}$. The final unmixing matrix, applied to $\mathbf{x}$, is therefore $\mathbf{W} = \mathbf{Q}\mathbf{V}$.

2.2 Mutual Information Approach

Bell and Sejnowski [6] have suggested a method for performing ICA based on mutual information. More traditional ICA methods have relied on kurtosis as a measure of non-Gaussianity and, by inference, statistical independence [7]. Mutual information, on the other hand, provides a direct measure of statistical independence between a joint probability density function (PDF) and the product of its marginals.

In the approach of Bell and Sejnowski, the ICA process is viewed as a neural network with input $\mathbf{x}$, output $\mathbf{y}$, and nodal weights $\mathbf{W}$. The goal, then, is to minimize the mutual information between the input and output of the network. This mutual information may be written

$$I(\mathbf{x}, \mathbf{y}) = H(\mathbf{y}) - H(\mathbf{y}|\mathbf{x}), \tag{2}$$

where $H(\mathbf{y})$ is the entropy of $\mathbf{y}$ and $H(\mathbf{y}|\mathbf{x})$ is the entropy of $\mathbf{y}$ relative to $\mathbf{x}$ [8]. Since $\mathbf{y}$ is a deterministic function of $\mathbf{x}$, the term $H(\mathbf{y}|\mathbf{x})$ is a constant (i.e., $-\infty$), independent of $\mathbf{W}$. Thus, minimization of the mutual information is equivalent to minimizing the entropy of the output.

To achieve greater sensitivity, the output is transformed via a sigmoidal function $g(\cdot)$ so that the output of the network is $\mathbf{y} = g(\mathbf{u})$, where $\mathbf{u} = \mathbf{W}\mathbf{x}$. The choice of this function is arbitrary, although the Bell

and Sejnowski suggest using the logistic transfer function, $g(\mathbf{u}) = 1/(1 + e^{-\mathbf{u}})$, a common choice for neural networks, and we adopt this choice in our work. Ideally, the transfer function would be formed from the cumulative distribution function (CDF) of the output $\mathbf{u}$.

Extremizing the mutual information with respect to $\mathbf{W}$ leads to the following gradient method: Starting with an initial value for $\mathbf{W}$, the matrix is updated according to the scheme $\mathbf{W} \mapsto \mathbf{W} + \Delta\mathbf{W}$, where

$$\Delta\mathbf{W} = \alpha \left[(\mathbf{W}^T)^{-1} + (1 - 2\mathbf{u})\mathbf{x}^T \right] \qquad (3)$$

and α is the learning rate. In our work, the initial value of $\mathbf{W}$ was taken to be the identity matrix and α was taken to be $0.001/n$, where n is the current iteration number.

2.3 Applicability to Marine Mammal Detection

To apply ICA to marine mammal detection, is it necessary to demonstrate the validity of two key model assumptions: statistical assumptions regarding the source signals and propagation assumptions under the linear mixing model.

For propagation we assume a linear medium such that the received signal is a sum of attenuated and time delayed sources. The receivers are assumed to be horizontally distributed so that time delays may exist between receivers from sources are different bearings. Thus,

$$x_m(t) = \sum_{p=1}^{P} A_{mp} s_p(t - \tau_{mp}) + n_m(t), \qquad (4)$$

where τ_{mp} is the time delay from source p to receiver m and $n_m(t)$ is the ambient noise in receiver m. (For simplicity, we ignore the presence of multipaths.) We assume that a dominant source signal is present which is spatially localized in bearing so that τ_{mp} is effectively independent of p. (Physically, this dominant source may correspond to an individual whale or an entire pod.) Then, using a correlation technique described in Sec. 3, the received data may be transformed by the estimated relative delay time, $\hat{\tau}_m$, so that $x'_m(t) = x_m(t + \hat{\tau}_m)$ may be expected to fit the ICA linear model.

The fundamental statistical assumption required for our work is that marine mammal calls can be distinguished from ambient background signals by their non-Gaussian statistics. Such behavior is well known in the field of human speech processing, were super-Gaussian statistics dominate.[6] In the case of right whales, this question may be answered by an examination of actual whale call recordings. In Fig. 1 we
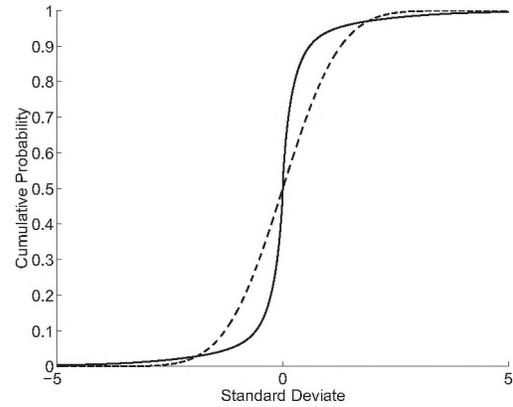
Figure 1: Plot of the cumulative distribution of right whale call statistics (solid line) versus that of a standard normal distribution (dashed line). The distribution is symmetric, but the sharp slope at the center indicates a high value of kurtosis.

have plotted the empirical CDF for a 110-sec recording of a North Atlantic right whale.[2] For comparison, the CDF for a standard normal distribution is included. It is clear from the figure that the statistics of the right whale calls exhibit a large, positive kurtosis excess (about 22.3) and thus may be considered super-Gaussian. Such statistical behavior has been found to be typical of human speech, as well as many other natural and unnatural sound sources, so this result is not at all surprising. On the other hand, and for this very reason, statistics alone cannot serve as a means of classification, though it may serve a role in detection and preprocessing.

3 DETECTION ALGORITHM DESCRIPTION

The detection algorithm may be described as a three step process. First, a sliding window along the time axis chooses a segment of the time series data to be analyzed. Next, independent component analysis is performed to extract the dominant non-Gaussian signal. Finally, a test statistic is computed from the dominant component and compared to a threshold to determine whether a detection is called. As a preprocessing step prior to selecting individual time segments, the time axes of the different receiver time series are aligned via cross-correlation under the assumption that all sound sources are co-located in bearing. Although not required by ICA, such an alignment is needed to ensure that a given call appears in the same time window for all hydrophones.

Time aligning the different receivers is equivalent to

---

[2]Data provided by Susan Parks of the Woods Hole Oceanographic Institute from recordings made by Scott Kraus of the New England Aquarium.

source localization in bearing. Since the expected source signals are transient and broadband, the usual narrowband subspace methods for bearing estimation are rejected in favor of an approach using cross-correlations. In this approach, a reference receiver, labeled m_0, is chosen, and cross-correlations are computed for each receiver paired with the selected reference. Rather than search for peaks in the receiver cross-correlations, a set of physically realizable time delays is computed based on the geometry of the receiver array and a hypothesized source bearing.

The delays are computed under the assumption of a constant sound speed, c, and direct path propagation. To do this, an asymptotic result is used. If the source range is much larger than the spatial extent of the array, then the relative time delay for receiver m is given approximately by

$$\tau_m = -\left(\Delta x_m \cos\theta + \Delta y_m \sin\theta\right)/c, \qquad (5)$$

where $\Delta x_m = x_m - x_{m_0}$, $\Delta y_m = y_m - y_{m_0}$, and θ is the hypothesized source angle. At short ranges, propagation in the vertical direction will cause this approximation to underestimate the magnitude of the time delays.

Each value of θ gives a different set of delays. If we consider the cross-correlation $\rho_m(\tau)$ between receiver m and the reference, then a measure of goodness of fit would be the sum of $|\rho_m(\tau_m)|^2$ over m; i.e., the beam intensity in the θ direction. For perfect alignment of data that differ only by a translation in time, this quantity will have a peak value of M. Thus, maximizing over θ gives an estimate of the source bearing and, with it, the delay estimates, $\hat{\tau}_m$, needed to perform time alignment.

Having aligned the receiver data, we next run a window of fixed width w and offset $t_n = n\Delta t$ such that $0 \leq t_n \leq T - w$, where T is the maximum time recorded. In general, Δt may be taken to be smaller than w. In our work we found that $w = 2$ sec, which adequately bounds the duration of a typical right whale call, and $\Delta t = w/2$ appear to work quite well.

The ICA algorithm described in Sec. 2 is applied to the data after preprocessing through PCA. The number of independent components to be extracted is variable (up to the number of receivers), but we have found that for detection purposes it is best to extract only two independent components (ICs). These are ranked according to their absolute kurtosis value, and the component with the largest such value is taken to be the one that would contain the source signal.

Detection is based on the non-Gaussianity of the first independent component extracted. For a detection statistic, we chose the Kolmogorov-Smirnov (KS) statistic, which measures the largest difference between
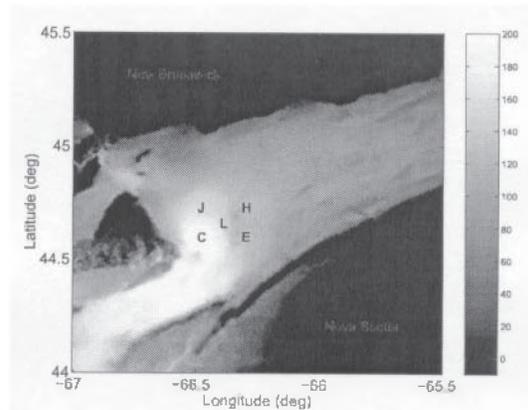
Figure 2: Bathymetry for the Bay of Fundy at 0.1-minute resolution. New Brunswick is at the top of the figure; Nova Scotia appears on the bottom right. The depth is in meters. The positions of the five OBHs are indicated in the figure, at the left and center of each letter. (Bathymetric data courtesy of the Naval Oceanographic Office)

the empirical CDF of the first IC, suitably standardized, and a standard normal distribution [5]. The KS statistic has the desirable property that, for large samples, its distribution is independent of that of the underlying data. This allows one to set a detection threshold for a desired probability of false alarm (PFA). For a standard PFA of 5%, the critical value of the KS statistic is about 1.36.

4 APPLICATION TO RIGHT WHALE DATA

4.1 Data Description

The detection algorithm was applied to data collected from the Bay of Fundy in September of 2002 on five Ocean Bottom Hydrophones (OBHs). The OBHs (OAS model E-2SD) are omnidirectional hydrophones that are moored about 0.9 m from the seafloor when normally deployed. The positions of the OBHs are shown in Fig. 2 in relation to the local bathymetry.

The locally measured sound speed profile (SSP) is shown in Fig. 3. The strongly downward refracting profile together with a shallow bottom implies that sound propagation will tend to be limited in range due to multiple bottom interactions. From reciprocity we may also observe that very little sound from the surface propagates to the bottom via a direct path.

The data set itself consists of 15 segments, each about 30 sec in length. All contain vocalizations from one or more right whales in the area. Of the 15 segments, the first five are so-called "gunshot" calls, which appear as broadband impulsive transients. The remaining segments contain low (segments 6–13) and mid-
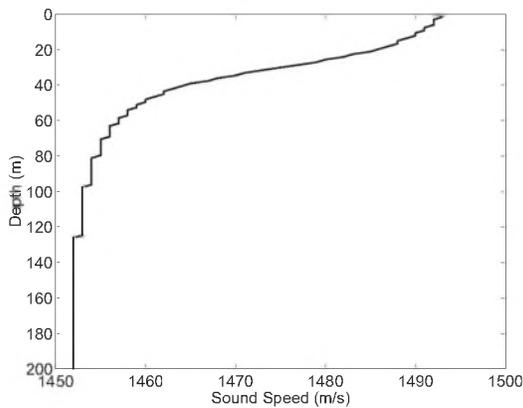
Figure 3: Plot of the sound speed profile in the Bay of Fundy during September 2002, as determined from local Expendable Bathythermograph (XBT) and Conductivity-Temperature-Depth (CTD) measurements . A mixed layer of some 40 m is evident.
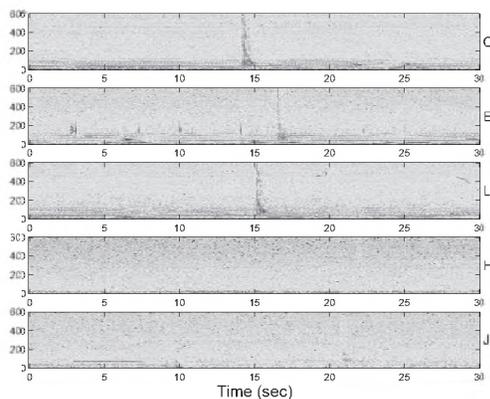
Figure 4: Plot of the spectrograms for the first data segment on all five OBHs. The gunshot is visible at about 15 sec on OBH L. OBH H and J did not register the gunshot well. The frequency, in Hz, is plotted along the vertical axis. The magnitude of the spectrogram is given in arbitrary units on a logarithmic scale.

frequency (segments 14 & 15) upsweeps. Fig. 4 shows the spectrograms for the first time segment, illustrating a gunshot call. Note that the data on OBHs H and J is very poor.[3]

4.2 Detection Results

Due to the nature of the data collected, ground truth is not available for a true assessment of detector performance. In lieu of this, detections based on human observations were used as truth to baseline the performance of the detector. For each of the 15 data segments, the start and stop times of each right whale call were determined by aural clues and visual inspec-

[3]OBH J was know to suffer from hardware problems. The low amplitude in OBH H may be due to its location relative to the right whale pod, which was known to be south of the array.
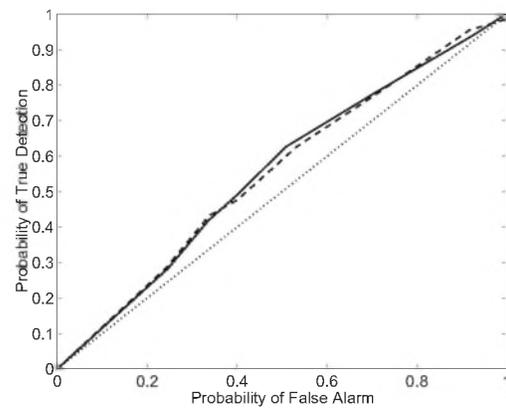
Figure 5: ROC curve comparing detection performance of PCA (solid curve) versus ICA (dashed curve). The dotted diagonal line represents random chance.

tion of the spectrograms. All such times were based relative to OBH L, which lay in the center of the array.

A called detection was considered correct (true positive) if the time window $[t_n, t_n + w]$ in which the KS statistic exceeded the threshold intersected with the truthed call interval described above. The probability of detection (PD) was then defined as the ratio of the number of true positives to the total number of true positives and true negatives. Similarly, the probability of false alarm (PFA) was defined as the ratio of the number of false positives to the total number of false positives and false negatives. By sweeping through a range of threshold values, a receiver operating characteristics (ROC) curve of PD versus PFA could then be generated. It is this ROC curve that we use as our metric of performance.

In Fig. 5 we compare the detection performance of ICA, with PCA as a preprocessing step, versus PCA alone. The results suggest that there is very little improvement gained by ICA over PCA, with the two ROC curves being almost identical. Clearly, the bulk of the work done in separating independent signals is done simply by linearly transforming the data so that the phones are uncorrelated. Since the PCA algorithm is much faster (about 100 times faster than ICA), this suggests that it may be the better choice for a real-time system.

Detection is performed by computing the KS statistic on the first independent (or principle) component estimated. If we compare the discriminating power of this first component to the second, we see from Fig. 6 that detection performance is severely degraded by using the latter. This suggests that the signal contained in the first component really does have information content useful for detection.

We have developed a method for time aligning data from spatially separated receivers. As shown in Fig.
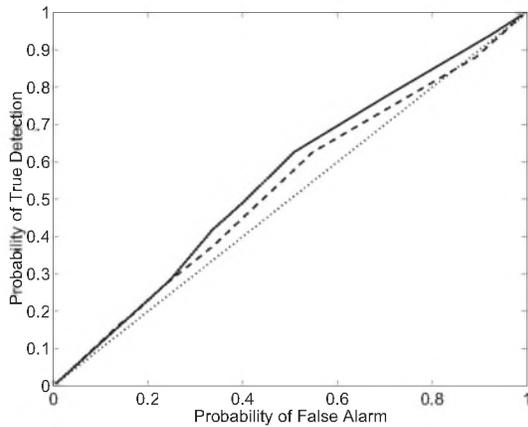
Figure 6: ROC curve comparing detection performance using the first (solid curve) and second (dashed curve) independent components.
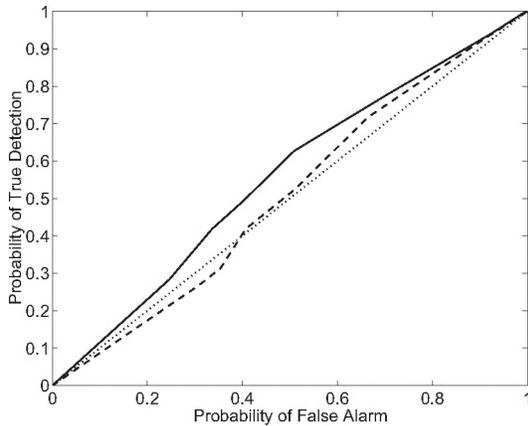
Figure 7: ROC curve comparing detection performance without time alignment (solid curve) and with time alignment (dashed curve).

Figure 8: Magnitude of normalized cross-correlations for the five OBHs relative to OBH L. The solid vertical line indicates the estimated time delay; the dotted line indicates the peak value.

7, however, this procedure has been found to actually decrease overall detector performance. The reason for this appears to be that the cross-phone correlations contain very little information with which to discriminate different lag times. This can be seen in Fig. 8, where we have plotted the cross-correlations for each OBH relative to OBH L. The cross-correlations are flat and noisy, making estimation of time delays difficult, and small errors in estimating the peak can translate into huge errors in the delays.

Finally, we considered the effect of removing the two problematic OBHs, H and J, from the analysis. Interestingly, though the two phones were very noisy, there is a significant drop in the performance of the detector. Furthermore, if time aligning is performed with these two OBHs removed, the results are comparable to that when no time aligning is performed.

Figure 9: ROC curve comparing detection performance using all five OBHs (solid curve) versus that when OBHs H and J are removed (dashed curve).

5 DISCUSSION OF RESULTS

In this paper we have described an algorithm for passive acoustic detection of marine mammals using independent component analysis. This algorithm was implemented and tested on data collected in the Bay of Fundy and containing a variety of right whale calls. The performance of the detector was such that it was possible to find an operating point on the ROC curve such that about three-fourths of whales are detected with about a third of all calls being false alarms.

In comparing ICA against PCA, it was found that ICA provides little improvement over PCA when the latter is used as a preprocessing step. This suggests that decorrelating the data goes a long way towards achieving statistical independence. PCA is much faster than ICA, at least in the implementation used based on mutual information, and is therefore recommended for use in place of ICA for detection purposes. Using PCA alone, the detection algorithm is able to run in well under real time.

Procedures for estimating relative delays and time aligning the receiver data were frustrated by hardware issues and noisy data. A consequence of poor delay estimation was that detector performance actually worsened when these corrective techniques were applied. Since the PCA component of the algorithm is sufficiently fast to allow for additional processing, it may be that inclusion of a better localization algorithm, such as matched field processing, is possible within a real-time system.

For marine mammal classification, the ICA algorithm presented may be suitable as a preprocessing step to feed into a large classifier. Non-Gaussianity as a metric for detection is useful for separating out ambient background sounds, but it alone is not suitable for classification. ICA may be useful in this regard by providing an estimate of the extracted source signal, which may then be used to classify the source into more specific categories.

REFERENCES

[1] S. L. Perry, D. P. Demaster, and G. K. Silber. Special issue: The great whales: History and status of six species listed as endangered under the U.S. Endangered Species Act of 1973. *Marine Fisheries Review*, 61:1–74, 1999.

[2] D. W. Laist, A. R. Knowlton, J. G. Mead, A. S. Colliet, and M. Podesta. Collisions between ships and whales. *Marine Mammal Science*, 17:35–75,

[3] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent Component Analysis*. J. Wiley, New York, 2001.

[4] Sheldon M. Ross. *Stochastic Processes*. John Wiley & Sons, New York, 1983.

[5] Alexander M. Mood, Franklin A. Graybill, and Duane C. Boes. *Introduction to the Theory of Statistics*. McGraw-Hill, New York, 3rd edition, 1974.

[6] Anthony J. Bell and Terrence J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1004–1034, 1995.

[7] Juha Karhunen. Neural approaches to independent component analysis and source separation. In *Proc. 4th European Symposium on Artificial Neural Networks (ESANN'96)*, pages 249–266, Bruges, Belgium, 24–26 April 1996.

[8] Athanasios Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Boston, 4th edition, 2002.

6 ACKNOWLEDGMENTS

A Comparison Of Methods For Detecting Right Whale Calls

David K. Mellinger
Cooperative Institute for Marine Resources Studies, Oregon State University
and
NOAA Pacific Marine Environmental Laboratory
2030 SE Marine Science Drive, Newport, OR 97365 USA
David.Mellinger@oregonstate.edu

ABSTRACT

North Atlantic, North Pacific, and southern right whales all produce the *up* call, a frequency-modulated upsweep in the 50-200 Hz range. This call is one of the most common sounds, and frequently the most common sound, received from right whales, and as such is a useful indicator of the presence of right whales for acoustic surveys. A data set was prepared of 1857 calls and 6359 non-call sounds recorded from North Atlantic right whales (*Eubalaena glacialis*) near Georgia and Massachusetts. Two methods for the detection of the calls were compared: spectrogram correlation and a neural network. Spectrogram correlation parameters were chosen two ways, by manual choice using a sample of 20 calls, and by an optimization procedure that used all available calls. Neural network weights were trained via backpropagation on 9/10 of the test data set. Performance was measured separately for calls of different signal-to-noise ratio, as SNR heavily influences the performance of any detector. Results showed that the neural network performed best at this task, achieving an error rate of less than 6%, and is thus the preferred detection method here. Spectrogram correlation may be useful in situations in which a large set of training data is not available, as manual training on a small set of examples achieved an error rate (26%) that may be acceptable for many applications.

SOMMAIRE

Les baleines franches de l'Atlantique Nord, du Pacific Nord et Sud produisent toutes une vocalisation montante, soit un balayage ascendant modulé en fréquence dans la région de 50 à 200 Hz. Cette vocalisation est un des sons les plus communs produit par les baleines franches et, par le fait même, est un indicateur très utile de la présence des baleines lors de sondages acoustiques. Un ensemble de données a été préparé avec 1857 vocalisations et 6359 sons non vocalisés enregistrés auprès de baleines franches de l'Atlantique Nord (*Eubalaena glacialis*) près de la Georgie et du Massachusetts. Deux méthodes de détection des vocalisations ont été comparées: la corrélation de spectrogramme et le réseau neuronal. Les paramètres de la corrélation de spectrogramme ont été choisis de deux façons: par choix manuel, en utilisant seulement 20 vocalisations, et par une optimisation de la procédure utilisant toutes les vocalisations. Les coefficients de pondération du réseau neuronal ont été établi par rétropropagation sur 9/10 des données de test. Les performances ont été mesurées séparément pour des vocalisations ayant des rapports signal sur bruit différents, le rapport signal sur bruit ayant une grande influence sur tout détecteur. Les résultats démontrent que le réseau neuronal performe mieux dans ce genre de tâche, atteignant un taux d'erreur de moins de 6% et, par conséquent, est défini ici comme la meilleure méthode de détection. La corrélation de spectrogramme peut être utile dans les situations où un grand nombre de données de formation ne sont pas disponibles. Le choix manuel sur de petite tranche d'échantillons a atteint un taux d'erreur (26%) qui pourrait être acceptable dans plusieurs applications.

1.   INTRODUCTION

Right whales (*Eubalaena* spp.) are the world's most highly endangered large whale, and among the most highly endangered marine mammal of any kind (Clapham *et al.* 1999; Hilton-Taylor 2000; IWC 2001). They have thus been the focus of intense conservation interest (Silber and Clapham 2001). Acoustic methods have been proposed for use in right whale conservation principally in two ways (Gillespie and Leaper 2001). Acoustic surveys can be used to determine seasonal movements, habitat requirements, behavior, and other characteristics of right whales. These surveys can be done using either towed arrays, real-time sonobuoys (Desharnais *et al.* 2000; McDonald and Moore 2002), or autonomous hydrophones (Clark *et al.* 2000; Waite *et al.* 2003; Wiggins 2003), instruments that record sound continuously for time periods of months to years. A second proposed application of acoustic methods is as part of a ship-strike avoidance system (Gillespie and Leaper 2001). In such a system, right whales are acoustically detected and localized in real time and their locations passed to ships, which can then be steered so as to avoid the whales. For either of these applications, a problem arises: how to find the sounds of right whales amid the thousands of hours of data. These sounds can be found by manual scanning of spectrograms, but in most cases this is labor-intensive and prohibitively expensive.

Automatic detection is often a better solution. This involves having a computer analyze a sound signal and determine the times at which a desired sound is present. Having sound analyzed automatically offers advantages over manual scanning besides cost: a computer is not subject to fatigue; a computer is unbiased, or rather its bias is constant and does not change over time; a computer typically works quite quickly, as for instance when it took only a few days to detect right whale calls in five hydrophone-years of data (Waite *et al.* 2003); and a computer method may be replicated exactly for different applications, ensuring comparability of the results.

A detection method is used for a sound of some desired type. In most cases, the desired sound is a stereotyped call made by a certain species, and this is true of right whales as well. One type of call frequently made by all three species of right whales is the low-frequency *up* call (Clark 1982), and indeed it is known to be one of the most common types of call in the species for which this has been quantified, Southern right whales (*E. australis*; Clark 1983) and North Pacific right whales (*E. japonica*; McDonald and Moore 2002). Note that the call under consideration is the lower-frequency *up* call between approximately 50 and 220 Hz (Clark 1982) rather than a higher-frequency call in the 300-600 Hz range that has also been referred to as the *up* call (Vanderlaan *et al.* 2003).

Because of the need for an automated method of detecting right whale calls, and because of the ubiquity of the *up* call in the sounds produced by right whales, it was decided to optimize a method for detecting *up* calls of North Atlantic right whales (*E. glacialis*). In this paper, we compare two principal methods of detecting right whale *up* calls, spectrogram correlation and a neural network. Two variations of the spectrogram correlation method are examined. The comparison is done on a test data set consisting of thousands of right whale *up* calls and other sounds recorded with them.

2.   METHODS

A comparison is done between two methods for detecting right whale *up* calls, spectrogram correlation (Mellinger and Clark 1997, 2000) and a neural network trained using backpropagation (Rumelhart and McClelland 1987). The spectrogram correlation method is developed separately in two different ways, by manual parameter choice and by an automated optimization procedure. Thus in effect there are three detection methods that are compared here: spectrogram correlation with manual parameter choice, spectrogram correlation with optimized parameter choice, and a neural network.

2.1.   Data Set

Data for this comparison is from recordings made in Dec. 1996 - Jan. 1997 from a cabled hydrophone array off Jacksonville, Florida; in May 2000 from "pop-up" autonomous hydrophones (Clark *et al.* 2000) in the Great South Channel, Massachusetts; and in March 2001 from pop-ups in Cape Cod Bay, Massachusetts. A spectrogram of each recording was made (frame size and FFT size 0.256 s, overlap 0.192 s, Hamming window) and the data were visually scanned for the presence of right whale *up* calls. Beginning and ending times of each call were marked, resulting in a set of 1857 total *up* calls.

The training and testing of a detection method also required a set of other, non-call sounds. These should be representative sounds from the entire set of the recordings, and as such a set of randomly-chosen times (with times of *up* calls removed) should suffice. However, a better approach than choosing times randomly is to choose times at which some significantly loud sound occurs in the frequency range of interest. This approach is better than using random times because it targets those parts of the recordings that are likely to cause difficulties for a detection method; a set of random times is likely to include a lot of instances when only background noise is present, and these instances are not likely to be helpful for developing a robust detector. Accordingly, a process was run to find sounds in a

Recording location	Date	# *up* calls	# non-call sounds
Off Jacksonville, Fla.	Dec. '96 - Jan. '97	124	210
Great South Channel, Mass.	May '00	169	1421
Cape Cod Bay, Mass.	Mar. '01	1564	4728
Total		1857	6359

Table 1. Recording locations and dates and the number of *up* calls and non-call sounds.

frequency band encompassing right whale *up* calls, 50-250 Hz. These sounds included the right whale *up* calls, so the *up* calls were removed. The resulting set contained some right whale sounds – calls other than *up* calls – that were retained in the "non-call sounds" category since the problem here is to detect *up* calls. The set also contained a handful of "uncertain" sounds, those for which it was unclear whether or not the call was an *up* call; this happened either because a sound was too faint to determine whether it was a call or merely a bit of background noise, or because a call had an odd frequency contour that was somewhat, but not definitively, like an *up* call. These unclear sounds were removed from the non-call sound set. The resulting set contained 6359 non-call sounds used for training and testing. Table 1 shows the number of sounds from each recording location.

2.2. Detection Process

For both methods, the overall detection process is as follows. An input sound signal is transformed into a spectrogram, to which a conditioning technique – spectrum level equalization and normalization – is then applied. The normalized spectrogram is then used as input to the *detection method* (spectrogram correlation or neural network), resulting in a *detection value D* – a number indicating the certainty that a right whale *up* call is present. A threshold is then applied, and the times at which the detection function goes over threshold are considered to be *detection events* – right whale *up* calls.

In more detail, the first step in the detection process is making a spectrogram. The exact parameters involved in making the spectrogram vary between the three detection methods and are covered in more detail below. For all three methods, the next step is spectrogram level equalization and normalization. This is done by time-averaged spectrum equalizing (Van Trees 1968), followed by hard-limiting the lower and upper bounds of spectrogram amplitudes. In other words, the time-averaged spectrogram value is calculated for each frequency band of the spectrogram; this is subtracted from the spectrogram at each time step, and floor and ceiling values are applied. More exactly, let $S(t, f)$ represent the spectrogram. Then the normalized spectrogram $\hat{S}(t, f)$ is given by

$$(1) \quad M(t, f) = kS(t, f) + (1 - k)M(t - \Delta t, f)$$

$$(2) \quad S_1(t, f) = S(t, f) - M(t, f)$$

$$(3) \quad \hat{S}(t, f) =$$

$$\max(S_{floor}, \min(S_{ceiling}, S_1(t, f))) - S_{floor}$$

where $M(t, f)$ represents the time-averaged spectrogram value at time t for frequency f, Δt is the time step between spectrogram frames, k is a time constant that determines how quickly this process responds to changes in level in the spectrogram, $S_1(t, f)$ is the normalized spectrogram, and S_{floor} and $S_{ceiling}$ are the minimum and maximum normalized spectrogram values. The values of k, S_{floor}, and $S_{ceiling}$ are chosen as explained below.

This equalization process (Fig. **1**) has two beneficial effects. It removes from the spectrogram any sounds lasting a sufficiently long time, including ship sounds, electrical noise, and wind noise. In effect, short-duration sounds – such as right whale *up* calls – are emphasized. It also normalizes average levels across frequency, relatively emphasizing fainter parts of the spectrogram.

The next step in the detection process is application of one of the three detection methods:

(1) *Spectrogram correlation by manual choice of parameters*. Spectrogram correlation (Mellinger and Clark 1997, 2000) operates by cross-correlating a synthetic kernel with a conditioned spectrogram of the input signal (Fig. **2**). Correlation is done in only the time dimension of the spectrogram, so the result is a one-dimensional signal – the detection function $d(t)$. An example is shown in Fig. **2**c. The synthetic kernel is constructed for a specific call type, in this case a right whale *up* call. The kernel (Fig. **2**b) has an axis matching the frequency contours of an *up* call; this part of the kernel is positive. Flanking areas of the kernel are negative, a design that results in the correlation's dot-product producing zero when interfering sounds intersect both the axis and flanking regions. Details about kernel design, including equations for making kernels, may be found elsewhere (Mellinger and Clark 2000).
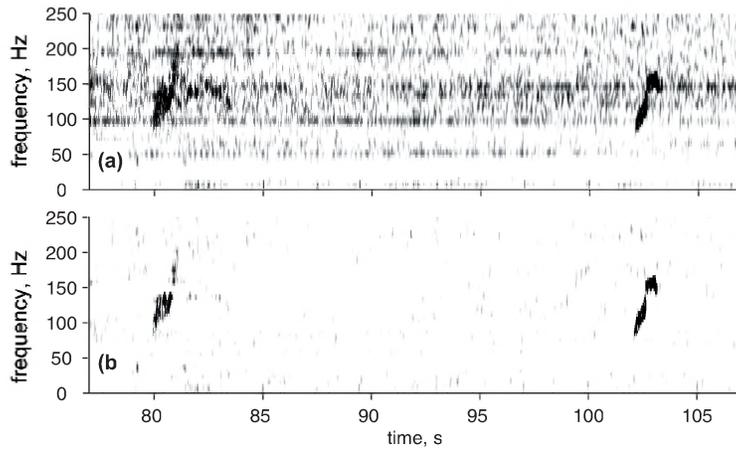
Figure 1. Example of spectrogram equalization and normalization. (a) Recording including two right whale up calls. Spectrogram parameters: frame size 0.128 s, FFT size 0.256 s, overlap 0.112 s, Hamming window. (b) The same spectrogram after equalization and normalization.
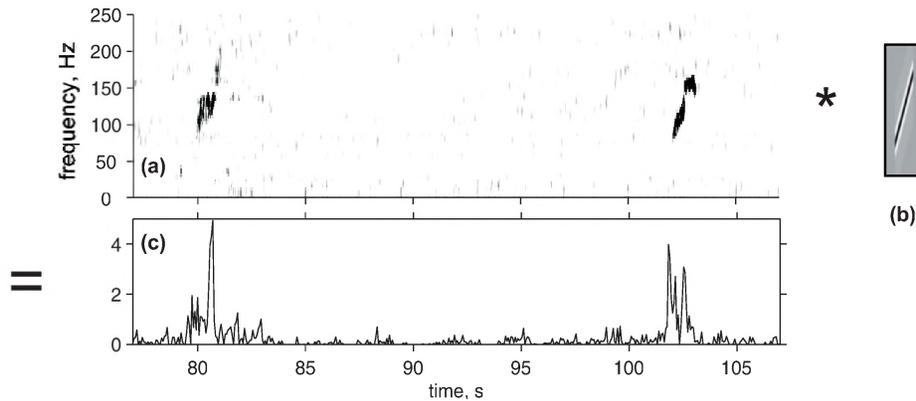
Figure 2. Schematic depiction of spectrogram correlation. (a) Normalized spectrogram, as in Fig.1. (b) Spectrogram correlation kernel. (c) Detection function produced by the spectrogram correlation function; its peaks correspond to the times at which right whale *up* calls are present.

The performance of spectrogram correlation is affected by the choice of the spectrogram parameters of frame size, FFT size, overlap, and window type; by the equalization parameters k and S_{floor}, and $S_{ceiling}$; and by the kernel parameters of start frequency, end frequency, duration, and bandwidth. With such a large number of parameters, the set of reasonable combinations of parameter values exceeds 10^6, far too many for exhaustive testing. Two approaches were taken to address this issue. In the first approach, the author visually examined and experimented with a random sample of 20 *up* calls from the set of 1857 marked calls.

Parameters controlling the spectrogram correlation process were chosen by hand, in a sequence of successive steps: First spectrogram parameters (frame size, FFT size, overlap, and window type) were chosen such that the right whale *up* calls appeared with reasonable clarity (in both time and frequency) in a spectrogram. Next, the time constant k was chosen for spectrogram equalization such that right whale *up* calls were relatively little affected, and common noise sounds such as from ships (Fig. **1**) were largely removed. The author's experience is that a good value for k is one that causes a given noise level to decay to $1/e$ of its original

value after a time period of five times or more the length of the target call type – i.e., for *up* calls, after 5 or more seconds.

Once the spectrogram and normalization parameters were chosen, it remained to decide on parameters for the spectrogram correlation kernel. This was done by measuring the start frequency, end frequency, duration, and bandwidth of the 20 example calls. In doing this, it was noted that the duration of the *up* calls was almost always less than about 1 s, but different *up* calls spanned different frequency ranges. For instance, one call might range from 70 Hz to 150 Hz, and another from 120 Hz to 200 Hz. It is possible to make one kernel that would detect both of these sounds, one with a kernel axis from 70 Hz to 200 Hz lasting nearly 2 s; such a kernel would match the frequency sweep rate of both of those examples. But that kernel would be especially susceptible to interfering sounds, since the positive region of the kernel would be especially large for the size of any given call. To solve this problem, separate kernels were made for the two halves of the frequency range: one sweeping from 70 to 150 Hz, the other from 120 to 200 Hz. This produces two detection functions, one per kernel. These were combined at each time step by using the maximum value of the two to produce $d(t)$. Another problem was that different *up* calls swept upwards at significantly different rates, and a given kernel did not perform well for all sweep rates. This problem was solved in a similar manner, by creating three kernels with three different sweep rates, cross-correlating them with the spectrogram, and taking the maximum value at each step. With kernels of different durations, a weighting factor proportional to the inverse of the kernel duration was needed so that cross-correlation values were comparably scaled. If $d_i(t)$ is the cross-correlation result for the ith kernel, and g_i is the duration of the ith kernel, then the overall detection function $d(t)$ was given by

$$(4) \qquad d(t) = \max_i(d_i(t)/g_i)$$

Not all combinations of kernel start/stop frequency and kernel duration were used, since only some of these combinations were observed among the 20 example calls. A total of five different kernels were ultimately used (Table 2), with the final detection function $d(t)$ at each time t being the maximum of the five cross-correlations. Table 2 shows the values that were finally chosen by manual choice of parameters.

Given the detection function $d(t)$, the detection value D for a given call or non-call sound was simply the maximum of $d(t)$ in a 2 s-long period centered on the call or non-call sound. A D value was calculated for each call in the training set to produce a collection of "call" detection values, and likewise for the non-call sounds to produce a collection of "non-call" detection values.

(2) *Spectrogram correlation with optimized parameter choice.* A second method of choosing the parameters for the spectrogram correlation detector was to run an optimization procedure to find the set of parameters that worked best. "Best" was defined as the smallest false positive proportion(e_p) at a fixed false-negative proportion (e_n) of 10%. This terminology is explained in detail in the "Performance evaluation" section below.

The optimization procedure used a fixed range of discrete possible values for each spectrogram correlation parameter (Table 3). This range for each parameter p_i was determined as all values that seemed even slightly reasonable in examination of example calls. For instance, the parameter p_9, the duration of the kernel, included values ranging from the measured durations of the shortest call found, 0.55 s, to that of the longest call, 1.14 s. Similar ranges were chosen for all nine parameters that determine the operation of the spectrogram correlation calculation. Parameter p_8, "number of segments", determines the number of kernels into which

Spectrogram	
frame size	0.256 s (512 samples)
FFT size	0.256 s (512 samples)
overlap	0.192 s (384 samples)
window type	Hamming
Equalization	
time to decay to $1/e$	10 s ($k = 0.0064$)
floor value S_{floor}	0.9
ceiling value $S_{ceiling}$	1.5
Kernel	
bandwidth	10 Hz
combinations of	(70 Hz, 150 Hz, 0.6 s)
(f_0, f_1, duration)	(70 Hz, 150 Hz, 1.0 s)
	(120 Hz, 200 Hz, 0.5 s)
	(120 Hz, 200 Hz, 0.7 s)
	(120 Hz, 200 Hz, 1.0 s)

Table 2. Parameter values for the spectrogram correlation detection method that were manually chosen by examination of 20 example calls. The floor and ceiling values S_{floor} and $S_{ceiling}$ are spectrogram amplitudes whose scaling is unknown, so units are not given.

the given frequency range is divided. For $p_8 > 1$, the frequency range from f_0 to f_1 is divided into p_8 separate, equal spans, and one kernel is constructed for each span These kernels are used as above: Each one is cross-correlated with the input spectrogram, and the detection function $d(t)$ at each time step t is the maximum of the cross-correlation functions.

The optimization procedure worked as follows. A set of the nine parameter values $\{p_i\}^{i=1..9}$ was randomly chosen, i.e., one value was randomly chosen from the "set of values" in each line of Table 3. Performance was evaluated using this set of parameters by running the spectrogram correlation process on the entire data set of all calls and all non-call sounds, and (as described above) measuring e_p at the point where $e_n = 10\%$. The initial performance score s_{init} was this value of e_p. Next, for the first parameter p_1 from Table 3 (spectrogram frame size), the value just below the randomly-chosen value was selected. For instance, if the randomly-chosen value for p_1 was 0.128 s, then the value of 0.064 s was selected. This value was substituted for p_1 in the

set $\{p_i\}$, and the spectrogram correlation process was run and evaluated again to get a new score $s_{1,low}$. (The subscripts indicate that the next-lower value for parameter 1 was used to calculate this score.) Then the next-higher value for p_1 was used instead of the next-lower value (in the instance above, 0.256 s), and the resulting score $s_{1,high}$ was calculated. This process of trying each next-lower and next-higher parameter value was repeated for each parameter in $\{p_i\}$, resulting in 18 scores $\{s_{1,low}, ... s_{9,low}, s_{1,high}, ... s_{9,high}\}$. The best – i.e., lowest – of these scores was examined. If it was better than s_{init}, then the parameter set corresponding to this best score was chosen, and the process was repeated. If that

Parameter name	Variable	Set of values for optimization	Optimized value
Spectrogram			
frame size	p_1	0.064, 0.128, 0.256, 0.512 s	0.128 s
FFT size*		same as frame size	0.128 s
overlap*		3/4 of frame size	0.096 s
window type*		Hamming	Hamming
Equalization			
time to decay to $1/e$	p_2	1, 2.5, 5, 10, 20 s	1 s
floor value S_{floor}	p_3	0.7, 0.8, 0.9, 1.0, 1.1	0.7
ceiling value $S_{ceiling}$	p_4	1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8	1.7
Kernel			
bandwidth	p_5	5, 8, 10, 15, 20 Hz	20 Hz
start frequency f_0	p_6	70, 80, 90, 100, 110 Hz	80
end frequency f_1	p_7	150, 175, 200, 230 Hz	175
number of segments	p_8	1, 2, 3	1
duration	p_9	0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2 s	0.8 s

Table 3. Parameters used in the optimization process for spectrogram correlation. Each parameter has specified here the variable name p_i, the set of possible values used in optimization, and the value of that parameter for the optimized set. Parameters marked with * are not independent, but rather are fixed or are determined by other parameter values.

best score was no better than s_{init}, then repetition stopped and the resulting parameter set $\{p_i\}$ was considered the maximally-performing one.

The procedure just described is a form of steepest-descent search in a discrete parameter space. Such searches are influenced, sometimes heavily, by the choice of starting point – by the randomly-chosen parameter set used. Accordingly, this optimization procedure was run 20 times, with the best-scoring parameter set chosen as the final result. This set is shown as the last column of Table 3.

(3) *Neural network*. Neural networks have been used for detection of tonal sounds (Potter *et al.* 1994, Murray *et al.* 1998, Deecke *et al.* 1999), but not, to the author's knowledge, heretofore for right whale calls. A feedforward neural network (Hagan *et al.* 1996) was constructed with 252 input elements, 10 hidden units, and 1 output unit. Each hidden unit consisted of a weighted sum with bias followed

by an arc-tangent nonlinearity (Hagan *et al.* 1996). The output unit was linear, consisting of just a weighted sum.

Input to the network was a small piece of a spectrogram (frame size and FFT size 0.256 s, overlap 0 s, Hamming window), here called a *minigram*. Each minigram spanned the frequency range from 70 to 230 Hz and lasted 1.5 s. Such a minigram has 252 cells; it was the values (amplitudes) in this minigram that were used as input values to the neural network. Figure **3** shows some examples of call and non-call minigrams.

Minigrams of 90% of the data set were used in training and testing the neural network. The training data for this network came from the set of 1857 *up* calls and 6359 non-call sounds. For each of these sounds, a minigram was made from a spectrogram of the recording containing the sound. For each *up* call in the set, the start- and end-time of the minigram were set such that the frequency contour of the

call passed through the 120-Hz frequency bin of the spectrogram 0.55 s after the start of the minigram. Adjusting the timing in this manner time-aligned all of the *up* calls in the minigrams such that if the minigrams were laid atop one another, their frequency contours would occur along the same diagonal line, regardless of the start and end frequencies of the calls. Non-call sounds were similarly time-aligned, but since there was no frequency contour in the non-call sounds to use for alignment, they were aligned by centering the time of maximum energy in the minigram. Training the neural network required target values, values that the network was supposed to learn to produce for the call and non-call minigram inputs. These target values were set at +0.5 and -0.5, respectively. The network was trained using these call and non-call minigrams. Actually not all of

the available minigrams were used for training; one-tenth of the calls and one-tenth of the non-call sounds, randomly chosen, were reserved for testing the trained network. This was done because neural networks have enough parameters, in the form of connection weights, that they are capable of learning their training set – basically, learning to identify specific minigrams by idiosyncratic characteristics of those minigrams. For this reason, it is better to test a network with "new" data absent from its training data set. (This problem does not exist with the optimization procedure for spectrogram correlation, because the number of parameters – nine – is far too small for the process to learn specific calls.)
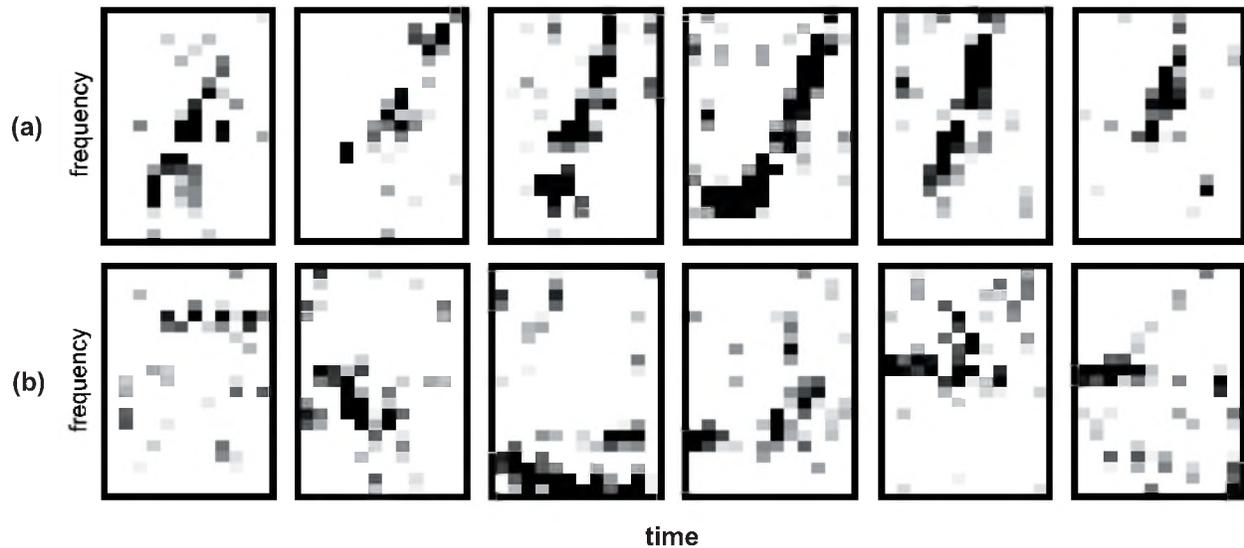
Figure 3. Examples of (a) *up* call minigrams and (b) non-call minigrams.

The network, coded using the Matlab neural network toolbox, was trained using standard gradient-descent backpropagation with an adaptive learning rate (Hagan *et al* 1996). Before starting training, the network weights were initialized to small random values so that different units would adapt differently. Training was done in "batch mode," i.e., all of the call and non-call minigrams were presented in parallel, output values for each input were computed, and all network weights were updated. This constituted one *epoch* of training. Training was done for 5000 epochs in all, a number chosen because at that point the rate of performance improvement per epoch had become very small.

After this training was complete, the network was tested with the one-tenth of the calls and non-call sounds that were not used for training. For a given network input (a minigram

of the call or non-call sound), the detection value D was simply the network output. The set of D values for the set of calls and the set for the non-call sounds were used in measuring the performance of the network.

2.3. Performance Evaluation

Performance was evaluated by raising and lowering a threshold and comparing the threshold to detection values produced by each detection method. For the spectrogram correlation detection method using either type of parameter choice, a given threshold was compared to the maximum of the detection function $d(t)$ in a 2 s long period centered on each *up* call or non-call sound. For the neural network, the threshold was compared to the output of the network for each call or non-call sound.

For either detection method and for a given threshold, two error measures were determined: the false negative proportion e_n, which is the proportion of missed calls as a fraction of all calls, and the false positive proportion e_p, the proportion of wrongly detected noise sounds as a fraction of all noise sounds. Raising the threshold makes the proportion of false negatives rise and the number of false positives fall, and inversely for lowering the threshold. By varying the threshold between the lowest and highest values produced by any given detection method, one could obtain a parametric curve – the *performance curve* – detailing the performance of the detection method. Figure 4 shows some examples of such curves. This curve is analogous to the Receiver Operating Characteristic (ROC) curve used in measurement of radar system performance.

A special point on the performance curve was used for comparison of methods. This was the point at which the false-negative proportion e_n was 10%. The 10% false-negative level was chosen because an *up* call detection method is probably useful even if it misses 10% of the calls present: Right whales make calls in clusters lasting a few minutes and containing an average of 2 calls (North Atlantic right whales; Matthews *et al.* 2001) to over an hour and containing 10-15 calls (North Pacific right whales; McDonald and Moore 2002). With a 10% missed-call rate – i.e., a 90% detection rate – and assuming that the probability of detection is independent from one detection to the next, the probability that a detector would miss a cluster ranges from 0.01 down to 10^{-10} or less. A detector might thus some calls but would be unlikely to miss a whale.

The false-positive proportion e_p corresponding to the 10% false-negative point was used as a performance metric, a metric named the *single-point score*. By choosing this single point on the curve, performance measurement for a given detection method and its configuration parameters was reduced to a single number, enabling direct comparison of disparate methods (and, as explained above, enabling the spectrogram correlation optimization procedure to choose the "best" parameter set).

The performance of any call-detection method depends critically on the signal-to-noise ratio (SNR) of the calls under consideration. The SNR of a given call was characterized as the ratio of the average power during the call in the 50-250 Hz frequency band to the average "noise power," the power in the 10 s before and 10 s after the call. Note that since this calculation is done before any kind of spectrum equalization, tonal background noise that fluctuates in intensity can make calls that are relatively apparent in a normalized spectrogram have an SNR of 0 dB or even less. The performance curve was calculated separately for calls with SNR <0 dB, calls with SNR from

0-10 dB, calls with SNR from 10-20 dB, and calls with SNR >20 dB.

3. RESULTS

The performance of spectrogram correlation with manually-picked parameters is shown in Fig. 4a. This figure shows a series of parametric performance curves, one curve for each range of call SNRs; each point on this curve corresponds to a certain threshold value, and the (x,y) location plotted is the point (e_p, e_n), i.e., the false-positive vs. false-negative error proportions. In such a plot, the lower-left area is the region of least error and hence of best performance. The single-point score on a given curve may be found by drawing a horizontal line from the 10% mark on the y-axis, seeing where this line intersects the curve, and determining the x-axis value of the intersection. Note that the axes of this plot are logarithmic, so constant ratios are the same size, and small distances on the plot can correspond to relatively large differences in performance. Figure 4b shows the performance curves for spectrogram correlation when the parameters are chosen by the optimization procedure (see also the rightmost column of Table 3), while Fig. 4c shows the curves for the neural network. Figure 4d is a comparison of performance curves for the three detection methods; it was made using all calls regardless of SNR. In this figure, the single-point score is indicated for the neural network (6%) and spectrogram correlation with manually chosen parameters (26%).

Figure 5 shows some of the calls that resulted in high output values when used as input to the neural network.

4. DISCUSSION

As expected, performance of all methods as measured by the ROC curve generally improved with increasing SNR, typically by a factor of 3-6p in going from the calls with SNR less than 0 dB to those with SNR greater than 20 dB. Also note that performance varied only somewhat between the calls with 10-20 dB SNR and those with >20 dB SNR. One reason for this may be that the calls with 10 dB SNR contain as much information as the spectrogram correlation method is able to use; the missed calls may be due to other effects such as odd frequency contours that do not match the usual *up* call contour.

The performance of the optimized spectrogram correlation method (Fig. 2b, and rightmost column of Table 3) was significantly better than that of spectrogram correlation with manually-chosen parameters. It is not surprising to find a difference, as the optimization procedure used the entire data set to choose its parameters, while the manually-chosen parameters were selected using only a small subset. In addition, the optimization procedure used many days of
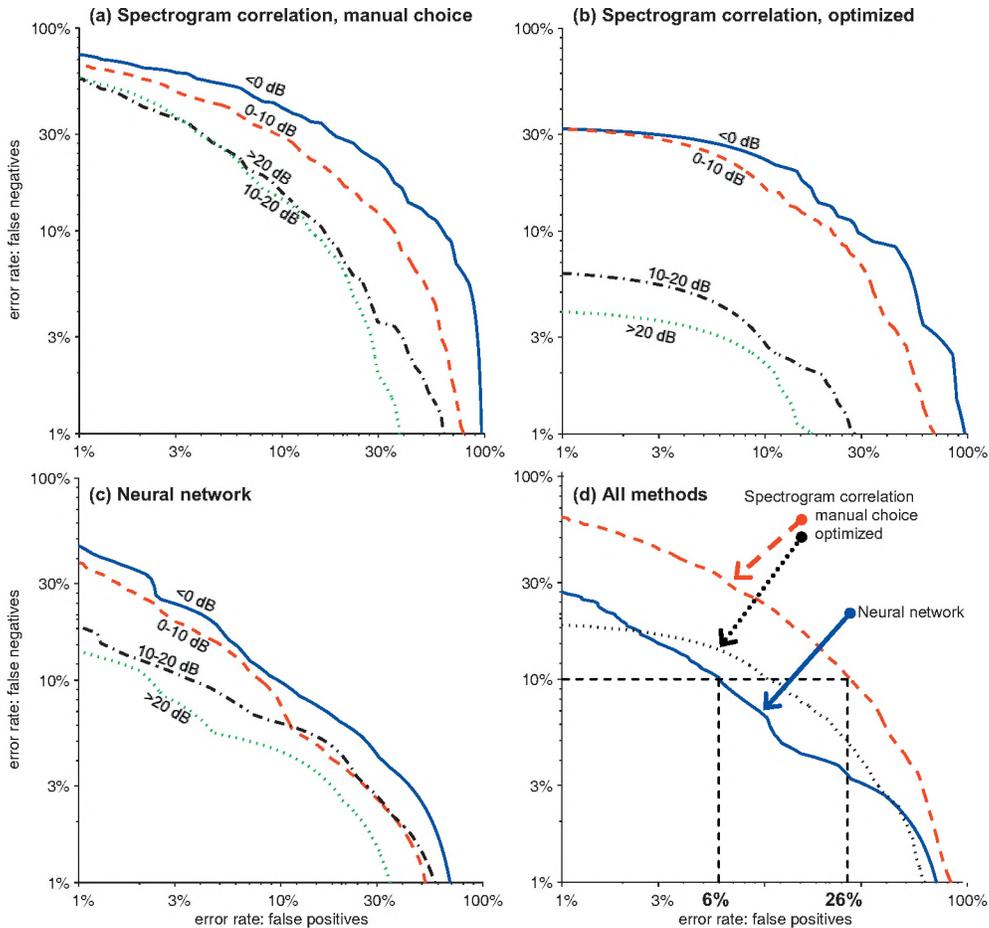
Fig. 4. Performance curves of the various methods. In each, the false positive rate e_p is plotted versus the false negative rate e_n. (a) Performance of spectrogram correlation using manually-chosen parameters. The labels on each curve shows the signal-to-noise ratio of the calls used for that curve. (b) Performance of spectrogram correlation using parameters found by the optimization procedure. (c) Performance of the neural network. (d) Performance comparison of the three methods, with calls of all signal-to-noise ratios lumped together.
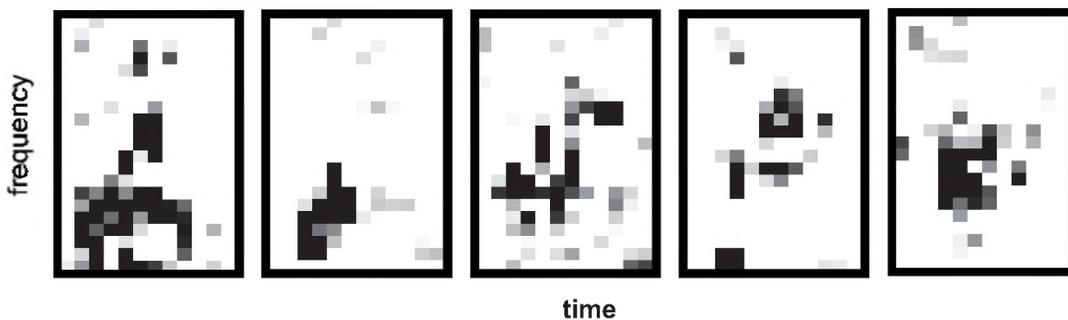
Figure 5. Examples of some of the noise sounds that resulted in large positive outputs of the neural network, i.e., outputs that are similar to those. Many of these have sound in a band from lower left to upper right, exactly where right whale *up* calls have sound.

computer time (though using a relatively inefficient algorithm), in contrast to the manual method which required only a few hours of the author's time. However, the degree of difference was surprisingly large, with the optimized method generally performing better by a factor of two to five.

Neural network performance (Fig. 4c) was somewhat better than the optimized spectrogram correlator for calls of poor SNR, though slightly worse for calls of high SNR. Stated another way, the performance curves of the neural network were more closely bunched than those of the optimized spectrogram correlator. This difference may be a reflection of the fact that the neural network had more tunable parameters, and was thus able to adapt to the types of calls – particularly faint calls – and non-call sounds better than the spectrogram correlator. Its range of variation between low-SNR and high-SNR calls was therefore smaller.

The performance curves of Figs. **4**a and 4c show anomalies in which the performance for calls of greater SNR was occasionally worse than that for calls lesser SNR. It is not known why this occurred; it may have something to do with the fact that SNR was measured before spectrum equalization, while the methods operate from a spectrogram to which equalization has been applied.

As shown in Fig. **4**d, the neural network performed better than either spectrogram correlation method – substantially better than with manually-chosen parameters, and somewhat better than with optimized parameters over most of the range of measurement. There are regions at either end of the curves (the low false-negative and low false-positive ends) at which optimized spectrogram correlation performed slightly better than optimized spectrogram correlation, but in the broad middle range, the neural network was better. This raises the question of whether spectrogram correlation is useful at all. There are three answers to this question:

1) No. The neural network with weights adjusted by the training process described above is plainly the preferred method for detecting the right whale *up* calls in this data set, and probably for detecting right whale *up* calls in other data sets.

2) Possibly. One open question is how well the neural network would work for right whales recorded in different locations or at different times of year. This network is optimized for the sounds (both calls and non-call sounds) with which it was trained, but its performance would probably degrade for data collected in a different sonic environment – where, for instance the types of transient interfering sounds were different. Performance of spectrogram correlation would probably degrade too, but it might degrade less, since the spectrogram correlation

process and parameters are less finely tuned to the set of training calls and especially non-noise sounds. Whether this is true is a matter for future research.

3) Yes. The spectrogram correlation method worked reasonably well when its parameters were chosen manually from only a few (20) example calls. Spectrogram correlation may be useful for detecting other call types from right whales, or calls from other animal species. In many cases, a large set of recordings containing thousands of training calls may not exist, or resources may not be available for a person to mark where in the recordings the desired calls are. Having such a set of marked calls is a prerequisite for training a neural network; in the absence of the marked calls, spectrogram correlation may well be a viable option. A single-point score of 26% false detections, as achieved by spectrogram correlation, is quite tolerable for many applications. For instance, if the application is determining whether calls of a certain type occur in a large body of recordings (*e.g.*, Clark *et al.* 2000, Waite *et al.* 2003), then using spectrogram correlation with manual choice of parameters could result in detection of most or all of the desired calls, in addition to a small set of undesired non-call sounds. Examining an extra 26% of a small set of detections is almost trivial. In such a case, spectrogram correlation would have worked well.

Another case in which spectrogram correlation might be useful is when minimizing the number of missed calls (false negatives) is the desired. If, for instance, one desired a missed-call rate of 1%, then optimized spectrogram correlation is the best-performing method. Setting the detection threshold for such a missed-call rate would necessarily lead to a large number of false detections (>50% in this case), but that could well be acceptable for some applications. One example is a real-time system for avoiding ship strikes, in which a human operator would check each possible detection before announcing the presence of a right whale. In such a system, the cost of a false detection could be only minimal, but that of a missed detection extremely high.

How would the neural network, which was trained using discrete minigrams, be used in applications such as real-time detection where the signal is continuous in time? The network would be applied once per spectrogram time slice, by extracting the time-frequency portion of the spectrogram – the minigram – that begins at that time slice and has the same frequency bounds and duration as the minigrams used for training. The network would produce an output value for this input minigram; the network's successive output values over time would constitute a detection function, similar to that of Fig. 2b. A threshold could then be applied to this function, and supra-threshold peaks in the detection function

would indicate where a right whale call was most likely to be present.

Neural networks offer promise as a high-performance method for call detection. There remain a number of research issues about them. A large training set is needed for the neural network, but how large? A better phrasing of this question is to ask how network performance would degrade with smaller training sets. A related question is whether selecting subsets of the training set – say, calls that span the perceived range of variation of right whale *up* calls – would lead to equally good performance. Related to both of these is the optimum design of the neural network: would networks of fewer or more hidden units, or different training regimes, have performed substantially differently? These questions are further topics for research.

In conclusion, the neural network was generally found to perform the best at the task of detecting right whale *up* calls in a data set consisting of 1857 *up* calls from North Atlantic right whales and 6359 non-call sounds. The network outperformed spectrogram correlation over most of the range of desired performance; this was true when the spectrogram correlation process used a manually-chosen set of parameters, as well as when parameters were chosen by an optimization procedure. The neural network is thus the best choice for the detection of right whale *up* calls from the given recording locations and probably elsewhere. Despite the superior performance of the neural network, spectrogram correlation is still a viable option for call types for which a large set of marked training examples is not available, or for when a very low number of missed calls is desired.

ACKNOWLEDGEMENTS

REFERENCES

Clapham, P.J., S.B. Young, and R.L. Brownell, Jr. 1999. Baleen whales: conservation issues and the status of the most endangered populations. Mamm. Rev. 29:35-60.

Clark, C.W. 1983. Acoustic communication and behavior of the Southern Right Whale (*Eubalaena australis*). In *Communication and Behavior of Whales*, R. Payne, ed. (Westview, Boulder), pp. 163-198.

Clark, C.W. 1982. The acoustic repertoire of the southern right whale, a quantitative analysis. Anim. Beh. 30:1060-1071.

Clark, C.W., D. Gillespie, A. Moscrop, T. Fowler, T. Calupca, and M. Fowler. October 26-27, 2000. Acoustic sampling of right whale vocalizations in the Great South Channel using sea-floor pop-up recorders. Report of the Right Whale Consortium, Boston, MA.

Deecke, V.B., J.K.B. Ford, and P. Spong. 1999. Quantifying complex patterns of bioacoustic variation: Use of a neural network to compare killer whale (*Orcinus orca*) dialects. J. Acoust. Soc. Am. 105:2499-2507.

Desharnais, F., M. Laurinolli, A. Hay, and J.A. Theriault. 2000. A scenario for right whale detection in the Bay of Fundy. Proc. IEEE Oceans 2000, pp. 1735-1742.

Gillespie, D., and R. Leaper. 2001. Report of the Workshop on Right Whale Acoustics: Practical Applications in Conservation. Technical Report, Intl. Fund for Anim. Welfare, 411 Main St., PO Box 193, Yarmouth Port, MA. ISBN 1-901002-08-X.

Hagan, M., H. Demuth, and M. Beale. 1996. *Neural Network Design*. Brooks/Cole: Pacific Grove.

Hilton-Taylor, C., eds. 2000. 2000 IUCN Red List of Threatened Species. Intl. Union for Conservation Nature Natural Resources: Gland, Switzerland and Cambridge, UK.

IWC. 2001. Report of the workshop on status and trends of Western North Atlantic right whales. J. Cetacean Res. Manage. (special issue) 2:1-60.

Matthews, J.N., S. Brown, D. Gillespie, M. Johnson, R. McLanaghan, A. Moscrop, D. Nowacek, R. Leaper, T. Lewis, and P. Tyack. 2001. Vocalisation rates of the North Atlantic right whale. J. Cetacean Res. Manage. 3:271-282.

McDonald, M.A., and S.E. Moore. 2002. Calls recorded from North Pacific right whales (*Eubalaena japonica*) in the eastern Bering Sea. J. Cetacean Res. Manage. 4:261-266.

Mellinger, D.K., and C.W. Clark. 1997. Methods for automatic detection of mysticete sounds. Mar. Freshwater Beh. Physiol. 29:163-181.

Mellinger, D.K., and C.W. Clark. 2000. Recognizing transient low-frequency whale sounds by spectrogram correlation. J. Acoust. Soc. Am. 107:3518-3529.

Murray, S.O., E. Mercado, and H.L. Roitblat. 1998. The neural network classification of false killer whale (*Pseudorca crassidens*) vocalizations. J. Acoust. Soc. Am. 104:3626-3633.

NMFS. 2002. Species listed under the Endangered Species Act of 1973. Technical report, National Marine Fisheries Service, http://www.nmfs.noaa.gov/prot_res/species/ESA_species.html.

Potter, J.R., D.K. Mellinger, and C.W. Clark. 1994. Marine mammal call discrimination using artificial neural networks. J. Acoust. Soc. Am. 96:1255-1262.

Rumelhart, D.E., and J.L. McClelland. 1987. *Parallel Distributed Processing*. MIT: Cambridge.

Silber, G.K., and P.J. Clapham. June 2001. Updated Recovery Plan for the Western North Atlantic Right Whale, *Eubalaena glacialis*. Technical report, Northeast Fisheries Sci. Center, National Mar. Fisheries Service, Woods Hole.

Vanderlaan, A.S.M., A.E. Hay, and C.T. Taggart. 2003. Characterization of North Atlantic right-whale (*Eubalaena glacialis*) sounds in the Bay of Fundy. IEEE J. Oceanic Engr. 28:164-173.

Van Trees, H.L. 1968. *Detection, Estimation, and Modulation Theory*, Vol. I. Wiley: New York.

Waite, J.M., K. Wynne, and D.K. Mellinger. 2003. Documented sighting of a North Pacific right whale in the Gulf of Alaska and post-sighting acoustic monitoring. Northwestern Naturalist 84:38-43.

Wiggins, S. 2003. Autonomous Acoustic Recording Packages (ARPs) for long-term monitoring of whale sounds. Mar. Tech. Sci. J. 37(2):13-22.

Detection Of Frequency-Modulated Calls Using A Chirp Model

Justin Matthews
International Fund for Animal Welfare,
87-90 Albert Embankment, London, SE1 7UD, UK
whalesong@ifaw.org

ABSTRACT

Many cetacean vocalisations are tonal and most are frequency-modulated. The detection algorithm presented here breaks the frequency contour into a sequence of elements. Each element is sufficiently short that a linear approximation to the frequency contour can be made. In this way the problem is simplified from that of detection of an unknown signal, to the detection of a known signal (a linear chirp) with unknown parameters. The method of estimation is based on maximum likelihood, and the start frequency, chirp rate and amplitude of each element are estimated. Further analysis is then carried out on groups of concatenated chirps (i.e. calls) to classify them.

Results are given on performance for the supplied test recording and for synthetic signals in white noise. The pros of the algorithm are: good detection performance, at least in white noise; high resolution; ease of interpretation; flexibility; data compression. The cons are: computational cost; deterioration of performance in non-white noise or with amplitude-modulated signals. Further development is needed to reduce errors with overlapping tonal or non-tonal signals. The algorithm is currently being applied to the problem of detecting right whale vocalisations and distinguishing them from those of humpback whales.

RÉSUMÉ

Plusieurs vocalisations de cétacés sont de type tonal et la plupart sont modulées en fréquence. L'algorithme de détection présenté ici sépare le contour de fréquence en une séquence d'éléments. Chacun des éléments est suffisamment petit pour qu'une approximation linéaire du contour de fréquence puisse être effectuée. Le problème est donc en ce sens simplifié de façon à ce que la détection d'un signal inconnu passe à celle d'un signal connu (une modulation linéaire de fréquence) avec des paramètres inconnus. La méthode d'estimation est basée sur le maximum de vraisemblance, et la fréquence de départ, le taux de modulation et l'amplitude de chacun des élements sont estimés. Des analyses plus poussées sont alors effectuées sur des groupes de modulations enchaînées (i.e. vocalisations) afin de classifier les sons comme étant du bruit ou comme faisant partis d'une espèce spécifique.

Les résultats sont tirés de la performance des données de test et de signaux synthétiques en présence de bruit blanc. Les avantages de cet algorithme sont: une bonne performance de détection, du moins à l'intérieur d'un bruit blanc; une haute résolution; la facilité d'interprétation; la flexibilité; la compression de données. Les désavantages sont: les coûts computationnels; la détérioration de la performance à l'extérieur d'un bruit blanc ou avec un signal modulé en amplitude. Des développements plus poussés sont requis afin de réduire les erreurs provenant de la superposition d'un son tonal sur un son non tonal. L'algorithme a été appliqué aux problèmes de détection des vocalisations des baleines franches ainsi qu'à celui de la distinction de leurs vocalisations avec celles des rorquals à bosses.

1. INTRODUCTION

The calls of marine mammals are highly variable, and even those species with prima facie stereotyped call repertoires, such as blue whales, have proved to be more varied under wider scrutiny (e.g. Rivers 1997, Stafford *et al.* 1999). Consequently, even a detection system directed at a single species must be capable of handling biological variation (intra- and inter-individual, geographic, seasonal etc.). Furthermore, the frequency contours of individual vocalizations can be complex and nonlinear. Both these points are illustrated in Fig. 1, which shows a sample of calls from a group of singing humpback whales (*Megaptera novaeangliae*).

In this intricate signal environment the signals of interest cannot usually be fully specified. Parametric models of sufficient flexibility and complexity to approximate real varying signals are rarely used in marine mammal bioacoustics. (The usual approach is to manually measure nonparametric features of the signal from a spectrogram, e.g. minimum and maximum frequency. These are useful quantities but don't represent the signals well.) What is more, even under a suitable model, the statistical distributions of the parameters are not generally available, because sampling the calls on a sufficiently large scale over time, space, behaviour etc. is difficult.

This paper describes a method for detecting marine mammals calls in which the calls being detected, or parts thereof, are approximated by constant-amplitude linear chirps. The received signals are in effect simplified by analyzing them in short sections; in this way a paramateric model *can* be specified. A detector can be devised based on the estimate of the signal amplitude, and the frequency-related estimates hold information useful for classification.

The performance of the system is examined by (a) simulations using synthetic signals and noise, with known properties (b) trials using real recordings of whale calls.

The investigations in this study were motivated by a project to examine the potential use of acoustic detections for detecting North Atlantic right whales (*Eubalaena glacialis*) (Gillespie and Leaper, 2001). Improved understanding of the whereabouts of these animals could reduce the high mortality rate due to ships and fishing gear. Computer assistance in acoustic detection and classification is very desirable with large volumes of data, and could also be used for remote sensing.

2. METHODS

2.1 General description of algorithm

The frequency contour of marine mammal tonal calls is usually nonlinear. However, subsequences of the data (frames) can be taken, and if they are sufficiently short, then the call contour in that frame will be approximated well by a linear chirp. When a signal is present, each data frame can be treated as possibly containing a 'partially known' signal (i.e., a linear chirp with unknown parameter values to be estimated). This framing approach is like that used in the short-time Fourier transform, but the underlying signal model is different.

A spectrogram of a synthetic tonal sound with a nonlinear frequency contour is shown in Fig. 2. The sound was divided into four parts, and each part was treated as a linear chirp. The parameters were estimated with the algorithm described in this paper; the chirps are overlaid in the figure.
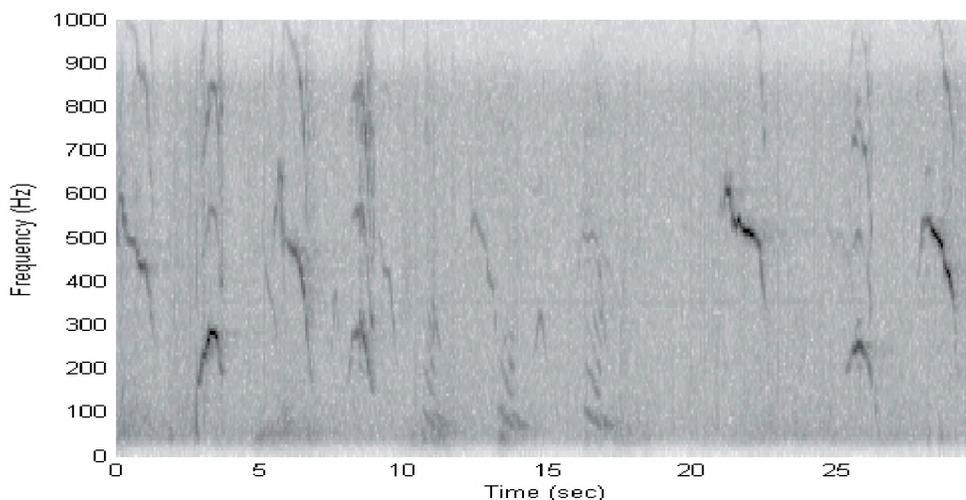
Figure 1. Spectrogram of a sequence of calls from a group of singing humpback whales (*M. novaeangliae*).
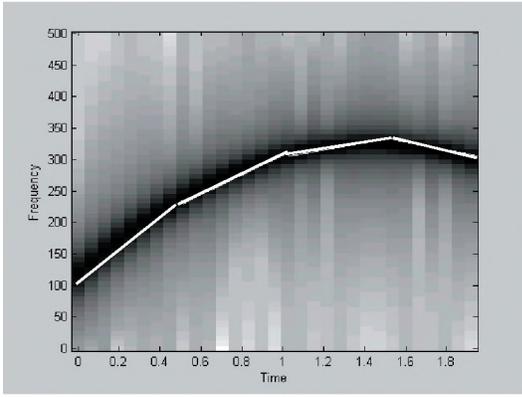
Figure 2. Spectrogram of a synthetic tonal signal with a nonlinear frequency contour, characterised by a sequence of four linear chirps (white lines).

If a chirp is present, its parameters may be estimated (amplitude, start frequency and chirp rate). The problem before us then is to find a good detector *for short signals and low SNR*. Together, these are particularly challenging conditions for detection. Boashash (1992) reviewed several methods under these circumstances, and with one exception, a maximum likelihood estimation (MLE) method worked best. This is the method used here.

<u>Detection and estimation in white noise</u>

In the case of white Gaussian noise (WGN), the combined (complex) signal and noise model is $\mathbf{y}=\mathbf{z} + \mathbf{w}$, where $\mathbf{y}$ is the observed signal and noise, $\mathbf{z}$ represents the underlying signal, and $\mathbf{w}$ the noise. The signal is modeled as a linear chirp so that:

$$z[n] = A \exp(j(a_0 + a_1\Delta n + a_2\Delta^2 n^2))$$

where Δ is the sampling interval; a_0, a_1 and a_2 are phase parameters; and A is the amplitude (assumed constant). The phase parameters are related to the start frequency f and chirp rate c by $f=a_1/2\pi$ and $c=a_2/\pi$.

The MLE estimates of a linear chirp in white noise are shown by Boashash (1992) to be obtainable using a 'dechirping' operation:

$$\max L(a_1, a_2) =$$

$$\max \left| \frac{1}{N}\sum_{n=0}^{N-1} y[n]\exp(-j[a_1\Delta n + a_2\Delta^2 n^2])\right|^2$$

(For those readers familiar with time-frequency analysis, there is a connection here with the Wigner distribution. Kay and Boudreaux-Bartels (1985) describe a detector in which the Wigner distribution of the observed data is integrated along all lines in the time-frequency plane. They show that this is the optimal detector for sufficiently long chirp signals in white noise. Li (1987) proves that the dechirping approach used here is equivalent.)

An example likelihood surface from a 1024-point synthetic chirp signal is shown in Fig. 3.
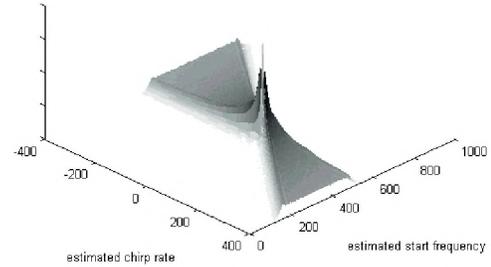
Figure 3. The likelihood surface from a 1024 point frame, for a synthetic, constant-amplitude linear chirp with a start frequency of 500 Hz and chirp rate of 100 Hz/second.

The likelihood can be maximised numerically over a grid of frequency and chirp rate values. The allowable combinations form a parallelogram: values chosen outside of this region would lead to negative frequency values. This region is given by:

$$-\frac{f}{T} \leq c \leq \left(\frac{f_s}{2T} - \frac{f}{T}\right)$$

where f_s is the sample rate, T is the duration of frame (seconds).

However, for reasons of efficiency the algorithm searches the entire rectangular region using FFTs:

$$\hat{a}_1, \hat{a}_2 =$$

$$\arg\max_{a_1,a_2}(\left| \frac{1}{N}\sum_{n=0}^{N-1} y[n]\exp(-j[a_1\Delta n + a_2\Delta^2 n^2])\right|$$

$$= \arg\max_{a_1,a_2}(\left| DFT(y'(a_2))\right|^2)$$

which uses the maximum of the discrete Fourier transform (DFT) of the dechirped signal $y'[n;a_2]=y[n]\exp(-ja_2\Delta^2n^2)$. Although the search is over the entire rectangular region, the likelihood values in the out-of-bound (negative frequency) regions will be small.

The MLE estimate of A is given by (Boashash 1992):

$$\hat{A}_1 = \left| \frac{1}{N}\sum_{n=0}^{N-1} y[n]\exp(-j[\hat{a}_1\Delta n + \hat{a}_2\Delta^2 n^2])\right|$$

but a better estimate of amplitude (see results, especially fig. 7) is:

$$\hat{A}_2 =$$

$$\max_{a_1,a_2} \left| \frac{1}{N} \sum_{n=0}^{N-1} y[n] \exp(-j[a_1\Delta n + a_2\Delta^2 n^2]) \right|$$

Detection and estimation in coloured noise

In the more general and probably more realistic case of coloured noise, we have the signal and noise model $\mathbf{y}=\mathbf{z} + \mathbf{w_c}$, where $\mathbf{w_c}$ is now distributed as $N(\mathbf{0},\mathbf{C})$, where $\mathbf{C}$ is the sample autocorrelation matrix of the background noise. The spectrum of coloured noise is not flat.

In this case, a whitening approach can be used as described by Kay (1993). The matrix $\mathbf{C}$ is factored to give a 'prewhitening matrix' $\mathbf{D}$ such that $\mathbf{C}^{-1}=\mathbf{D}^T\mathbf{D}$. Applying $\mathbf{D}$ to the combined signal and noise:

$$\begin{aligned} \mathbf{Dy} &= \mathbf{Dz} + \mathbf{Dw_c} \\ &= \mathbf{z'} + \mathbf{w'} \end{aligned}$$

where $\mathbf{w'}$ is distributed as $N(\mathbf{0},\mathbf{I})$, i.e. the noise has been whitened. The resulting signal $\mathbf{z'}$, the parameters of which are now estimated, is a distorted version of the original. The potential advantage of prewhitening is that the flat noise spectrum means that consecutive estimates when the signal is absent or weak will not be correlated due to correlation in successive noise spectra.

Efficiency of estimators

MLEs are usually statistically efficient, i.e. as N tends to infinity and at a sufficiently high SNR, the estimates are unbiased and have variances that attain the Cramer-Rao Lower Bound (CRLB) (a lower bound on the variance of all unbiased estimates). In other words, under these ideal conditions, the resolution of the MLE is as high as can be. Peleg and Porat (1991) give the CRLB variance bounds for constant-amplitude linear chirps in white noise.

DFT-based estimates, on the other hand, do not attain this resolution even under high SNR, high N conditions (see e.g. Boashash 1992, fig 3). Of course, whether high resolution is required for a classification problem needs to be assessed on a case-by-case basis.

However, in the present study the estimates do not necessarily have these desirable MLE properties, because of the following conditions used:

(i)   N is not large because short frames are used to obtain a linear approximation to the frequency contour.
(ii)   The SNR can be low for long-range signals.
(iii)   The use of the DFT for computationally efficient estimation of frequency introduces a limit to the resolution.
(iv)   With the prewhitening transformation of a signal in coloured noise, the underlying signal is distorted.

(v)   The assumption of constant amplitude is almost certainly violated.

The performance of MLE estimators at small N cannot usually be found analytically (Kay 1993). Simulations were therefore carried out using synthetic signals and noise to provide an understanding of the bias of the estimators as a function of SNR and N.

After detection of individual chirps, sequences of chirps can be concatenated to represent the entire signals or calls. The information can then be used jointly for call detection (as opposed to chirp detection) or call classification.

2.2 Testing of performance

The following simulations used 1000 runs each and a sample rate of 2kHz. Results were obtained as a function of signal-to-noise ratio (SNR) and signal length (N). The SNR is defined as A^2/σ^2, where σ^2 is the noise variance.

Bias of estimators

Synthetic linear chirps (start frequency 100 Hz, chirp rate 200 Hz/second) were created and embedded in WGN. The biases in estimates of start frequency, chirp rate and amplitude were examined.

Detection using synthetic signals and noise

Synthetic linear chirps were created and embedded in noise. The parameter values were randomly chosen in such a way that there were no negative-going frequencies. Some examples are shown in Fig. 4. The algorithm was used to estimate the signal amplitude, and this was used as a test statistic for detection. The probability of false and true detections was determined by simulation.

Two types of noise were used: WGN and a coloured spectrum based on a sample from the ocean. Synthetic noise, with a similar magnitude spectrum to the sample of ocean noise, was generated by passing white noise through a specially designed filter. An example power spectrum is shown in Fig. 5. The filter was designed using the Yule-Walker method (MATLAB signal processing toolbox).
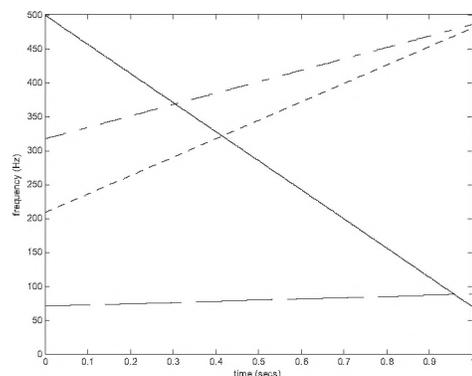
Figure 4. Examples of four linear chirp signals with randomly chosen phase parameter values. The parameter values are chosen so that there are no negative-going frequencies.
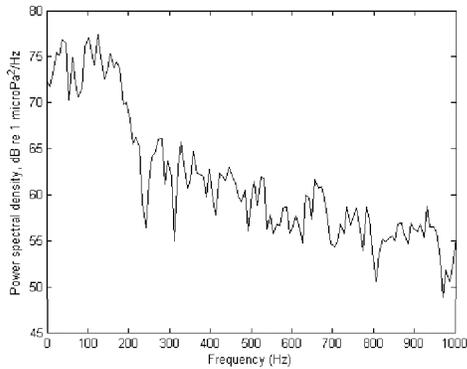
Figure 5. Power spectrum of a sample of synthetic ocean noise.

Detection using real recordings

In order to allow a dynamic response to changing noise levels and spectra in real noise conditions, the noise variance (σ^2) and the prewhitening matrix were re-estimated periodically, based on a block of sound. The signal samples used to calculate these quantities are based on order statistics.

Let

$\{s_i\}$, i=1,...,m be a set of sound vectors each of length N, selected randomly from the current sound block.

$v_i = var(s_i)$, the variance of vector i

e_j and e_k = the jth and kth percentiles of the distribution of $\{v_i\}$.

e_j provides an estimate of σ^2. The sample vector corresponding to e_k is used to estimate the prewhitening matrix **D**.

Table 1 shows the parameter settings used for the analysis presented in this paper. In addition, files were high-pass filtered at 50 Hz. The allowable chirp rate was set between +/- 250 Hz per second and searched over 100 points. Chirps were detected and estimated with start frequency up to 800 Hz. These settings were chosen after some initial experimentation showed they produced reasonable fits to call contours (see Fig. 11) and few false detections on the spectrogram, but they are in no sense optimal.

Table 1. Parameter settings for analysis of real recordings used in this study.

Parameter	Description	Value
f_n	Sampling rate (per second) for noise blocks	1/60
m	Size of set of sample noise vectors	400
j	Percentile for noise estimation	0.9
k	Percentile for prewhitening matrix	0.5
N	Frame length	512
H	Hop length	256

After detection and estimation of chirps, sequences of contiguous chirps that did not differ in frequency by more than 30 Hz were combined into detected 'calls'. The characteristics of calls from recordings of right whales, humpback whales, and a recording thought to be free of both, were then compared. Information about these recordings is shown in Table 2.

The comparison of humpback and right whale call characteristics showed that right whales commonly produced upsweeping calls. An upsweep detector was then made; this simply selected the subset of upsweeps from the detected calls.

The upsweep detector was optimised as a function of the SNR threshold in the following sense. The number of upsweep detections was maximized in recordings where right whales were known to be present, and simultaneously minimized in the recording where right whales were thought to be absent.

3. RESULTS

3.1 Simulations using synthetic signals and noise

Bias of estimators

Fig. 6 shows the bias in frequency and chirp rate as a function of N and SNR, from simulations using a synthetic chirp in WGN. Below a threshold SNR for all N, bias increases rapidly. Above the threshold, the bias is generally small.

Table 2. Information about recordings used for analysis.

Location	Date	# hours recorded	No. of channels	Human listening	Visual survey
Cape Cod Bay	16/3/01	4	3	Many right whales heard.	Right whales seen on 17/3/01.
Great South Channel	26/5/00	4	6	Right whales and humpbacks heard.	Right whales seen on 26/5/00.
Great South Channel	01/05/01	12	1	Few right whale calls heard by human listeners; of these, none definite.	None
Great South Channel	16/5/00	4	6	Many humpbacks heard. No right whales heard.	None
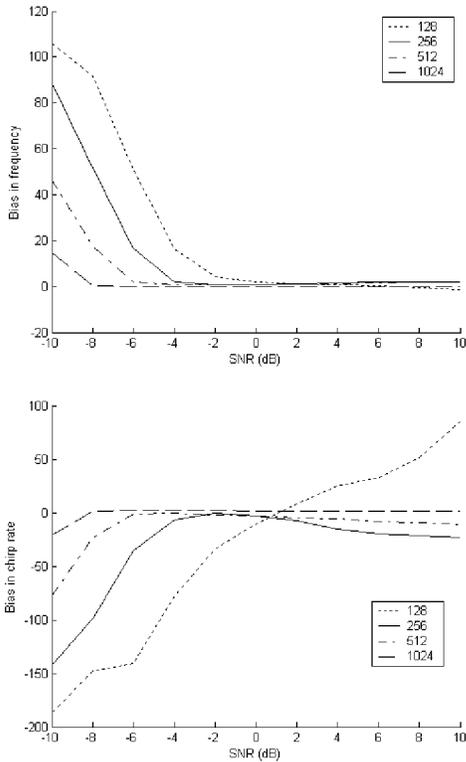
Figure 6. Simulation results showing biases of start frequency (top) and chirp rate (bottom) for linear chirps in WGN, as a function of SNR and N. When the number of samples and SNR are sufficiently high, the bias approaches (but does not necessarily reach) zero. The threshold SNR decreases as N increases. N is shown in the legend.

Figure 7. Simulation results showing biases of two estimates of signal amplitude: $\hat{A}_1$ (top) and $\hat{A}_2$ (bottom), for linear chirps in WGN, as a function of SNR and N. See text for details of these estimators. $\hat{A}_2$ has less bias and appears to stabilise as SNR increases. N is shown in the legend.

The bias of the amplitude estimator $\hat{A}_1$ is shown in Fig. 7 (top), and does not approach zero for these values of N and SNR. The estimator is sensitive to the biases in the estimates $\hat{a}_1$ and $\hat{a}_2$. The bias of estimator $\hat{A}_2$ is shown in Fig. 7 (bottom). The bias stabilises above a threshold SNR and is relatively small. Therefore, $\hat{A}_2$ is used for detection purposes in this study.

Detection performance

The performance of the detector on random linear chirps is shown in Fig. 8. The results show that, as anticipated, the increase in N leads to improved performance at low SNR. However, the real cost of increased N is hidden in these simulations. Real signals usually have a nonlinear contour, and increasing N in that case would make the linear approximation poorer, leading to a trade-off in performance.

Figure 8. Performance of detector on synthetic linear chirps in WGN, for N=512 (top) and 1024 (bottom). Detection performance improves as N increases. SNR (dB) is shown in the legend.

Fig. 9 shows the results for a detection of linear chirps in WGN by peak-picking the DFT. Performance is poorer than the chirp detector. However, these results may exaggerate what would happen using real signals, because modulation rates are likely to be more limited.

The performance of the detector in coloured noise with and without prewhitening is shown in fig. 10. Prewhitening in this case improves the detector.
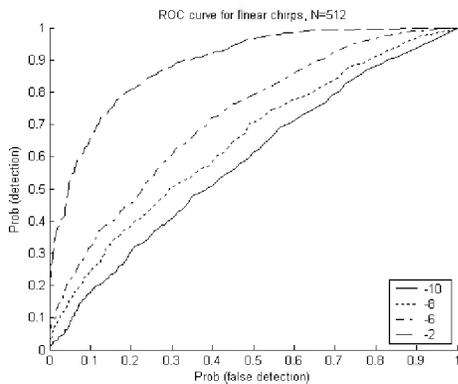
Fig. 9 Detection performance by peak-picking the DFT using synthetic linear chirps in WGN (N=512). SNR (dB) is shown in the legend.
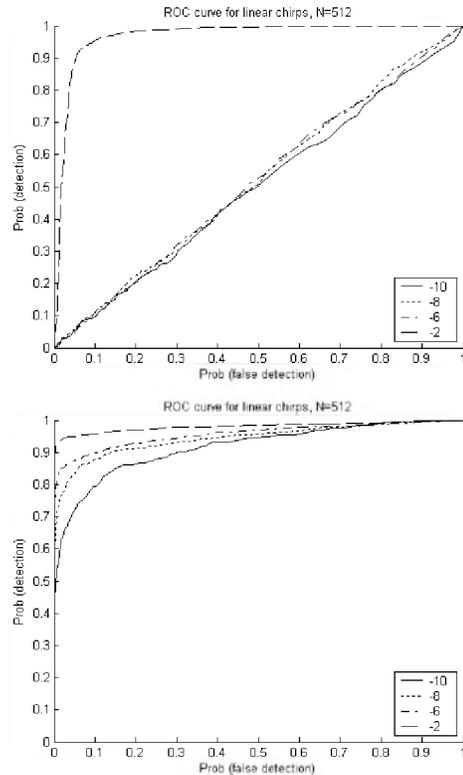
Figure 10. Detection performance in synthetic ocean noise with (bottom) and without (top) prewhitening. SNR (dB) is shown in the legend.

3.2 Tests using real recordings

The spectrogram of an underwater recording is shown in Fig. 11, with the detector results overlaid in white. In the figure are the sounds of a disk drive, two right whale upsweeps and another, more complex whale call. The detector has found and estimated short chirps and these are shown in white. After detection and estimation of chirps, nearby chirps are joined together into calls or sounds as described in the methods section.

The start frequency and sweep of calls from the recordings of right whales and humpbacks is shown in Fig. 12. A cloud of points representing upsweeps is evident from the right whales at about 50-150 Hz start frequency and sweeping up by about 30 Hz or more. This type of call is in fact well known from previous studies as a low-frequency 'contact' (upsweep) call (Clark 1982, McDonald and Moore 2002). As the scatter of points in Fig. 12 shows, they vary considerably in start frequency and sweep rate. Some example spectrograms are shown in Gillespie (2004). Calls with these characteristics were not found in the humpback recording.
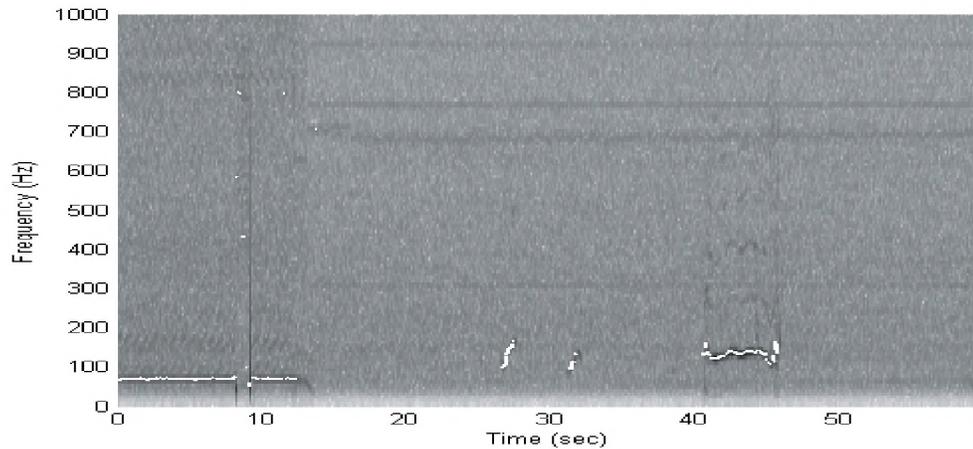
Figure 11. Illustration of the detection and estimation algorithm. The spectrogram of the received sound is shown with the detected and estimated chirps overlaid in white. The algorithm has detected the sound of a disk drive over the first 12 seconds; right whale upsweeps at about 26 and 32 seconds; and a more complex call at about 40 seconds.

There is considerable overlap elsewhere in the parameter space, though humpback signals with start frequencies >200 Hz tend to have greater sweep magnitudes. A large proportion of the signals with small sweep are false detections of ambient noise. In fact, the magnitude of the frequency modulation is a good indication of a biological signal.

The results show that, at least in the recordings tested here, upsweeps are a common and fairly distinctive right whale call. An upsweep detector was created to find calls with start frequency between 50 and 200 Hz, and with an end frequency at least 30 Hz higher than the start frequency (i.e. >30 Hz upsweeps). The duration of the signals was between 0.5 and 3 seconds. It is possible, however, that calls with inflexions in the contour may also be selected i.e., which are not strictly upsweeps.

The results of applying the upsweep detector to real recordings are shown in Fig 13. The figure shows the number of upsweeps detected per hour per channel as a function of threshold SNR. The y-axis in each plot shows the rate from a recording when right whales were present, and the x-axis shows the rate from a recording where right whales were thought to be absent. As the SNR threshold is decreased, more right whale calls are detected (y-axis) but there are also more false detections (x-axis). The appropriate threshold setting depends on the false detection rate required by the application. In the case of the northern right whale, a very low false alarm rate is likely to be required.

The effect of the prewhitening process is also shown in Fig. 13. Prewhitening appears to reduce the performance of the detector, except at high SNR thresholds in the Cape Cod Bay recording. No general conclusions should be drawn about using prewhitening though: its usefulness or otherwise will depend on the nature of the noise; and furthermore only one particular setting was tested here (Table 1).
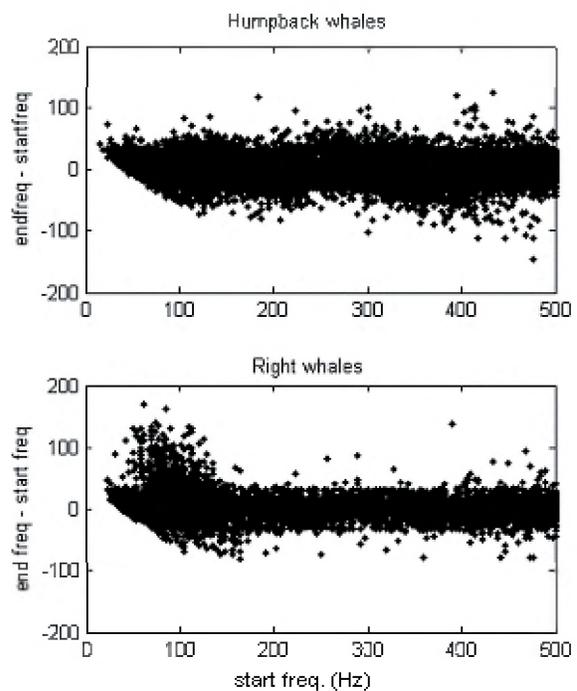
Figure 12. Scatterplot of start frequency versus sweep for calls of duration between 0.5 and 3 secs with an SNR threshold of –6 dB. The top plot is from the recording of humpback whales (Great South Channel 16/5/00) and the bottom from right whales (Cape Cod Bay 16/3/01). There are many upsweeping calls from right whales.

4. DISCUSSION

In an ideal detection environment of known, fixed signals, template-based detection methods such as matched filters perform well. In biologically realistic environments, however, signals are highly diverse and variable, and often only partially known. Methods are required which handle signals of this nature. The problem of detection and estimation of real animal signals will often be intractable

under a fully parametric approach, because this requires (i) a sophisticated and flexible parametric signal model, and (ii) large, high-quality datasets for determining the statistical distributions of the parameters. The system described here attempts to simplify the problem somewhat by 'forcing' FM tonal signals to be approximately chirp-like using a short data frame. The result is a highly flexible system, which allows any slow-varying frequency-modulated signal to be detected and characterised as a sequence of linear chirps.

4.1 Detection and estimation of linear chirps

A problem introduced by the deliberately short frames is that detection performance generally falls with signal length. Short signal length and low SNR are particularly challenging detection conditions. Nevertheless, performance may still be superior to techniques based on the DFT because the underlying model in that case is that of stationary (sinusoidal) signals, while the signals of interest are known to be frequency-modulated. The simulation results appear to confirm this.

It is likely that the method will usually give a reasonable fit to the actual phase of the original signal (with sufficiently high SNR) because of the linear approximation. There are biases because the data frames are short (i.e. N is not large), but these biases are small.

However, the model assumes a constant amplitude signal, which is certainly incorrect. Real signals will have at the very least a finite attack and decay. No results have been obtained in this paper on the effects of amplitude modulation.

A potential disadvantage of the method described here is that it only searches for the global maximum of the likelihood surface, and therefore can only detect a single signal in any frame. Generalising the technique for use with multicomponent signals should be possible but would require more processing. The method of simulated annealing may be a useful way of finding multiple local maxima (Press *et al.*, 1992). The current inability of the algorithm to detect multicomponent signals means that it is not practical for analysing the whistles of wild groups of odontocetes, in which simultaneous calls are common. But the call rate of baleen whales, with the exception of singing humpbacks but including right whales, is generally low and simultaneous calls may be unusual.

Further signal processing may be required to eliminate detections of broadband sounds. The algorithm will trigger on these but they may be eliminated using a further measure of spectral concentration.

Extending the present estimation method for use with a higher order polynomial is fairly straightforward but necessitates further computational costs. Some initial simulations have also indicated that the detection performance deteriorates with this type of model, although further investigation is needed. In comparisons of methods for estimating nonstationary signals, Boashash (1992) found that for estimation with short signals and low SNR, the 'cross-Wigner-Ville' method outperformed the ML method, but this method requires more intensive processing.
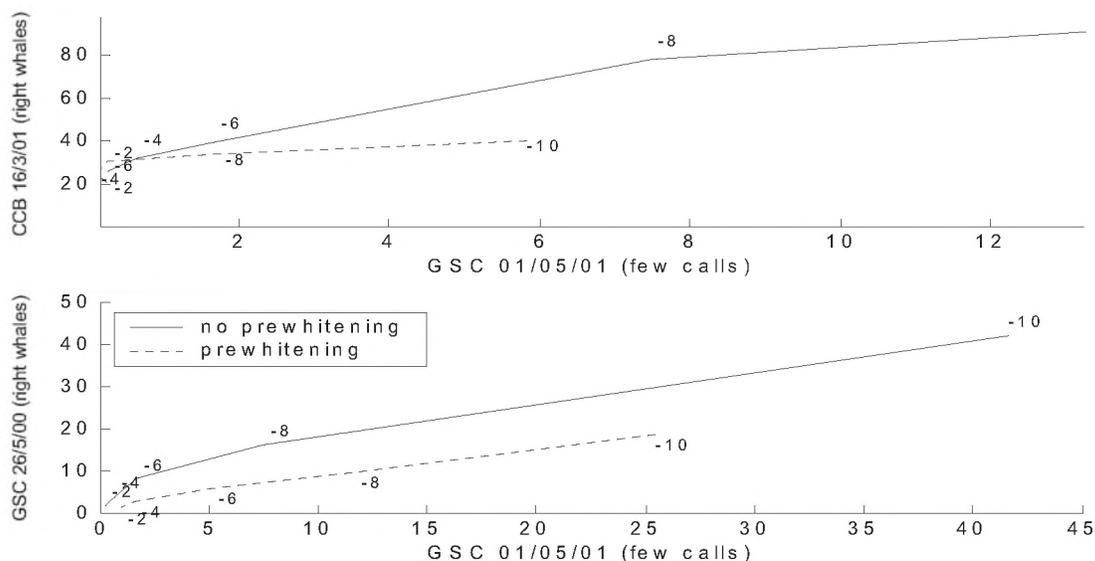
Figure 13. Performance of the upsweep detector using real recordings. Numbers of calls per channel per hour are shown for the two workshop test datasets (y-axes), in which right whale calls are present, against a dataset thought to be free of right whales (x-axis). The curves show the effect of varying the SNR threshold from –2 to -10 dB, and applying or not applying the prewhitening procedure. The characteristics of the calls are described in the text.

4.2 Formation of calls

Two further processes are required after the detection/estimation of chirps. Firstly, a method of forming calls from sequences of linear chirps is needed. Secondly, classification requires the allocation of subsets of calls with the required characteristics to the appropriate classes. In this study, these secondary processes were purposefully kept simple. Contiguous chirps that did not differ in frequency by more than some upper limit were combined into calls, and calls with appropriate start and end frequencies and durations were selected. Both of these processes could easily be improved to make further use of phase and amplitude information, perhaps also using continuity conditions.

4.3 Application to right whale detection

The numerical representation using the linear chirp model is very compact. In the example of Fig. 2, a ~2 second signal (2048 samples at 1000 Hz sample rate) was characterized by four 512-point linear chirps, requiring storage of 12 numbers (4 chirps x 3 estimates). This reduced the original digitized sound information by a factor of the order 100. In practical terms, satellite transmission is much more feasible after processing in this way.

In this study, right whale detection has focused on the relatively simple (yet variable; see Fig. 12) contact call or upsweep. Recent studies, including this one, have now shown this type of call to be reasonably common from northern right whales (McDonald and Moore 2002, Laurinolli *et al.* 2003).

For this particular signal type the 512-point linear chirp model, as applied in the present paper, may not be ideal. Firstly, a longer chirp model ought to provide better detection performance if the linear approximation to the contour holds fast. Secondly, the search space of the algorithm could be reduced to only include the region corresponding to upsweeps.

A more realistic model for variable upsweeps is a polynomial phase signal. Clark (1982) estimated polynomial coefficients by regression on digital images of spectrograms and used these as descriptors of call shape in southern right whales. Since then, computer functionality has developed considerably to allow this kind of parametric approach. For right whale upsweeps, an extension of the method used here to higher-order polynomials ought to be possible, though the additional computational cost could cause practical problems.

On the other hand, there are advantages to using the general, short chirp detector applied in this paper. Although not optimal for detecting longer upsweeps, information on other signals present over a larger frequency range (50-800 Hz, say) could be used to identify characteristics of humpback calls. This more general approach might, for example,

provide suitable quantitative information within which to search for periodic, similar signals, and for practical purposes these may be sufficiently distinctive features of humpback song.

REFERENCES

Boashash, B. 1992. Estimating and interpreting the instantaneous frequency of a signal - part 2: algorithms and applications. *Proc. IEEE* 80: 540-568.

Clark, C. 1982. The acoustic repertoire of the southern right whale, a quantitative analysis. *Animal Behaviour* 30: 1060-1071.

Gillespie, D. 2004. Detection and classification of right whale calls using an 'edge' detector operating on a smoothed spectrogram, Canadian Acoustics, Vol 32 #2, June 2004.

Gillespie, D. and Leaper, R. 2001. Report of the workshop on right whale acoustics: practical applications in conservation. Paper SC/53/BRG2 presented to Scientific Committee of International Whaling Commission, London, 2001.

Kay, S. 1993 *Fundamentals of Statistical Signal Processing I: Estimation Theory.* Prentice-Hall, New Jersey. 595 pp.

Kay, S., and Boudreaux-Bartels, G. 1985. On the optimality of the Wigner distribution for detection. *IEEE ICASSP Proc.* 3:1017-1020.

Laurinolli, M., Hay, A., Desharnais, F. and Taggart, C. 2003. Localization of North Atlantic right whale sounds in the Bay of Fundy using a sonobuoy array. *Mar. Mamm. Sci.* 19:708-723.

Li, W. 1987. Wigner distribution method equivalent to dechirp method for detecting a chirp signal. *IEEE Trans. Acoust. Speech and Sig.Proc.* 35: 1210-1211.

McDonald, M. A., and Moore, S. 2002. Calls recorded from North Pacific right whales (*Eubalaena japonica*) in the eastern Bering Sea. *J. Cet. Res. Manage.* 4: 261-266.

Peleg, S., and Porat, B. 1991. The Cramer-Rao lower bound for signals with constant amplitude and polynomial phase. *IEEE Trans. Sig. Proc.* 39: 749-752.

Press, W., Teukolsky, S., Vetterling, W. and Flannery, B. 1992. *Numerical Recipes in C.* CUP.

Rivers, J.A., 1997. Blue whale, *Balaenoptera musculus,* vocalizations from the waters off central California. *Mar. Mamm. Sci.* 13: 186-195.

Stafford, K., Nieurik, S., and Fox, C. 1999. Low-frequency whale sounds recorded on hydrophones moored in the eastern tropical Pacific. *J. Acoust. Soc. Am.* 106: 3687-3698.

ACKNOWLEDGEMENTS

THE REAL-TIME DETECTION OF THE CALLS OF CETACEAN SPECIES

E J Harland[1] , M S Armstrong[2]
[1]QinetiQ Ltd, Winfrith Technology Centre, Dorchester Dorset, UK. DT2 8XJ.
[2]Kaon Ltd, 5 Wey Court, Mary Road, Guildford, Surrey, UK. GU1 4QU

ABSTRACT

The UK is currently implementing the next generation of military surface ships active sonars. As part of the environmental protection work associated with the new sonars, and also in support of existing sonars, the UK Ministry of Defence has funded a programme of work to improve the capability to detect, classify and localise marine mammals. QinetiQ Ltd have been contracted to provide a software package, which can process the raw acoustic data from a number of sonar systems, including towed arrays, hull-mounted arrays and sonobuoys. The software needs to be able to adjust to the local environment and provide a cetacean presence/non-presence decision in real-time and with a very low false alarm rate. The first version, running on a standard PC, has now been completed and tested at sea during the NATO SIRENA 03 cruise. This paper describes the processing method employed and the results achieved during testing using a number of datasets.

RÉSUMÉ

Le Royaume-Uni implante présentement la prochaine génération de navires de surface avec sonar actif. En ce qui a trait au travail de protection de l'environnement associé avec les nouveaux sonars, et aussi en support aux sonars existants, le MoD a fondé un programme de travail visant l'amélioration des capacités de monitorage des vaisseaux de la marine royale afin de détecter, classifier et localiser les mammifères marins. Le QinetiQ Ltd a été mandaté afin de fournir un système pouvant traiter les données acoustiques brutes à partir d'un certain nombre de systèmes sonar, incluant les réseaux remorqués, les réseaux montés sur coque, et les bouées acoustiques. Un tel système doit pouvoir s'ajuster à l'environnement local et doit fournir une décision concernant la présence/non-présence de cétacés en temps réel avec un très faible taux de fausse alarme. La première version peut être utilisée sur un PC standard, et a été complétée et testée en mer durant la croisière expérimentale SIRENA 03. Cet article décrit la méthode d'évaluation employée et les résultats des tests en laboratoire en utilisant un certain nombre d'ensembles de données.

1. INTRODUCTION

The UK is currently implementing the next generation of military surface ship active sonars. The Ministry of Defence (MoD) has recognised the potential problems associated with the use of active sonar systems and has in place a research programme looking at minimising the possible risk to the marine environment. As part of this work QinetiQ Ltd has been tasked to provide a software package, which aims to detect, classify and localise the calls of all marine mammals. The first version of this package is known as the Marine Mammal Automated Detection System (MMADS) and is reported in this paper. Kaon Ltd were contracted by QinetiQ to write the real-time implementation. This initial version does not fully implement echolocation pulse processing and only implements the detection and classification aspects.

It is recognised that animals may not always vocalise so this package will be part of a broader system, which will integrate marine mammal detections from visual, infra-red and radar sensors and combine these with the passive acoustic detection data to provide a twenty four hour detection capability for marine mammals.

The research described here aims to detect the calls from all marine mammals. The method used extracts parameters from the detected sound that allow marine mammal calls to be distinguished from other natural and anthropogenic sounds. While the software aims to recognise calls from cetaceans, pinnipeds and sirenia, it also needs to be able to identify sounds from sonars and other

anthropogenic sources in order to eliminate these sounds during the classification process.

2. BACKGROUND

Previous attempts to produce automated detection software have been reported by a number of groups. This earlier work (e.g. Potter et al. 1994; Sturtivant and Datta 1997a; Sturtivant and Datta 1997b; Mellinger and Clark 2000) looked at recognising individual species or closely-related groups of species. This work has normally been associated with behavioural research or animal censusing where it is important to be sure that the calls are from just the required species. Other workers in the field have used marine mammal calls to demonstrate the efficacy of novel signal processing algorithms (e.g.Helweg and Moore 1997; Tiemmann et al. 2001) or to demonstrate better/faster signal processing hardware (e.g.Jones et al. 1997). A number of workers have investigated the possibilities of using a number of call parameters to identify species (e.g. Wang et al. 1995; Oswald et al. 2003).

Some of the currently available software packages for cetacean acoustic research include an automated detection capability. ISHMAEL, produced by NOAA/PMEL (Mellinger 2002; Mellinger 2004), includes three forms of automated detection software. Energy summation sums the energy across all frequencies over a limited range in time and is useful for detecting echolocation pulses. The matched filter method correlates the incoming signal with a user generated reference waveform and is useful for searching for signals with little variation from pulse to pulse. The spectrogram correlation method (Mellinger and Clark 2000) cross correlates the spectrogram with a time-frequency kernel. This method is more tolerant of variability in the call than the matched filter method, but still requires that the expected signal be constrained within fairly tight limits.

RAINBOW CLICK is available from IFAW (IFAW 2004) and is used to detect and track sperm whales (Gillespie and Chappell 1998). The software processes the incoming datastreams from two hydrophones to identify sperm whale clicks and then uses the time delay between the hydrophones to estimate the bearing of each pulse. The results are then presented to an operator for manual interpretation.

WHISTLE, also available from IFAW, searches the incoming acoustic datastream for whistle calls. The contours of the whistles are then displayed to an operator and stored in a file for further analysis.

All of the above packages were designed to assist cetacean research and this constrains their potential use in a military environment where operator involvement has to be kept to the absolute minimum and the processed data cannot be further interpreted. The MMADS package is designed from the outset to require the minimum possible operator intervention and to have a simple presence/non-presence indication as its output. The first version described here still has some research facilities i.e. the spectrogram displays and the file capture capability, in order to be able to assess how well the software is performing.

3. REQUIREMENTS AND ASSUMPTIONS

The automated detection system must be able to detect all marine mammal calls while achieving low false alarm rates and a high probability of classification. If the system is to be useful in the mitigation role it must give detections and classifications that the sonar operators and command team can trust. It must therefore detect at adequate range, classify in a timely and correct manner and with a very low false alarm rate. The software must also be able to operate with a variety of sonar sensors e.g. towed arrays, hull-mounted arrays and sonobuoys. It should also be capable of multi-sensor operation with only the minimum of re-configuration

The MMADS software package implements a number of algorithms which are based on assumptions about the calls of marine mammals. These are:-

- That animals rarely vocalise just once
- That there are sufficient vocalisations to discard crossing and corrupt calls and still be able to make a correct and timely decision
- That it is possible to identify features in the calls that allow the calls to be discriminated from all other sounds in the sea.

In preparation for this work a reference set of acoustic data has been assembled covering examples of the cetacean species groups defined below. This data has been gathered from a number of military sonar sensors, from research hydrophones and from pop-up and acoustic tag deployments. Additional data was also gathered in areas where it was known that no cetaceans were present and in areas of high levels of anthropogenic noise from a variety of sources. This data was used to visually check that the above assumptions were valid, to define the algorithms described below and then to set the limits for the parametric tests in the decision-making software. Further data became available when the implementation work was well advanced and this was used to test the implemented algorithms.

4. IMPLEMENTATION

MMADS processing splits the incoming acoustic signals into five processing channels appropriate for five groups of calls. These are:-

Odontocete echolocation	15-150 kHz
Odontocete tonal	1-22 kHz
Low-frequency echolocation	1-22 kHz
HF mysticete	150-1000 Hz
LF mysticete	10-150 Hz

For the first implementation of MMADS it was decided to use readily available commercial technology resulting in the choice of standard PC hardware with the Linux operating system and the C++ programming language. The audio input card was chosen to be a SoundBlaster-compatible audio input card.

Figure 1 MMADS Hardware

The audio card limited the sample rate to 48 kHz and this meant that the odontocete echolocation pulse processing channel was severely limited in capability as most of the available energy is above the 22 kHz bandwidth. For many of the species there is still sufficient energy remaining to make a classification decision but animals like the harbour porpoise (*Phocoena phocoena*), which transmit limited-bandwidth pulses centred on 140 kHz cannot be detected by this method. A fourth processing channel optimised for echolocation calls is being investigated and will be added in a later version of the software. The 22 kHz bandwidth limit will not accept the highest frequencies in whistles from some of the small odontocetes, but this represents a very small percentage of total energy available and is not considered a significant limitation on the performance of the detector.

Figure 2 shows a typical screen display produced by MMADS. The three spectrograms correspond with the three frequency ranges used by the processing software. Alongside each spectrogram window is a vertical bar, which displays the confidence of there being marine mammal calls in that frequency range. A background scrolling text window also displays the confidence level of the five processing chains. The top window is the 1-22 kHz window and the side bar is a combination of the confidence levels from the odontocete tonal, odontocete echolocation and sperm whale processing channels. The lower left window is the LF mysticete channel and the lower right window is the HF mysticete channel.
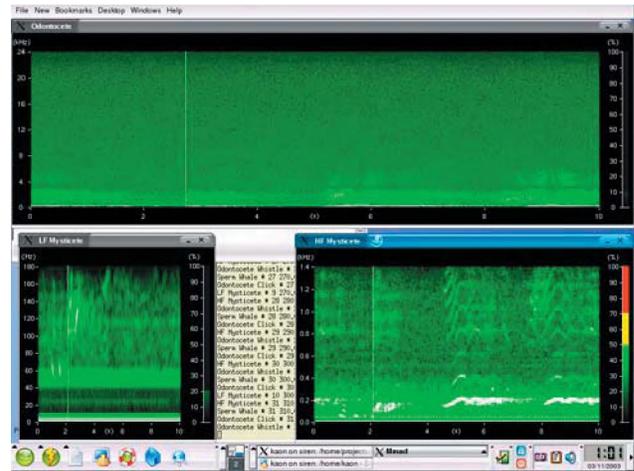
Figure 2 MMADS display

5. ALGORITHMS

5.1 Odontocete echolocation pulses

This class of signal includes the echolocation pulses emitted by animals ranging from common dolphins to killer whales. Incoming signals are digitised at a 48 kHz sample rate. If amplitude clipping is detected, the time sequence is rejected and a message generated in the text box. Good data is then transformed into the frequency domain using a 256 point FFT with Hamming window and 50% overlap. Frequency domain data is accumulated into 10 second patches. The patch is then normalised using the background mean. The resulting normalised data is converted to a binary spectrum using a fixed threshold 8 dB above the mean. The initial detection criteria is that 40% of points within the required bandwidth must exceed the threshold. The centre of the pulse is located in time using a split window test and the duration of the pulse has to be appropriate for echolocation pulses. The spectral slope is then tested to confirm the signals are echolocation pulses. A confidence level is calculated based on the number of echolocation pulses per patch.

5.2 Low-frequency echolocation pulses

For this class of signal, which is primarily clicks from sperm whales, all processing is as for the odontocete echolocation pulses except that the spectral test looks for the lower frequencies used by sperm whales.

5.3 Odontocete tonal

This class of signal includes the tonal signals emitted by animals ranging from common dolphins to killer whales. The incoming data stream, sampled at 48 kHz, is transformed to the frequency domain using a 1024 point FFT with Hamming window and 75% overlap. This spectral

data is accumulated into patches ten seconds long and then normalised using a mean value offset in time. The data is converted to a binary spectrogram using a threshold 8 dB above the mean. The binary data is then searched for connected components using an 8-connectivity neighbourhood. The first test is that each connected component must have at least twenty time-frequency bins.

Each connected component is then tested to extract the following parameters:

    i. Minimum frequency
    ii. Maximum frequency
    iii. Start frequency
    iv. Stop frequency
    v. Duration
    vi. Bandwidth
    vii. Instantaneous bandwidth
    viii. Area ratio
    ix. Porosity
    x. Mean centre

The area ratio is the ratio of the number of time-frequency bins in the signal to total number of bins within the rectangle formed by start and stop times and minimum and maximum frequencies.

$$AreaRatio = \frac{SignalArea}{TotalArea} \qquad 1$$

The porosity is the ratio of enclosed non-signal time-frequency bins to signal bins within the rectangle defined by min and max frequency and start and stop times.

$$Porosity = \frac{ZeroArea}{SignalArea} \qquad 2$$

The porosity value is used to reject complex signals such as crossing tones or noisy signals.

The measured parameters are then tested against expected values for odontocete tonals. Successful classifications are accumulated across three successive ten second patches and used to calculate the confidence levels.

5.4 HF mysticete

This class of signals includes tonal sounds emitted by humpback whales and some pinnipeds. The data, sampled at 48 kHz, is low-pass filtered at 1.3 kHz and then decimated by a factor of 16 to a sample rate of 3 kHz. This data is then transformed into the frequency domain using a 1024 point FFT and 75% overlap. In each FFT, 256 points are real, the rest are zeroes. The resulting spectra are accumulated into patches ten seconds long and then normalised using a 3 second window median normaliser. The output is then converted to binary using a 10 dB threshold. The rest of the processing is then as described for odontocete tonals in 4.3 above, except that the parameter testing is optimised for this class of sounds. Some calls from the mysticetes are much broader in bandwidth than pure tonals. Provided the porosity

value stays below the rejection threshold, the instantaneous bandwidth parameter allows this type of call to be classified.

5.5 LF mysticete

This class of signals includes the sounds made by fin, blue and right whales. The incoming data, sampled at 48 kHz is low-pass filtered at 140 Hz and then decimated by a factor of 128 to give a sample rate of 375 Hz. This data is then transformed to the frequency domain using a 512 point FFT with Hamming window. 64 points of real data are overlapped by 75%, the other 448 points in each FFT are set to zero. The resulting spectrogram is normalised using the median of a 5 second window with no offset and converted to a binary spectrogram using a threshold of 8 dB. The remainder of the processing is as described for the odontocete tonals in paragraph 4.2 above.

5.6 Operator display/control

The package also produces a display so that the operator can monitor the output from the MMADS software. Figure 2 shows a typical display. In addition, it is possible to automatically capture the raw audio into files of type WAV whenever a preset confidence level is exceeded. To allow evaluation of the algorithm there is a facility to capture the partially processed data in a form suitable for importing into MATLAB. This was used to produce the displays in Figures 3-8.

6. KNOWN DEFICIENCIES

Testing of the software against the reference dataset and using the NATO SIRENA 03 and DRDC/Dalhousie datasets has suggested that this initial version of the MMADS software has a number of deficiencies.

The algorithm to detect sperm whales was designed around a limited dataset obtained by using sonobuoys. Data subsequently obtained from wideband hydrophones suggests that the implemented algorithm, optimised for use with sonobuoys suffers from an unacceptably high false alarm rate on higher bandwidth data. This will be addressed in the next version of the software.

The odontocete echolocation pulse detector is missing the majority of available energy in the pulse. This means that the detection range for these pulses is significantly reduced compared with that achievable using the full bandwidth. The limitation is due to the choice of 48 kHz A/D converters. Use of the next generation of A/D cards operating at 192 kHz would alleviate this problem, although it would still not work with harbour porpoise pulses.

The odontocete tonal detector works well with animal calls and has successfully detected calls from a range of animals. However, when used in a sonar environment it can be partially activated by complex transmit sequences. A processing channel which identifies sonar waveforms and

removes them from the input to the odontocete tonal classifier is currently being investigated.

The testing of the HF mysticete channel has not been as extensive as the authors would like due to the lack of good quality data. Alternative sources of data are currently being sought to allow this testing to continue.

The LF mysticete channel was tested extensively against fin whales (*Balaenoptera physalus*) during the NATO SIRENA 03 cruise. This revealed an unexpected problem. The algorithm worked well with individual calls but when a group of animals started calling in the confines of the Ligurian Sea the reverberation levels built up to the point where the normaliser was suppressing all of the calls. The algorithm will need to be modified to ensure the normaliser does not cancel out continuous calling.

7. MMADS TEST RESULTS

DRDC and Cornell University issued two acoustic datasets gathered using autonomous seabed recorders for use by participants in *the Workshop on Detection and Localisation of Marine Mammals using Passive Acoustics* held in Halifax, Nova Scotia, in November 2003. These datasets featured right whale and fin whale calls.

The data was read into the ISHMAEL software to extract single channel data from the multi-channel data supplied. The single files were then processed using MMADS in file-input mode to process the data. The processed data is then extracted at three stages through the processing chain and displayed using MATLAB. The figures that follow illustrate this processing of the data. In each group of three figures the first figure is the raw spectrogram of the data. The second figure is the spectrogram overlaid with the detected connected-components. The third figure is the spectrogram overlaid with the connected components that the decision-making process has chosen as originating from marine mammals.
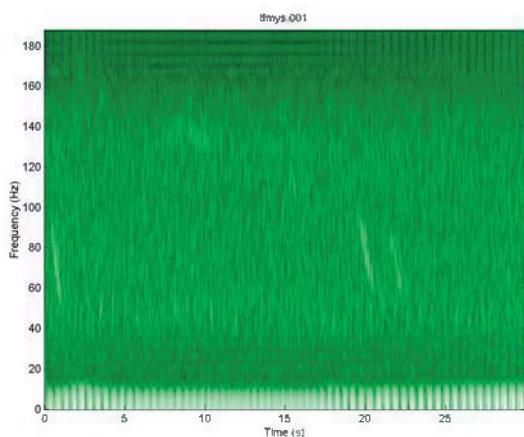
Figure 3 Raw spectrogram
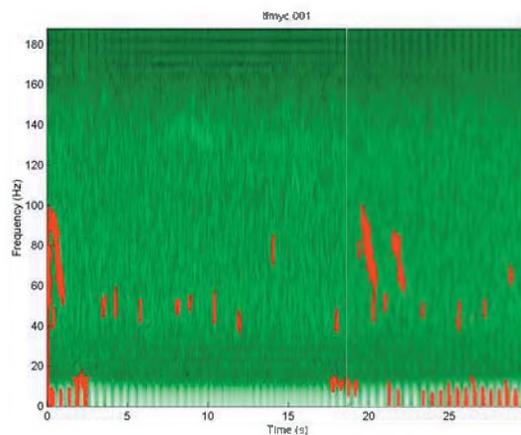
Figure 4 Connected components overlay
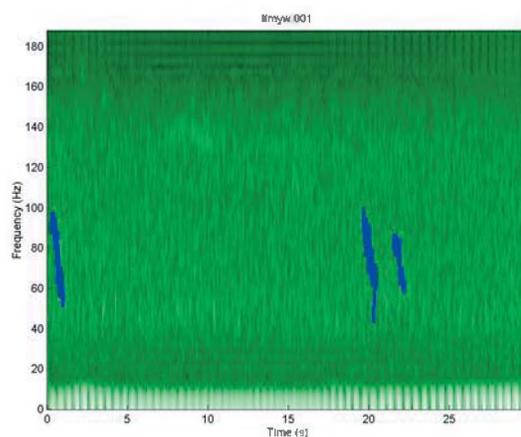
Figure 5 processed data

The white band at low frequencies appears to be cable strumming. The three pictures are visually inspected to determine a) how many animal calls there are in the raw spectrogram, b) how many animal calls are identified in the connected component (C-C) scan, c) how many non-animal sounds are identified in the C-C scan, d) how many animal call C-C are classified as marine mammals and e) how many non-animal C-C are classified as marine mammals. The data was processed using the LF Mysticete channel and both right whale and fin whale calls were processed. The table below includes the calls from both species.

Referring to the categories of detection (a-e) above, the results for a number of files from the Cornell data set are shown in Table 1 below.

The majority of the DRDC dataset was not suitable for processing through MMADS because the files were too short. Only one file is long enough to process and this is the L138b file. Again using the same detection categories (a-e) defined above the results for this file are shown in Table 2 below.

	Detection categories				
Data file	a	b	c	d	e
GSC0420	33	33	250	19	4
CCB0800	46	46	196	12	0
CCB0810L	37	36	182	16	1
CCB0810H	65	65	351	23	2
CCB0825	44	44	178	25	1
GSC0650	20	20	183	14	0
Totals	245	244	1340	109	8

Table 1. Analysis of Cornell data.

	Detection categories				
Datafile	a	b	c	d	e
L138b	168	128	276	68	22

Table 2. Analysis of DRDC data.

From these results it can be seen that just over 40% of calls are classified correctly while 3% of classified calls are false for the Cornell dataset and 10% for the DRDC data set.

8. SIRENA 03 DATA

During the NATO SIRENA 03 cruise in August/September 2003, the MMADS system was deployed and used to process hydrophone data from the SACLANTCEN towed array and the University of Pavia two-hydrophone array. This provided an opportunity to test the system against fin whales, sperm whales, striped dolphins and common dolphins. This data has yet to be fully analysed but Figures 6 - 8 show the sequence of data for a striped dolphins call. The vertical bars are echolocation pulses but in this example only the output from the odontocete tonal channel is shown.
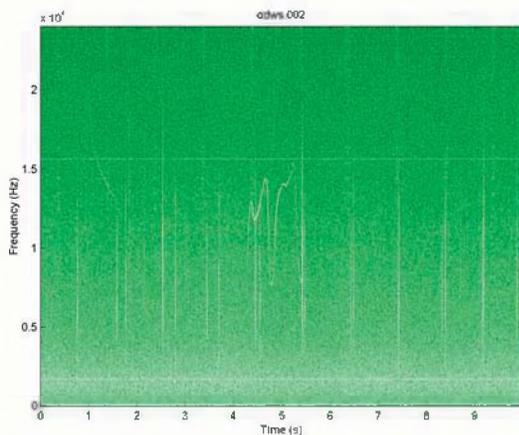
Figure 6 SIRENA 03 data spectrogram
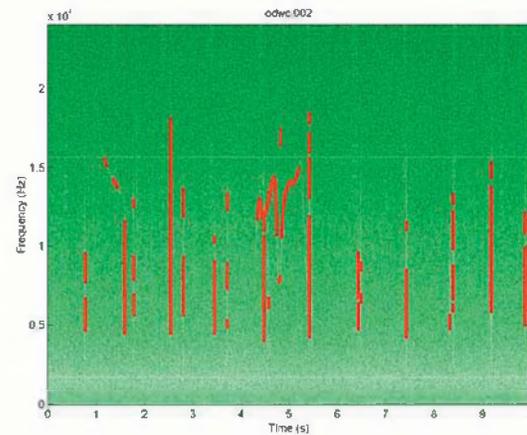
Figure 7 SIRENA 03 Connected components
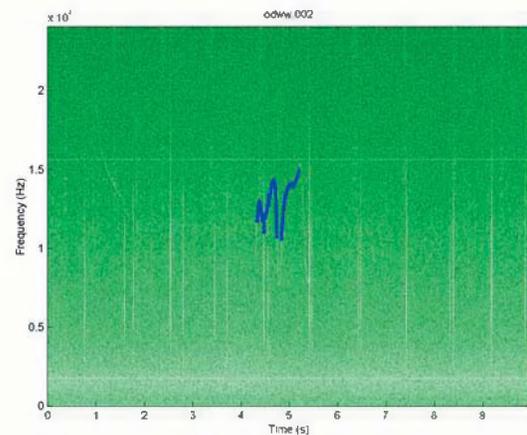
Figure 8 SIRENA 03 classified components

9. DISCUSSION

This work has demonstrated that algorithms based on classic spectral mapping techniques can provide a useful automated system for the detection and classification of marine mammal calls. Although the present system is constrained in performance by the limited reference data set used to choose the parameter tests, the system has still performed well in laboratory and at-sea testing. The 40% success at classifying calls is well within acceptable limits but the false alarm rate at up to 10% needs to be further reduced. The processing within the MMADS package to combine the individual call detections to form confidence levels provides this false alarm reduction but this has not been tested in detail yet and will be reported separately. It is still desirable to improve the individual call classification processing to reduce the load on the confidence calculation. This area will be addressed during the next phase of the work.

The reference data set has been expanded considerably since the MMADS processing was designed and will now

be reviewed to ensure the parameter tests can provide the correct classification for an increased range of species.

The current software detection and classification is based on parametric extraction from single calls, be they tonals or echolocation pulses. In the case of sperm whales this leads to high false alarm rates because of all the other sounds, both natural and anthropogenic, having very similar characteristics to sperm whale pulses. The main thrust of future work will investigate the possibilities of context processing in which the whole ensemble of calls is examined to further characterise the signals and enhance rejection of non-animal sounds. The acoustic background will also be examined to ensure there are no frequency or time effects, which could skew the classification results. To be useful an automated system needs to know how well it will perform in order to warn the user when the acoustic environment may adversely affect its ability to detect marine mammals.

If a very high quality reference dataset is available, such as those obtained by tagging studies or work with captive animals, and the acoustic environment parameters, such as sound velocity profile, surface roughness etc, are known then it becomes possible to refine the classification testing by predicting the characteristics of the calls likely to be received after propagation through the medium. This could include time dispersion due to multi-path or frequency dependent velocity of propagation, and frequency dependent absorption. This could significantly improve detection performance in some acoustic environments.

10. ACKNOWLEDGEMENTS

11. REFERENCES

Gillespie D, Chappell OP (1998) Automated cetacean detection and monitoring. In: Tasker ML, Weir C (eds) Proceedings of the seismic and marine mammals workshop, London, 23-25 June 1998. Sea Mammal Research Unit, St Andrews.

Helweg DA, Moore PWB (1997) Classification of aspect-dependent targets by a biomimetic neural network. Naval Command, Control and Ocean Surveillance Center, San Diego

IFAW (2004) Rainbow Click. IFAW, www.ifaw.org/ifaw/general/default.aspx?oid=35704.

Jones J, Browne RW, Di Meglio A, Wang LS (1997) Dolphin vocalisation analysis using an ADSP21020. Institute of Electrical Engineers, London.

Mellinger D (2004) Ishmael. PMEL/NOAA, http://cetus.pmel.noaa.gov/cgibin/getinfo.pl?dirname=ishmael

Mellinger DK (2002) Ishmael 1.0 User's Guide. NOAA/PMEL, OAR PMEL-120, Seattle, USA.

Mellinger DK, Clark CW (2000) Recognizing transient low-frequency whale sounds by spectrogram correlation. Journal of the Acoustical Society of America 107: 3518-3529.

Oswald JN, Barlow J, Norris TF (2003). Acoustic identification of nine delphinid species in the Eastern Tropical Pacific Ocean. Marine Mammal Science 19: 20-37.

Potter JR, Mellinger DK, Clark CW (1994) Marine mammal call discrimination using artificial neural networks. Journal of the Acoustical Society of America 96: 1255-1262.

Sturtivant CR, Datta S (1997a) Automatic dolphin whistle detection, extraction, encoding and classification. In: Goodson AD, Harland EJ (eds) Underwater bio-sonar and bioacoustics symposium. Institute of Acoustics, Loughborough University, pp 259-266.

Sturtivant CR, Datta S (1997b) Dolphin whistle classification with the "dolphin" software. In: Goodson AD, Harland EJ (eds) Underwater bio-sonar and bioacoustics symposium. Institute of Acoustics, St Albans, pp 149-156.

Tiemmann CO, Porter MB, Frazer LN (2001) Automated model-based localisation of marine mammals near Hawaii. In: Holland RC (ed) OCEANS 2001. Marine Technology Society, Washington, pp 1395-1400.

Wang D, Wursig B, Evans W. (1995) Comparison of whistles among seven odontocete species. In: Kastelein RA, Thomas JA, Nachtigall PE (eds). Sensory systems of aquatic mammals. De Spil Publishers, Woerden, The Netherlands. ISBN 90-72743-05-9. pp 299-323.

DETECTION AND CHARACTERIZATION OF MARINE MAMMAL CALLS BY PARAMETRIC MODELLING

A. T. Johansson[1] and P. R. White[2]

Signal Processing and Control Group, Institute of Sound and Vibration Research (ISVR),
University of Southampton, Highfield, Southampton SO17 1BJ, UK.
Email: tj@isvr.soton.ac.uk[1], prw@isvr.soton.ac.uk[2].

ABSTRACT

In this paper, we describe a parametric modeling method for detection and characterization of tonal signals and its application to marine mammal calls. The method tracks dominant frequencies with an adaptive notch filter (ANF), and couples this to a novel, simultaneous detection step. The detection statistic is derived from a measure of tracking reliability, obtained as a by-product of the tracking algorithm. Detection therefore comes at little extra computational cost from an algorithm that is fast, simple, and capable of dealing with multiple signals in low signal-to-noise ratios. Frequency estimates are derived directly from the time domain waveform, avoiding the resolution trade-off and other short-comings of the commonly used spectrogram. The performance of the algorithm is demonstrated on both simulated signals and recordings of right whale calls. The method is found to be noise robust and capable of extracting right whale and other calls with a low false alarm rate.

SOMMAIRE

Le présent article décrit une méthode de modélisation paramétrique pour la détection et la caractérisation des vocalisations tonales de mammifères marins. La méthode consiste à poursuivre les fréquences dominantes à l'aide d'un filtre à encoche adaptatif (ANF), couplé à une étape de détection simultanée innovatrice. La statistique de détection est dérivée d'une mesure de la fiabilité de poursuite, un sous-produit de l'algorithme de poursuite. La détection entraîne un coût computationnel additionnel minime, et est à la fois rapide, simple, et capable de traiter des signaux multiples et à de bas rapports signal sur bruit. Les estimations de fréquence sont dérivées directement du domaine temporel et fréquentiel, évitant ainsi les compromis de résolution de la technique spectrogramme utilisée fréquemment. Dans une application sur un fichier d'une durée de 18-min de l'ensemble de donnée de l'atelier, il est possible de détecter quatre vocalisations probables de baleines franches.

1. INTRODUCTION

Passive acoustic detection of cetaceans is an important and growing research field. The mitigation of several different threats to cetaceans, such as collisions with ships and ensonification by high-power active sonar, require the detection, localization and, ideally, identification of cetaceans in the vicinity. Visual observation requires daylight, a reasonably calm sea, and that the cetaceans are at the surface. These drawbacks do not apply to a system based on processing the cetaceans' vocalizations. Here, to be detectable an animal must vocalize, but this is not a major drawback as most cetaceans are highly vocal. One must also be able to process the calls in strong background noise and other difficult environmental conditions and handle the great variability of most cetacean calls.

The Workshop on Detection and Localization of Marine Mammals Using Passive Acoustics, held in Dartmouth, NS, 19-21 November 2003, answered to the growing need of collecting and comparing different algorithms for passive acoustic processing of marine mammal sounds. The conference focused on processing the sounds of the North Atlantic right whale, a critically endangered species of which there less than 300 individuals left [1]. Strong protection measures are already in place for this species, for example the US Marine Mammal Protection Act states that no vessel is allowed within 500 meters of a North Atlantic right whale [2]. To implement such

protection measures it is crucial to be able to detect right whales in the vicinity.

Accompanying the workshop was a dataset of several underwater recordings containing right whale calls. These recordings were all made on moored single hydrophones in the Bay of Fundy, a key North Atlantic right whale habitat. One file, called L-138, was specifically meant for testing detection algorithms, and will be used for this purpose here. The file is 18 min long and sampled at 1250 Hz.

In this paper, we use model-based signal processing to detect and characterize right whale sounds. Model-based signal processing is a popular research field that has found applications in many fields, such as biomedical signal processing, speech recognition and economic forecasting. The idea here is to apply a model, controlled by a small number of parameters, to the signal. Given old samples, the model produces a prediction of the next signal sample. By minimizing the difference between the model-predicted and the actual signal, we fit the model to the signal and force information about it into the model parameters. These can subsequently be used to characterize the signal, and, as we show here, also to detect occurrences of the signal immersed in broadband noise.

We use a specific type of model known as an adaptive notch filter (ANF), which expresses the prediction error by a filtering operation on the input signal. The transfer function magnitude response of the notch filter is that of a deep notch at one or more frequencies, and a relatively flat level away from notches. On fitting the notch filter to a recording, we minimize the filter output, which forces the notches to attempt to cancel the signal frequencies at each time. After model fitting one can the use the notch frequencies as estimates of the dominant frequencies of tonal components of the signal. The frequencies can then be fed to a classifier, but this step is not reported here. The ANF model is tailored to fit narrowband signals, and is capable of modeling simultaneous signals of time-varying frequencies and amplitudes. The authors have recently reported on this for cetacean whistles [3][4]. Because it works directly on the signal waveform, the ANF model avoids the resolution problems of characterization methods that are applied to the spectrogram or another time-frequency distribution. Also, it is simple to implement and use, requires little user tuning, and can be run in real-time.

This paper exploits a novel architecture where the ANF is used both for detection and characterization. The algorithm is an adaptive scheme with a fading memory. This allows it to estimate parameters based on a finite observation interval, so giving it the ability to track time-varying parameters. We propose to run a parametric model along the whole signal and detect signals from a measure of the reliability of the parameter estimates. This measure is an internal variable in the adaptation scheme, so detection comes at very little extra computational cost. The detection decision could possibly also be made on the basis of an analysis of the frequency estimates themselves. But for time-varying signals this is difficult - for instance we cannot just use the parameters' variances because a time-varying signal will naturally impose its own variation in the parameters.

The layout of this paper is as follows: In Sections 2 and 3, we describe the theory of ANF modeling and establish theoretical grounds for a suitable detection statistic. Then, in Section 4 we describe how to apply our method to detect and characterize tonal sounds in oceanic background noise. Section 5 reports on the results of application to a simulated signal and the workshop dataset L-138. Finally, in Section 6 we draw conclusions from the findings.

2. REVIEW OF ADAPTIVE NOTCH FILTER THEORY

For a signal composed of one or more slowly evolving narrowband signals, an AR model is appropriate. However, when such a signal is observed in moderate or strong background noise, the AR model does not perform well. Better performance can be obtained by including the noise in the model. To this end, we first pre-whiten the noise by estimating and equalizing its spectrum. This will be further discussed in Section 4. Including white background noise in the model results in an ARMA model with equal AR and MA parts. However, such a model is not identifiable as its transfer function is undefined at the signal frequencies. The standard way of dealing with this is to contract the poles slightly towards the origin using the pole contraction factor ρ, $0<\rho\leq1$. The pole contraction factor controls the notch bandwidth, and therefore implicitly the trade-off between tracking ability and noise robustness. This is because the algorithm is only able to track a signal if its instantaneous frequency falls within the current position of a notch, but the wider the notch width the more noise energy slips into the notch and influences the estimation.

The adaptive notch filter model expresses the prediction error $\varepsilon(n)$ as

$$\varepsilon(n) = \frac{1+\sum_{i=1}^{P} a_i(n)q^{-i}}{1+\sum_{i=1}^{P} \rho^i a_i(n)q^{-i}} y(n) = H(q^{-1},n)y(n) \qquad (1)$$

where $y(n)$ is the recording, $H(q^{-1},n)$ is the transfer function of the notch filter, and $a_i(n)$ is the ith AR coefficient at time n. The model order P is twice the number of components M tracked by the model, $P=2M$. The model order is a user-defined parameter, however one for which we argue that the exact value is not critical. Since we shall couple the estimation to detection, wherein we shall determine when a component is locked on to a signal, it suffices to choose the model order as the maximum number of simultaneous tonals that one wishes to track. If more tonals are present, the model will track the strongest ones at each time.

Several different filter parameterizations can be used.

Estimating the AR coefficients $a_i(n)$ directly is the most intuitive approach, but it is better to use parameters that relate to only one tonal each. This can be accomplished by either using a direct frequency parameterization based on the fact that the transfer function pole angles are equal to the normalized angular notch center frequencies [5], or by writing the notch filter in cascaded form [6][7]. Here, the direct frequency parameterization is chosen. Classification of a tonal marine mammal sound is usually based on its frequency contour, that is the evolution of its dominant frequency with time [8]-[10], so this is suitable for the application. Moreover, by defining the model through the poles one obtains a representation in which the parameters are nearly independent. This permits the development of the detection methodology described in Section 3. However, cascaded filter forms promise better convergence properties [7] and will be studied in the future.

For slowly evolving tonals, the AR polynomial $A(q^{-1},n)$

$$A(q^{-1},n) = 1 + \sum_{i=1}^{P} a_i(n)q^{-i} \tag{2}$$

is necessarily monic symmetric. This implies that the transfer function of each cascaded filter stage only has M free filter parameters (ρ is usually taken as a user parameter).

We use the popular Gauss-Newton type recursive prediction error (RPE) algorithm [11] to estimate the model parameters of a direct frequency parameterized adaptive notch filter. This algorithm has several attractive properties, including a fast operation (it has been implemented in real-time), good convergence properties, and a minimal parameter variance when applied to stationary signals [11]. The properties of ANFs estimated with the RPE algorithm and applied to both stationary and non-stationary signals have also been much studied [6], [12] -[14]. With the RPE algorithm, estimation works by stepwise minimization of a cost function $\beta(n)$, which is a weighted sum of squared prediction errors,

$$\beta(n) = 2t(n)\sum_{m=1}^{n} \Gamma(n,m)\varepsilon^2(m) \tag{3}$$

Here, $\Gamma(n,m)$ is a weighting function that defines the

1. Initialize: $\boldsymbol{\chi}(0) = \mathbf{0}, \boldsymbol{\psi}^c(0) = \mathbf{0}, \mathbf{S}(0) = \left[100/\overline{y^2}\right]\mathbf{I}, \omega_i(0) = \pi i/(M+1)$. Design parameters: $\rho(n), \lambda(n), P(=2M)$.

2. For $n=1,2,\ldots,N$ do:

$$\chi_i(n) = -y(n-i) - y(n-P+i) + \rho^i(n)\varepsilon(n-i) + \rho^{P-i}(n)\varepsilon(n-P+i) \quad ,1 \le i < M$$

$$\chi_M(n) = -y(n-M) + \rho^M(n)\varepsilon(n-M)$$

$$\varepsilon(n) = y(n) + y(n-P) - \rho^P(n)\varepsilon(n-P) - \boldsymbol{\chi}^T(n)\mathbf{a}(n-1)$$

$$\psi_i^c(n) = -y_F(n-i) - y_F(n-P+i) + \rho^i(n)\overline{\varepsilon}_F(n-i) + \rho^{P-i}(n)\overline{\varepsilon}_F(n-P+i) \quad ,1 \le i < M$$

$$\psi_M^c(n) = -y_F(n-M) + \rho^M(n)\overline{\varepsilon}_F(n-M)$$

$$\mathbf{J}(n) = J\{\boldsymbol{\omega}(n-1), \mathbf{a}(n-1)\}$$

$$\boldsymbol{\psi}(n) = \mathbf{J}(n)\boldsymbol{\psi}^c(n)$$

$$\mathbf{S}(n) = \frac{1}{\lambda(n)}\left[\mathbf{S}(n-1) - \frac{\mathbf{S}(n-1)\boldsymbol{\psi}(n)\boldsymbol{\psi}^T(n)\mathbf{S}(n-1)}{\lambda(n) + \boldsymbol{\psi}^T(n)\mathbf{S}(n-1)\boldsymbol{\psi}(n)}\right]$$

$$\boldsymbol{\omega}(n) = \boldsymbol{\omega}(n-1) + \mathbf{S}(n)\boldsymbol{\psi}(n)\varepsilon(n)$$

$$\mathbf{a}(n) = G\{\boldsymbol{\omega}(n)\}$$

$$\overline{\varepsilon}(n) = y(n) + y(n-P) - \rho^P(n)\overline{\varepsilon}(n-P) - \boldsymbol{\chi}^T(n)\mathbf{a}(n)$$

$$\overline{\varepsilon}_F(n) = \overline{\varepsilon}(n) - \rho^P(n)\overline{\varepsilon}_F(n-P) - \rho^M(n)\overline{\varepsilon}_F(n-M)a_M(n) -$$
$$- \sum_{i=1}^{M-1}\left[\rho^i(n)\overline{\varepsilon}_F(n-i) + \rho^{P-i}(n)\overline{\varepsilon}_F(n-P+i)\right]a_i(n)$$

$$y_F(n) = y(n) - \rho^P(n)y_F(n-P) - \rho^M(n)y_F(n-M)a_M(n) -$$
$$- \sum_{i=1}^{M-1}\left[\rho^i(n)y_F(n-i) + \rho^{P-i}(n)y_F(n-P+i)\right]a_i(n)$$

Table 1. Gauss-Newton RPLR estimation algorithm for adaptive notch filter with direct frequency parameterization [5].

analysis window and $t(n)$ is a normalization factor equal to the inverse of the window sum. The RPE algorithm is in essence an algorithm for stationary signals, so we use the window function $\Gamma(n,m)$ to define a short analysis window within which the signal is nearly stationary. The width of the window is controlled by a design parameter known as the forgetting factor, λ. The name arises because λ controls the duration of the algorithm's memory. For a constant forgetting factor, the window is exponential and given by

$$\Gamma(n,m) = \lambda^{n-m} \tag{4}$$

The Gauss-Newton RPE-type algorithm for estimation of the pole angles of an adaptive notch filter is given by Table 1. The algorithm was first reported by Chen *et al* [5] and is based on the ANF of Nehorai [5] and the findings of Nehorai and Starer on a pole-parameterized AR model [13]. In Table 1, $\psi^c(n)$ denotes the negative gradient of the prediction error with respect to the vector of transfer function coefficients $a(n)$, $\psi(n)$ the gradient with respect to the angular notch center frequencies $\omega(n)$,

$$\psi_i(n) = -\frac{\partial \varepsilon}{\partial \omega_i}(n) \tag{5}$$

and $S(n)$ the inverse of the pseudo-Hessian matrix (see [11], [5] and Section 3). $G(n)$ is the mapping from $\omega(n)$ to $a(n)$. One can implement it by denoting $a_i = a_i^{(M)}$ and for $m=0,1,..,M$ calculating [13]

$$a_i^{(m)} = a_i^{(m-1)} - 2a_{i-1}^{(m-1)}\cos 2\pi\omega_m + a_{i-2}^{(m-1)} \tag{6}$$

for $1 \leq i < P$, given that $a_0^{(0)} = 1$ and $a_i^{(m)} = 0$ for all other i and m. $J(n)$ is the Jacobian of this mapping, which can be estimated iteratively from [13]

$$\begin{cases} \dfrac{\partial a_0}{\partial \omega_p}(n) = 0 \\[2mm] \dfrac{\partial a_1}{\partial \omega_p}(n) = 2\sin \omega_p(n) \\[2mm] \dfrac{\partial a_i}{\partial \omega_p}(n) = 2\cos \omega_p(n)\dfrac{\partial a_{i-1}}{\partial \omega_p}(n) - \dfrac{\partial a_{i-2}}{\partial \omega_p}(n) + \\[2mm] \qquad\qquad + 2a_{i-1}(n)\sin \omega_p(n), \qquad 2 \leq i < P \end{cases} \tag{7}$$

for $1 \leq p < M$. Further, $y_F(n)$ and $\varepsilon_F(n)$ are the recording and prediction error, respectively, filtered by the AR part of the notch filter, and $\bar{\varepsilon}(n)$ is the a posteriori prediction error. Using $\bar{\varepsilon}(n)$ instead of $\varepsilon(n)$ where possible improves the convergence properties of adaptive estimation algorithms [11]. See [5] for a more thorough discussion. The tonal frequencies $f_i(n)$ can be estimated from

$$f_i(n) = f_s \frac{\omega_i(n)}{2\pi} \tag{8}$$

where fs is the sampling frequency.

The selection of the design parameters λ and ρ is important. A higher forgetting factor gives a better noise robustness, but a reduced tracking ability, and vice versa.

This trade-off is similar to that which controls the choice of the pole contraction factor, wherefore a relationship between them can be determined. Previous authors have studied how to choose this relationship according to different criteria [15]–[18]. For simplicity, we choose to adopt the result of Dragosevic and Stankovic [15], which is that the optimal pole contraction and forgetting factors are equated. This result was derived assuming that components are strictly narrowband and that their frequencies evolve according to a random walk model, which describes frequency increments as small and normally distributed with zero mean. Frequency increments on tonals are not well described by a random walk model, although by the central limit theorem if we average the increments over a great many signals we might expect this to result in a Gaussian pdf. Whilst this forms a partial justification for our choice we make no claim about absolute optimality. Our limited information about the signals of interest precludes determining a relationship that is certain to give better performance.

In this study, we keep the forgetting factor λ (and, consequently, the pole contraction factor ρ) constant during the course of the whole recording. This is not the common approach on previously detected signals [12], [16]-[18]. There, one usually increases λ and ρ exponentially from a low starting value. But we do not beforehand know where we will find detections, and have no reason to change our trade-off between tracking ability and noise robustness during the recording.

3. USING THE MODEL OUTPUT FOR DETECTION

Near an extremum point of a one-dimensional function, the second derivative provides a measure of how "sharp" the stationary point is. Equivalently, a measure of the width of the extremum peak or trough is given by the inverse of the second derivative. In multiple dimensions, the analog to the second derivative is the Hessian matrix P_{ij} of second order partial derivatives. The Hessian of the cost function of equation (3) is

$$\mathbf{P}_{ij}(n) = \frac{\partial^2 \beta(n)}{\partial \omega_i \partial \omega_j} = \\ = t(n)\sum_{m=1}^{n}\Gamma(n,m)\left\{\psi_i(m)\psi_j(m) + \varepsilon(m)\frac{\partial^2 \varepsilon}{\partial \omega_i \partial \omega_j}(m)\right\} \tag{9}$$

The diagonal elements of the inverse pseudo-Hessian $S(n)$ provide a measure of the peak or trough width. In the pseudo-Hessian, the second term of equation (9) has been discarded, but this is a good approximation close to an extremum point [11]. It also ensures that the diagonal elements of $S(n)$ are positive. $S(n)$ can then be used to measure the reliability of the current parameter estimate. This has also been formalized in the Cramer Rao theorem

on a lower bound to the variance of any unbiased estimate of a stationary parameter. The Cramer-Rao lower bound is estimated from the diagonal elements of the inverse of the Fisher information matrix, which in white background noise and for stationary parameters is closely related to the Hessian. Although the formalities are not detailed here, we note that for stationary signals the diagonal elements of the inverse Hessian are related to the variance of the parameter estimates. This nice property does not strictly hold for the non-stationary signals of interest here, but if they are indeed nearly stationary within the analysis window there should still be a strong relation between the diagonal of $S(n)$ and the estimate variances.

To use the above reliability estimation method in practice, we need to know that the estimate is close to the extremum point. There is no guarantee for this, but in the RPE algorithm updating step it is assumed that the previous parameters actually minimized the cost function [11]. If this approximation were not a good one at each time, the algorithm would lose tracking and drift off into noise space. The fact that it is much used and recognized for its good tracking performance [11], as is also found here, constitutes a heuristic validation for the approximation.
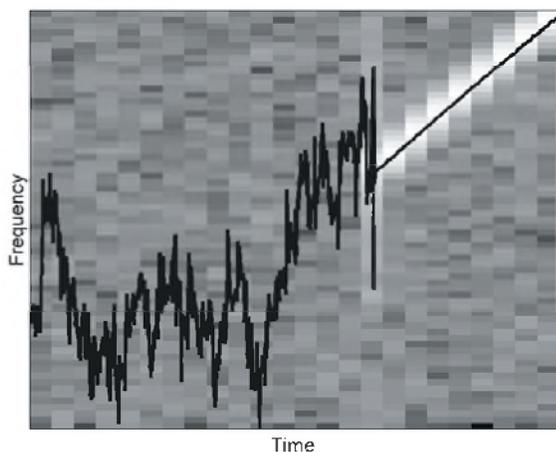
Figure 1. Gray-scale spectrogram of linear chirp in white noise, with frequency estimates overlaid (thick solid line).

For detection from a reliability measure of the parameter estimates, we need to convince ourselves that a reliable estimate only occurs when the adaptive notch filter is tracking a signal. The model is constantly looking for locally dominant frequencies, and even in white noise there are local time-frequency regions where the noise power is stronger. When signals are absent, the noise therefore produces a fluctuating parameter estimate. Figure 1 shows the evolution of a single component frequency upon moving from a noise-only section to tracking a signal (a linear chirp). It is apparent that the estimate fluctuates until the signal starts, but that as the algorithm starts to track the signal it becomes stable.

The background noise adds a stochastic component to the reliability estimate. In zero-mean noise, we can theoretically remove the effect of this by calculating the expectation value of the Hessian before inverting it. However, this appears very difficult in any practical application. It is commonplace in adaptive estimation schemes to simply ignore the expectation operation and use the "raw" quantity instead, accepting that the estimate becomes more variable. This approach is also taken here.

4. APPLICATION

To apply our detection and characterization method to a signal, we first need to pre-whiten it. Here the background noise spectrum is equalized and normalized, and constant frequency tonals are also attenuated. This is necessary because the model describes the background noise as white. It also helps to remove unwanted components such as ship noise.

In this study, pre-whitening is implemented by estimating the noise magnitude spectrum and then dividing the total magnitude spectrogram by it, thus preserving the phase information. The noise power spectrum is estimated from spectrograms of long data blocks using order statistics such as the median and trimmed mean [19]. This approach allows us to estimate the spectrum from the noise-dominated smaller values of the spectrogram only. An alternative approach more suitable for streaming data is to estimate the noise power spectrum from a moving average on the recording spectrogram. If the window is exponential no memory of previous data is required to update the noise spectrum estimate with current data. This method is fast and simple but its spectrum estimates are easily influenced by the presence of signals. It is therefore not used here.

The pre-whitening is the only processing step that operates on the spectrogram. We therefore subsequently inverse transform to obtain the pre-whitened time waveform. Then, the RPE parameter estimation algorithm of Table 1 is run on the whole signal. It is then interrogated for the diagonal elements of the inverse pseudo-Hessian matrix, $S(n)$, at each time instant. The detection statistic used is developed from each of these elements on a logarithmic scale,

$$\alpha_{k,raw}(n) = \log_{10} S_{kk}(n) \qquad (3)$$

This is referred to as the *raw* detection statistic. The logarithm makes its range more manageable. Note that the inverse Hessian diagonal elements decrease when the model starts to track a signal, so detections are made from small values of the detection statistic.

In white Gaussian noise (WGN), the raw detection statistics on different components are independent. Figure 2 shows histograms of the detection statistics on components 1, 2, and 3 for 100000 samples of WGN. The line shows the mean of the data fitted to a normal probability density function (pdf). As the figure shows, raw detection statistics

on different components all have approximately the same distribution, which is well approximated by a normal pdf.

As predicted in Section 3, the raw detection statistic fluctuates rapidly and may be difficult to threshold. Therefore, a smoothing filter is applied to it. The smoothed detection statistic is a weighted linear combination of raw detection statistics, so is also approximately normally distributed on WGN input.
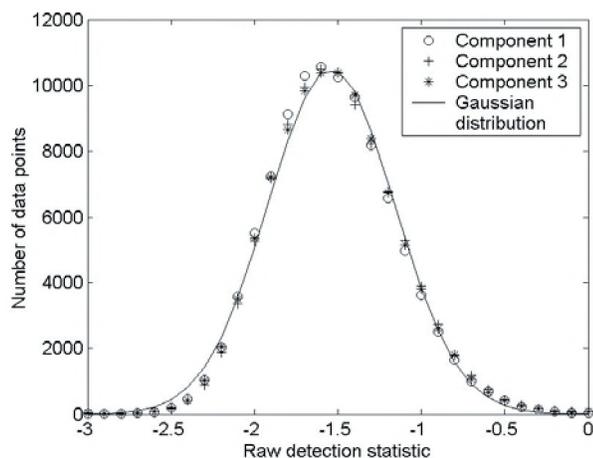
Figure 2. Histograms of raw detection statistics on 100,000 samples of WGN, fitted to a Gaussian distribution.

The means and standard deviations of the smoothed detection statistics are used to define a simple detection threshold; for detection, we require that the smoothed detection statistic is lower than a specified number of standard deviations below the mean. Here, the Page test can be employed to improve the performance [20], and this will be investigated in the future.

To as far as possible prevent signals from affecting the threshold levels, we define an equal detection threshold for all components from the highest mean estimate and the smallest estimated standard deviation. A signal lowers the detection statistic, so we expect these estimates to be the least influenced by signal presence. We estimate the mean and the standard deviation of the detection statistic by averaging over the whole recording. Since the distribution of each detection statistic is approximately normal on WGN input, we use the median as a (nearly) unbiased estimator of the mean.

On streaming data one could instead estimate the mean and standard deviation of the detection statistic via sliding window averaging. It would then be possible to prevent signals from affecting the detection thresholds by only updating the threshold estimates from non-tracking components. This approach was not taken here.

Despite the smoothing of the detection statistic, the algorithm also picks up short duration transients such as clicks. Cetacean clicks can be so much stronger than tonals that in the short processing window applied to the recording, they can contain more energy even in narrow frequency bands. Therefore, even if the model is tracking one or more tonals there is a high risk of it switching to tracking the dominant frequencies of the click. (Note that it is probably not possible to describe the click as a sum of constant frequency tonals even within our short analysis window, so the term "dominant frequencies" should be interpreted as the peak frequencies in the spectrum of the current analysis window.)

Disturbance to the tracking by clicks is of course undesirable. It could be avoided by lengthening the analysis window to reduce the ratio of click power to tonal power, but that would also reduce the tracking ability. An alternative is to introduce a pre-processing step, which detects clicks and reduces their influence. However, none of these measures have yet been implemented.

5. RESULTS

To illustrate the use of the proposed detection and characterization method, we commence by applying it to a simulated signal consisting of one linear and one non-linear chirp immersed in white noise. This signal is not intended to directly simulate a marine mammal call, although dolphin whistles usually have a narrow bandwidth and a smooth frequency evolution.

The amplitude of the linear chirp is constant at 7.9, whereas the non-linear one varies quadratically from 19.1 to 9.6 (RMS value 16.2). White noise of variance 7.9^2 is added so that the effective average SNRs for the linear and non-linear components are –3.0 and 3.2 dB, respectively.

The detection threshold is determined from the noise-only sections at the start and end of the signal. It is set at 2.5 times the estimated standard deviation below the estimated mean. This gives a low false alarm rate. The normalized cut-off frequency of the smoothing filter is 0.01, corresponding to an averaging length on the order of 100 samples.
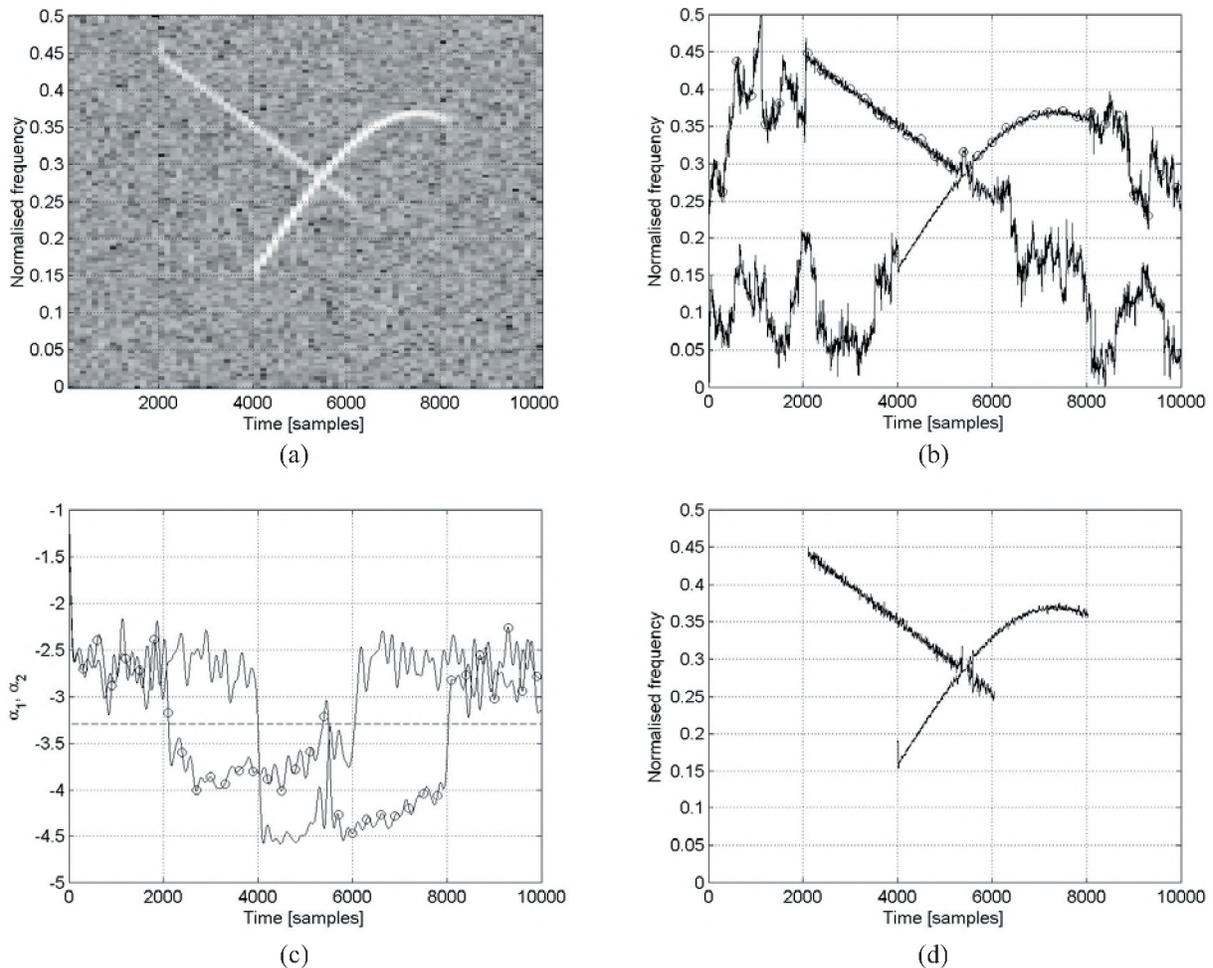
Figure 3. The proposed algorithm applied to the simulated two-component signal. (a) Spectrogram of the signal. (b) Estimated instantaneous frequencies. (c) Detection statistics and threshold level (dashed). (d) Detected components.

The results are shown in Figure 3. Subfigure (a) shows a spectrogram of the simulated signal. Estimated instantaneous frequencies can be found in subfigure (b). The evolution of the detection statistics on components 1 and 2 are shown in subfigure (c). Finally, subfigure (d) shows the detected components. It is evident from Figure 3 that the frequency estimates are highly variable in noise-only sections, but quickly lock on to signals when they appear. The proposed detection statistic provides a good measure of the estimate reliability. Note that the estimate is less variable on the stronger component. This is to be expected, and is also reflected in lower detection statistics on this component. Concluding, as subfigure (d) shows, the method is capable of detecting and characterizing simultaneous components in strong background noise.

We now turn the attention to the problem of detecting marine mammal vocalizations in general, and North Atlantic right whales in particular. The "Report of the Workshop on Right Whale Acoustics: Practical Applications in Conservation" [21] classifies right whale calls as "gunshot",

"low frequency", or "high frequency". A gunshot call is what is usually referred to as a "click". It is an impulsive, broadband sound of duration less than 0.5 s. The low frequency sounds are narrowband, with duration of 0.2-5.0 s and frequencies around 70 Hz. Finally, the high frequency calls have durations of 0.5-3.0 s and fundamentals at 100-600 Hz. A specific common type of high frequency call is the "FM upsweep". The duration of such a call is 0.5-1.5 s and its frequency rises monotonically in the band 100-400 Hz. The FM upsweep call is thought to be used as a contact call. It is the most well known call of the North Atlantic right whale and to the best of the authors' knowledge the only one species-specific enough for detection and discrimination from other whales with a reasonable degree of certainty.

Applying the present algorithm to the workshop dataset file L-138 results in a total of 486 detections, using the same threshold and cutoff frequency as for the simulated signal. Among these 486 detections, many are gunshot or click sounds, and many others are low-SNR calls split up into
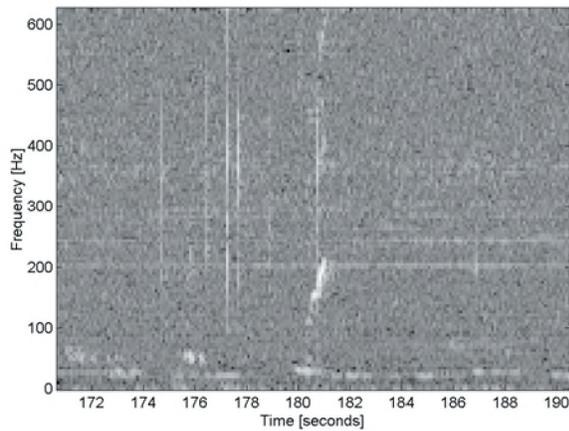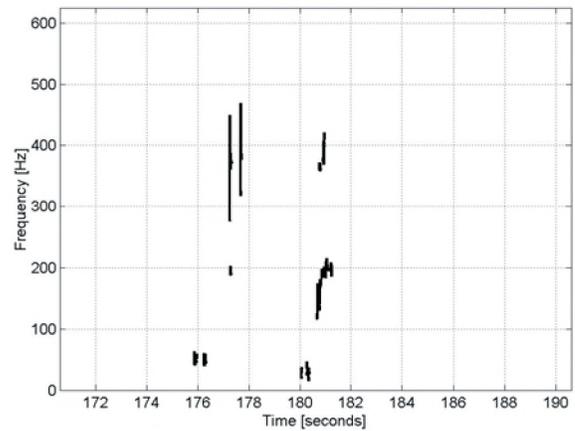
Figure 4. (a) Spectrogram of 20 s of data from dataset file L-138, centered at the detected right whale call starting at 180.6 s. (b) Detections extracted from this 20 s data batch.
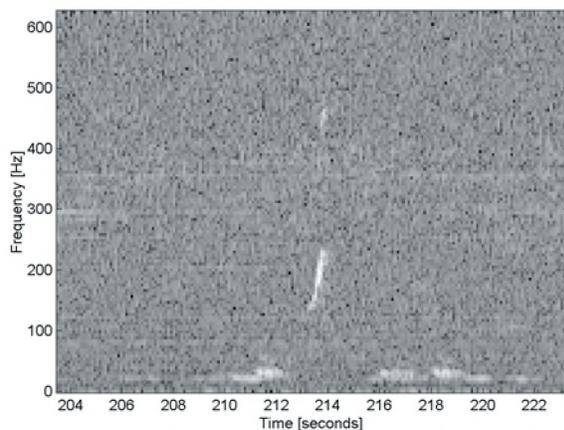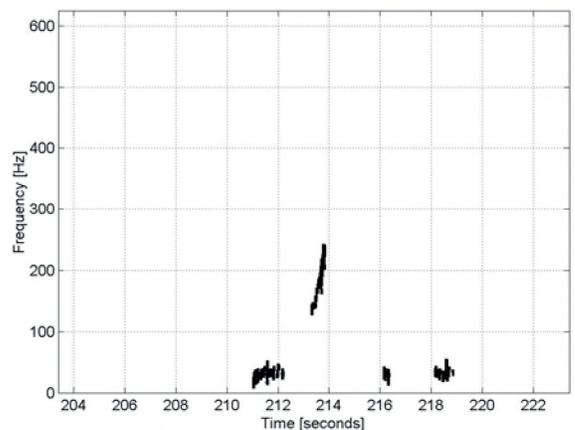
Figure 5. (a) Spectrogram of 20 s from dataset file L-138, centered at the detected right whale call starting at 213.4 s. (b) Detections extracted from this 20 s data batch.

many detections. These are not of interest here as they cannot directly be used to identify HF upsweep calls. Only detections lasting more than 0.2 seconds – 155 in total – are included in the search for right whales. Among these, 86 have frequencies below 50 Hz and are likely to be fin whale calls. Out of the remaining 69 calls, 8 calls are identified as candidate right whale calls. These are between 0.3 and 1.5 seconds long, start and end within 50 to 450 Hz, and sweep up at least 50 Hz.

These selection criteria are based on the reported characteristics of HF upsweep calls [21], "loosened up" to allow for algorithm imperfections. Three of these 8 candidate right whale calls could directly be discarded because their frequency evolutions started low and almost immediately jumped to a nearly constant higher frequency. This is probably caused by the detection firing too early on a strong and suddenly onset call. Future fine-tuning of the algorithm should alleviate this problem.

Four of the remaining sounds sweep up from 120-140 Hz to 200-220 Hz in 0.4-0.6 seconds. The authors believe that these are right whale calls. Their start times are approximately 180.6, 213.4, 505.6, and 536.0 s. The last remaining candidate call, starting at 63.3 s, sweeps up from approximately 80 to 150 Hz in 0.4 s. This is probably too low frequency for a right whale HF upsweep call.
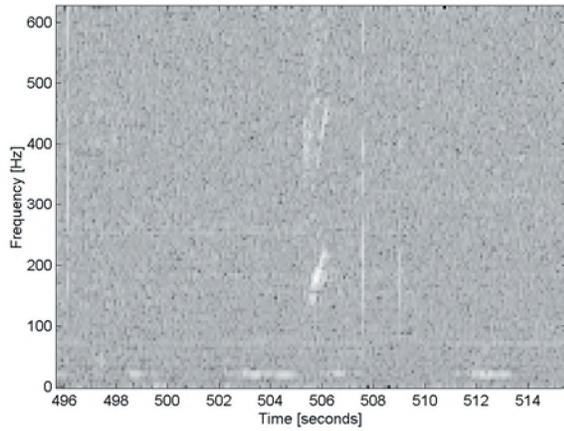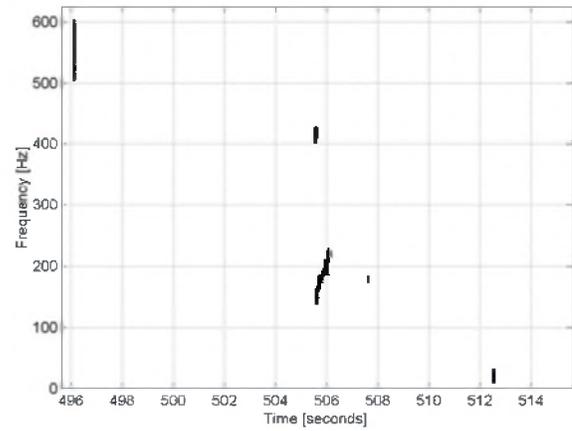
Figure 6. (a) Spectrogram of 20 s from dataset file L-138, centered at the detected right whale call starting at 505.6 s. (b) Detections extracted from this 20 s data batch.
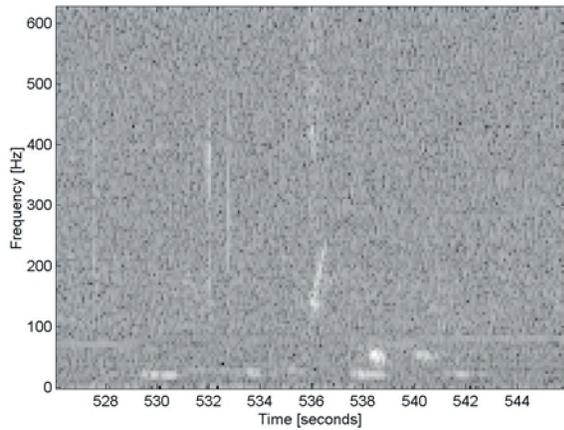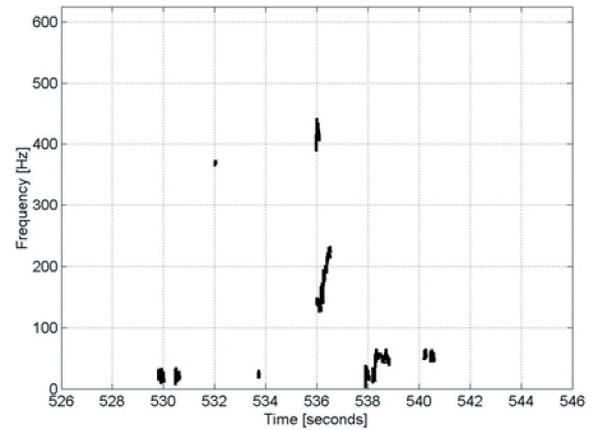
Figure 7. (a) Spectrogram of 20 s from dataset file L-138, centered at the detected right whale call starting at 536.0 s. (b) Detections extracted from this 20 s data batch.

Subfigures (a) of Figures 4-7 show spectrograms of 20 seconds of data centered on each right whale call; the data has been pre-whitened. Estimated frequency evolutions of all detections, that is also those shorter than 0.2 s, within these 20 seconds of data are shown in subfigures (b). These figures show that the proposed algorithm is able to track the frequency contour of the right whale calls. There are also several fin whale detections and some brief click detections. In Figures 4,6, and 7, note that the algorithm has also picked up on what are probably harmonics of the right whale call.

6. CONCLUSIONS

In this paper, we have described a new detection and characterization method for tonal marine mammal vocalizations, and have shown that the method works well with simultaneous sounds, in low signal-to-noise ratios, and with sounds, such as right whale calls, that do not appear to be strictly narrowband.

The method is simple to use and controlled by only a small number of user parameters. It has not yet been implemented in hardware, but in an off-line software

implementation it processes data at a rate that exceeds that necessary for real-time implementation at sample rates of 50-60 kHz.

The algorithm picks up and is disturbed by click sounds, so for a fully automatic operation it is necessary to attenuate these prior to application. Also, despite the smoothing of the detection statistic, calls are sometimes divided into several detections. To counteract this, one can apply a detection-merging algorithm, or change the detection criteria. These improvements will be studied in the future.

ACKNOWLEDGEMENTS

REFERENCES

[1] M. Fujiwara and H. Caswell. Demography of the endangered North Atlantic right whale. Nature, 414(6863):537–541, Nov 2001.

[2] 500-yard protection zone to prevent human activity from disturbing endangered right whales. United States National Oceanic and Atmospheric Administration (NOAA) press release 97-R107, Nov 1997.

[3] A. T. Johansson and P. R. White. Characterisation of cetacean whistles by parametric modelling. In Proceedings of the 17th annual conference of the European Cetacean Society (ECS), Las Palmas de Gran Canaria, Spain, March 2003.

[4] A. T. Johansson, P. R. White, and R. E. Sel- way. Automatic cetacean sound classification –application to sonar transient false alarm reduction and marine environmental monitoring. In Proceedings of Underwater Defense Technology (UDT) Europe, Malmo, Sweden, June 2003.

[5] B. S. Chen, T. Y. Yang, and B. H. Lin. Adaptive notch filter by direct frequency estimation. Signal Processing, 27:161-76, 1992.

[6] K.-H. Rew, S. Kim, I. Lee,and Y. Park. Real- time estimations of multi-modal frequencies for smart structures. Smart materials and structures, 11(1):36-47, Feb 2002.

[7] G. Li. A stable and efficient adaptive notch filter for direct frequency estimation. IEEE Transactions on Signal Processing, 45(8):2001-9, Aug 1997.

[8] V. M. Janik. Pitfalls in the categorization of behaviour: A comparison of dolphin whistle classification methods. Animal Behaviour, 57(1):133-43, Jan 1999.

[9] D. M. Mellinger and C. W. Clark. Recognizing transient low-frequency whale sounds by spectrogram correlation. Journal of the Acoustical Society of America, 107(6):3518-29, June 2000.

[10] S. Datta and C. Sturtivant. Dolphin whistle classification. Signal Processing, 82(2):251-8, 2002.

[11] L. Ljung. System identification: Theory for the user. 2nd Edition, Prentice-Hall, 1999.

[12] A. Nehorai. A minimal parameter adaptive notch filter with constrained poles and zeros. IEEE Transactions on Acoustics, Speech, and Signal Processing, 33(4):983-96, 1985.

[13] A. Nehorai and D. Starer. Adaptive pole es- timation. IEEE Transactions on Acoustics, Speech, and Signal Processing, 38(5):825-38, 1990.

[14] T. S. Ng. Some aspects of an adaptive digital notch filter with constrained poles and zeros. IEEE Transactions on Acoustics, Speech, and Signal Processing, 33(2):158-61, Feb 1987.

[15] M. V. Dragosevic and S. S. Stankovic. An adaptive notch filter with improved tracking properties. IEEE Transactions on Signal Processing, 43(9):2068-78, 1995.

[16] P. Stoica and A. Nehorai. Performance analysis of an adaptive notch filter with constrained poles and zeros. IEEE Transactions on Acoustics, Speech and Signal Processing 36(6): 911-9, 1988.

[17] P. Tichavsky and P. Händel. Two adaptive algorithms for adaptive retrieval of slowly time-varying multiple cisoids in noise. IEEE Transactions on Signal Processing, 43(5):1116-27, 1995.

[18] P. Händel and A. Nehorai. Tracking analysis of an adaptive notch filter with constrained pole and zeros. IEEE Transactions on Signal Processing, 42(2):281-91, 1994.

[19] T. S.-T. Leung and P. R. White. Mathematics in Signal Processing IV, pages 369-82. Clarendon Press, Oxford, 1998.

[20] D. A. Abraham. Analysis of a starting time estimator based on the Page test. IEEE Transactions on Aerospace and Electronic Systems 33(4): 1225-34, 1997.

[21] D. Gillespie and R. Leaper, eds. Report of the Workshop on Right Whale Acoustics: Practical Applications in Conservation. Woods Hole Oceanographic Institution, March 2001.

Detection and Classification of Marine Mammals Using an LFAS System

S.P. van IJsselmuide and S.P. Beerens
TNO Physics and Electronics Laboratory,
P.O. Box 96864, 2509 JG The Hague, The Netherlands, E-mail: Beerens@fel.tno.nl

ABSTRACT

World wide a concern is emerging about the influence of man-made sound in the sea on marine life, and particularly about high power active sonars systems. Most concern lies with marine mammals, which fully depend on sound in their natural behaviour (foraging, navigation and communication). One of the sonars under debate is the Low Frequency Active Sonar (LFAS). This type of system is designed for long range detection of submarines. It consists of a powerful source and a towed array receiver. Incidents with marine mammals could be avoided if the receiver that is dedicated to detection of submarine echoes, is equipped with Detection, Classification and Localisation capabilities for marine mammals as well.

In this paper the development of a prototype transient detector and classifier for the TNO-FEL LFAS array (named CAPTAS) is described. A broadband beamformer is developed that creates 8 beams (sectors) that are equally wide over the whole frequency band. A multi-beam LOFAR display is presented. On the normalised data a Page's test detector is applied that is "optimum" for signals with unknown duration. Detected transients are sent to a classifier that tries to discriminate between biological and man-made or natural transients. Time-frequency analysis is performed and in the resulting time-frequency plot structures are determined by means of cluster analysis after which the sound is classified. Detection results of the prototype are very good, the Classification module is under development and the Localisation module is part of future research. Part of this research is sponsored by the Royal NetherLands Navy (RNLN).

RÉSUMÉ

L'impact des sons d'origine artificielle sur l'écosystème sous-marin soulève un intérêt mondial croissant. Plus particulièrement, cet intérêt se porte sur l'impact des systèmes sonar actifs à forte puissance sur les mammifères marins, dont le comportement est entièrement basé sur l'utilisation du son (aussi bien pour s'alimenter, s'orienter ou communiquer). Un des systèmes sonar concernés est le Sonar Actif à Basses Fréquences (LFAS). Ce type de système est conçu pour la détection longue distance de sous-marins. Il consiste généralement d'une source puissante et d'une antenne de réception remorquée. Les accidents causés par l'interaction de ces systèmes sur les mammifères marins pourraient être évités si l'antenne réceptrice dédiée à la détection d'échos de sous marins était munie de capacités de Détection, Classification, et de Localisation (DCL) des mammifères marins environnants.

Cet article décrit un prototype de détecteur/classificateur de transitoires développé pour l'antenne LFAS de TNO-FEL, l'antenne CAPTAS). Un algorithme de formation de voies est appliqué sur l'ensemble de la bande de fréquence, créant 8 voies de largeur égale (secteurs). Une visualisation LOFAR multi-voies est alors proposée et les données normalisées sont soumises à un détecteur élaboré a partir d'un test de Page, optimal pour les signaux de durée indéterminée. Les transitoires détectés sont transmis à un classificateur qui tente de discriminer les signaux suivant leur origine biologique, artificielle ou naturelle: après une analyse temps-fréquence, les images obtenues sont soumises à une analyse de clusters. Les structures temps-fréquence résultant de ce traitement permettent alors de classifier le son précédemment détecté. Les résultats de détection sont excellents, le classificateur progresse rapidement et le développement d'un algorithme de localisation est amorcé. Cette recherche est en partie sponsorisée par la Marine Royale Néerlandaise (RNLN).

1. INTRODUCTION

World wide a concern is emerging about the effects of anthropogenic (man-made) noise in the marine environment. At present most concern lies with marine mammals [1], [2], [3], [4], [5], *i.e.* cetaceans (whales and dolphins) and pinnipeds (seals, etc.), but there is also an increasing interest in effects on sea turtles and fish [6].

1.1 The problem

Marine mammals fully depend on sound in their natural behaviour (foraging, navigation and communication). For these animals, knowledge of the physiological effects of anthropogenic noise on the auditory system is more developed than for 'lower' animal species. However, precise knowledge on acoustic disturbance and/or injury of marine mammals is still very limited, and the same holds for detailed information on marine mammal hearing systems. Intense sound can have severe negative effects on marine animals. The effects may vary between 'audible', via 'change in behaviour' and 'severe disturbance' up till 'hearing injury/death'.

Figure 1: A stranded juvenile Fin whale, found in North France by W.C. Verboom (photo by M. Verboom).

One of the sonars under debate is the powerful low-frequency sound source of Low Frequency Active Sonar (LFAS) systems. Besides a sound source, these systems also consist of a towed array receiver. Incidents with marine mammals could be minimised or even avoided if this receiver is equipped with Detection, Classification and Localisation (DCL) capabilities for marine mammals. The development of a prototype marine mammal detector and classifier is described in this paper. A first version of the detector was already used during the combined TNO-FEL/NURC ADULTS 2003 trial in the Mediterranean, where many whales and dolphins were encountered. Using passive acoustic monitoring as developed in this project, together with adequate mitigation measures should minimise the impact of LFAS on marine mammals.

1.2 Mitigation measures: the solution?

It is clear that mitigation procedures to reduce the impact of anthropogenic noise are at least recommendable to protect marine life. Also in the Netherlands mitigation measures are defined for active sonar. These procedures aim to prevent any damage in the hearing system of marine mammals in the vicinity of military sonar equipment. Three types of measures are commonly applied in mitigation procedures:

Careful mission planning: Before planning a mission in which an active sonar is operated, it is verified whether the area is inhabited by marine mammals in that season.

Monitoring of marine mammals in the best possible way before using the sonar: Not using the sonar if marine mammals are present is a very efficient mitigation measure. But, how do we know whether marine mammals are present? Marine mammals can be monitored in two ways:

- *Visual* monitoring can be successful, but it is problematic at night (although the use of infra-red is considered). Besides, marine mammals spent most of their hours underwater, hidden from eyesight. At high sea-state even a well-trained whale watcher can easily misses a sighting.
- *Passive sonar* can help to detect and is probably the most promising monitor. However, not all species of marine mammals produce sound (some types of pinnipeds), while other mammals produce sound outside the frequency band of the sonar (Cuvier's beaked whales). It is not known if all endangered species vocalise. Moreover, passive sonar does not (directly) provide the animal's range, which is important in all mitigation measures.

Ramp-up schemes to scare marine life away: Slowly raising the source level, so that the animal can swim away and keep the received sounds to acceptable levels (well below the *temporary threshold shift* level of the animal) may work. It prevents the mammal from being ear-damaged; however, it may still impact on the animal's natural behaviour.

Currently at TNO-FEL a tool (named SAKAMATA [7]) is under development that supports all three mitigation measures. Additionally this tool supports the sonar operator with passive acoustic monitoring.

1.3 Outline of this paper

In this paper the development of a transient detector with prototype classifier for the LFAS array of TNO-FEL (named CAPTAS) is described. In Section 2 a broadband beamformer is discussed that creates 8 beams (sectors) that are equally wide over the whole frequency band. Results are presented on a multi-beam LOFAR display. On the normalised data a power-law/Page's test detector is applied that is robust for signals with unknown frequency content and duration.

Detected transients are sent to a classifier that tries to discriminate between biological and man-made or natural transients. The proposed method is based on pattern

recognition in the time-frequency plot (which is a visualisation of the time-frequency distribution). The time-frequency distribution of a transient signal gives valuable information on the nature of the signal. Its bandwidth, duration and other spectral and temporal characteristics can be derived from the time-frequency plot, from now on denoted as *tf*-plot. Other common names for the *tf*-plot are "spectrogram", "LOFAR-gram" or just "gram".

Section 3 describes how time-frequency analysis is performed by means of conventional short-time FFT processing. In the resulting *tf*-plots, structures are determined by means of image processing (clustering). Dedicated cluster analysis classifies the sound as biological or mechanical. In the former case, it is also specified whether the mammal is large or small and whether it is a baleen whale or toothed whale.

2. DETECTION OF MARINE MAMMAL TRANSIENTS

Detection of marine mammals within the *danger-zone* of a sonar system is essential in avoiding exposure of those marine mammals to high-level sounds. The danger-zone is defined as the area where receive levels on animals are higher than "acceptable". What is "acceptable" in this respect is still under heavy debate, but more and more legislation is formulated. The radius of the danger-zone strongly depends on the hearing sensitivity of the species present and on the used sonar source. Typically danger-zones have a range on the order of 0 to 5 nautical mile.

The idea is that active sonar systems should have a sub-system that warns the operator for the presence of marine mammals within the danger-zone. The problem with detecting marine mammals is the wide range of species, where each species produces different sounds with different duration, frequency band and source level. For example, the very large baleen whales produce low frequency calls, around 10-20 Hz, which can last for several minutes. The much smaller porpoises produce very short clicks in the order of less than a millisecond and frequencies up to 160 kHz. As an example, *tf*-plots are shown in **Figure 2** for a Humpback whale and a Bottlenose dolphin. Making a robust detector for all these different sounds is a challenging task.

Several papers and reports are of direct relevance to the current work on the passive detection of marine mammals or the more generic problem of transient detection. A TNO paper [8] describes a transient detector developed for the Active Low Frequency (ALF) towed array. The detector was based on energy detection in a *tf*-plot. Another, very interesting report [9] and paper [10] from the NATO Undersea Research Centre (NURC) describes the combination of a power-law integrator [11] and a Page's test [12] for the passive detection of marine mammals. The power-law integrator is robust against varying signal bandwidth and the Page's test detector is a robust detector for signals with an unknown duration. This

seems to be a very useful method for detecting marine mammals with their wide variety of sounds.
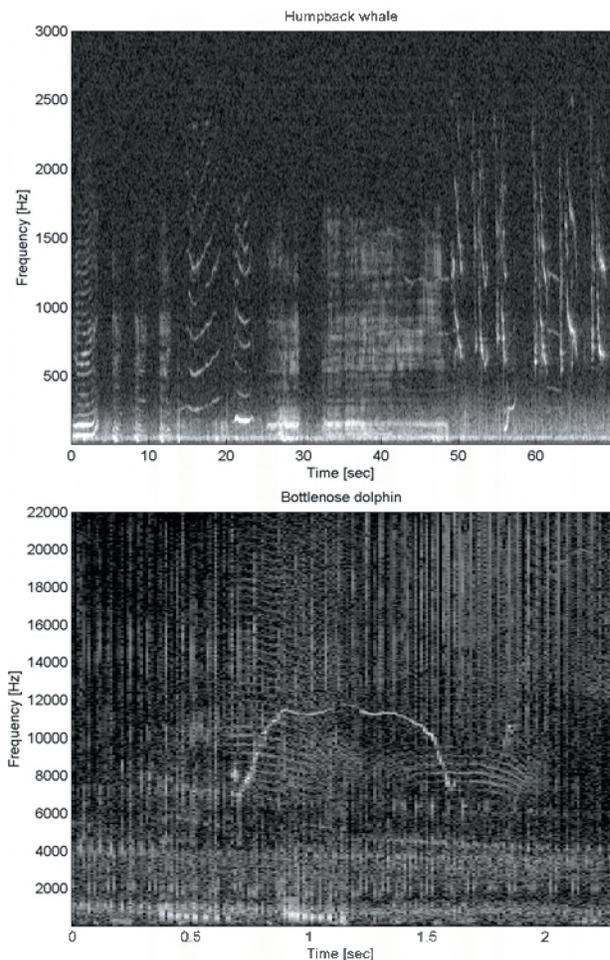
Figure 2: tf-plots of a Humpback whale (top) and a Bottlenose dolphin (bottom); note the very different frequency and time scales.

The remainder of this section describes the development of a prototype marine mammal (or more generic transient) detector for the CAPTAS towed array. The CAPTAS array is a modern LFAS receiving array with a capability for instantaneous left-right discrimination through the use of hydrophone triplets. The array consists of 64 triplets and operates in the frequency band from 10 to 2080 Hz; see [13] and [14]. This detector is strongly based on the already available processing software and structure developed for Anti Submarine Warfare (ASW) tasks (active/passive detection of submarines). A small schematic overview of the proposed marine mammal detector is shown in *Figure 3*.
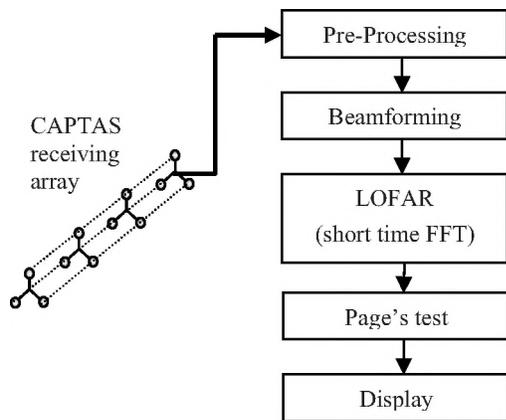
Figure 3: Structure of the proposed marine mammal detector.

The two major processing steps, pre-processing-beamforming and the detection processing, are described in Sections 2.1 and 2.2 respectively. This section ends with a short performance evaluation, based on recorded dolphin sounds and low frequency clicks, in Section 2.3.

2.1 Pre-processing and Beamforming

Pre-processing is the preparation of the hydrophone data so that they can be beamformed. It consist of the following steps:

- Detection and reparation of malfunctioning hydrophones,
- Roll stabilisation of the triplet structure,
- Fourier transformation of the hydrophone signals into the frequency domain.

These pre-processing steps are the same as used in ASW processing.

The beamformer developed for the marine mammal detector is rather different from the beamformer used in the ASW processor. Beamforming, the coherent summation of hydrophones signals, is normally used to improve the Signal-to-Noise Ratio (SNR) resulting in better detection of sound sources. Furthermore, it provides information on the bearing (direction of arrival) of the sound source.

For the detection of marine mammals we are not overly interested in maximising the SNR, since marine mammals generally make more noise than submarines and are relatively easy to detect. However, information on the (horizontal) direction of the target is very useful, *e.g.* for cueing visual observers in the right direction, and is essential for the choice of an appropriate mitigation procedure. When for instance mammals are detected in the forward sectors they will probably close in on the sonar system. This calls for other actions than when the mammals are in the aft sectors where the distance between sonar and mammal increases in time.

Beamforming is therefore an essential step to start with. However, a complicating factor is the wideband nature of the signals, which cover the total frequency band

of the CAPTAS array (10-2080 Hz). Applying a straightforward Delay and Sum Beamformer (DSBF) to this frequency band results in a frequency dependent angular resolution [15]. This has several practical drawbacks like the large number of beams that have to be made at the higher frequencies, while at lower frequencies these beams will overlap. Furthermore, a large number of beams will require a highly automated detection process since the amount of beamformed data will be far too large to be presented on a screen.

A proposed solution for these problems is the use of a constant beamwidth beamformer. This beamformer has a frequency dependent array shading which keeps beamwidth constant for all frequencies. Mathematically, beamforming can be seen as a spatial filter, to which standard filter theory applies. To achieve constant beamwidth, a complex-Remez filter design algorithm [16] was used to compute the filter coefficients (equivalent to array shading coefficients) for each frequency. Applying these filter coefficients results in a constant beamwidth output, as shown in *Figure 4*.
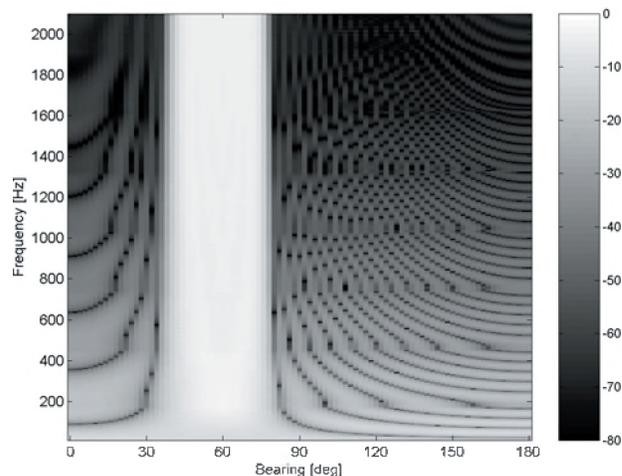
Figure 4: Constant beamwidth beamformer response for the total frequency band and a sector at 60°.

In this case a synthetic beam at 60° has been made with a constant beamwidth for all frequencies between 300 and 2080 Hz. A drawback of this method is the increasing sidelobe levels at lower frequencies. For frequencies below 300 Hz the filter coefficients are all set to one, to avoid excessive sidelobes. This changes the beamformer into an ordinary DSBF, as a consequence the beamwidth starts to increase for frequencies below 300 Hz.

The desired number of beams (look directions) is a compromise between the desired beam resolution, performance and display properties. Initially four beams are formed. In the consequent triplet processing the Port-Starboard (PS) ambiguity is solved, see [13] and [14], and number of beams is doubled. Finally, eight beams are made

directed in eight compass directions: 0°, 60°, 90°, 120°, 180°, 240°, 270° and 300°.

The output of this newly designed beamformer is used as an input for the Page's test detector, treated in more detail in Section 2.2.

2.2 Page's test detector

After frequency-domain beamforming the eight beams are converted back to the time-domain by means of an Inverse Fourier Transform (IFFT). To each of these beams a transient detector is applied. *Figure 5* shows an overview of the detector. The detector consists of a power-law integrator and a Page's test, which seems to be a good combination for detecting the capricious marine mammal signals. This algorithm is very well described in [9] and this section is largely based on this report.
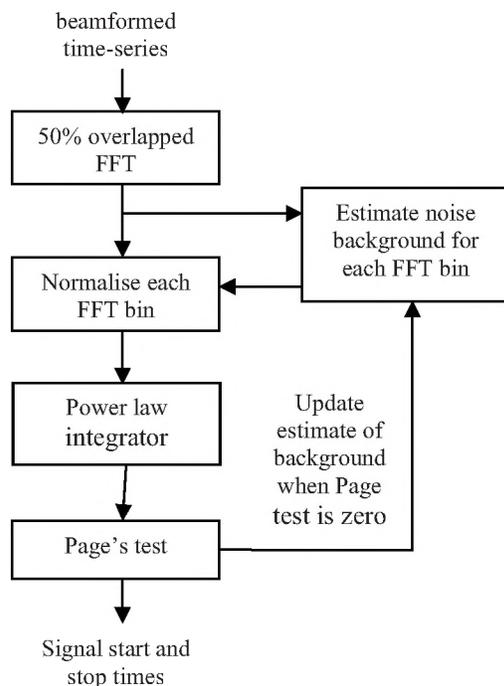
Figure 5: Block diagram of the detection scheme.

Note that in this application the Power law/Page's test detector is applied to beamformed data, but it can be applied to any time-series. In fact it can be applied to almost any receiving system. A good example is the application to sonar buoys in [17]. The following "walk-through" of the detector is based on one channel with acoustic data in the time domain.

As a first step the time-domain data are converted to the frequency domain by means of 50% overlapped short-time Fourier transforms (STFT). The integration time for the Fourier transform is always a compromise between spectral and temporal resolution. Generally a high temporal resolution (short integration time) will improve the detection of short signals like click (bursts) and sweeps.

One of the most important steps in the detector is the estimation of the background noise and interference for each frequency bin. The background consists of (wideband) ambient noise and (narrowband) shipping noise. As shown in *Figure 5* the detector exploits the Page's test to isolate data that is believed to be signal free. These data are then exponentially averaged over time using the following equation.

$$\lambda_{t+1} = \alpha\lambda_t + (1-\alpha)X_t \qquad (1)$$

In this equation λ_t is the old and λ_{t+1} the updated estimate of the background, α the time constant for the exponential averager and X_t the latest signal free power spectrum.

The following two steps are the actual normalisation of the power spectrum and application of the power-law integrator to the normalised spectrum. These two steps are shown together in a detailed overview of the Page's test as shown in Appendix A. The power law integrator sums the normalised frequency bins to a scalar, which is an indication of the energy level (for $p=1$ it is the energy). After a proper normalisation and in a *noise only* case this sum (denoted by Z in Figure) is approximately zero, while during a *signal present* case this sum is positive. Here several thresholds (h_0, h_1, b_0 and b_1) start to play a role. The used Page's test has separate thresholds for the onset detection of the signal (h_0) and the termination detection of the signal (h_1). Associated to these thresholds are biases in order to reduce the sensitivity (b_0 and b_1).

Based on trial and error the bias, threshold and power law parameters were set with the following values:

p — power law ($p=2$)
h_0 — threshold for start of signal detection ($h_0 = 8$)
b_0 — Page's test bias for start of signal detection ($b_0 = 2$)
h_1 — threshold for end of signal detection ($h_1 = 10$)
b_1 — Page's test bias for end of signal detection ($b_1 = 3$)
α — time constant for exponential averager ($\alpha = 0.99$)

These values were "optimised" for a proper operation on the experimental data evaluated, see Section 2.3. For different receiving systems and/or environments other settings might work better. As a rule of thumb the following guidelines can be used for optimising the different parameters:

• Low values for the power law (p) make the system relative more sensitive to wideband signals, while higher values make the system sensitive to narrow band signals.
• A high value for the power-law (p) and low values for the bias (b_0 and b_1) can make the system very sensitive to the small noise bursts that are always present in the underwater environment.
• Increasing the bias and threshold values decreases the sensitivity so that a signal needs a higher SNR to be detected.

For the current application, the detection of marine mammal vocalisations within the sonar danger-zone, SNR is not a real problem. Therefore, relative high values for the biases and thresholds are chosen.

2.3 Performance evaluation

In this section the previously described beamformer and detector are tested on two recorded marine mammal vocalisations. Both recordings were made using the TNO-FEL CAPTAS triplet array. The first recording has been made during an ASW-LFAS trial in 1999 near the Spanish coast of La Coruña and consists of several "high" frequency dolphin sweeps. This trial was conducted in co-operation with the Royal NetherLands Navy (RNLN) and Thales Underwater Systems (TUS). The second recording consists of several low frequency clicks and was made during another ASW-LFAS trial in the Autumn of 2003 near the coast of Sardinia. This trial was conducted in co-operation with the RNLN and NATO Underwater Research Centre (NURC). The recordings represent two different signals (high frequency opposed to low frequency and different characteristics) and are therefore very suitable for testing the marine mammal detector.

Both trials (and other trials performed in the intervening period) were dedicated ASW trials with the focus on testing new active sonar concepts. During these trials marine mammals were only rarely seen and even more rarely recorded with our towed array. Actually the two presented signals are, up to now, the only known marine mammal vocalisation recorded with the CAPTAS array.

The presented results in the following two subsections are intended to illustrate the functioning of the described beamformer and detector. In the future, a more thorough investigation on the detection performance has to be made, preferably with data from a dedicated marine mammal trial.

2.3.1 Detection of dolphin clicks

During an LFAS trial in 1999 several common dolphins approached the towing vessel within visual range, see *Figure 6* for a picture of the dolphins. An example of the acoustic recording made during this approach is shown in Figure 7. The upper panel shows a time-series of a single hydrophone containing several dolphin sweeps and an array artefact, the short and high peak around the 15 second time stamp. The lower panel depicts a *tf*-plot of the same recording. The dolphin sweeps are clearly visible in the upper frequency band (1000-2500 Hz). Also visible are some tonals from the tow ship, the horizontal lines.

Figure 6: Picture of the common dolphins that approached the sonar during the 1999 trial.

Applying the special beamformer of Section 2.1 reveals the direction of the vocalisation. Furthermore, beamforming rejects noise from directions other than the look direction. This is especially helpful in suppressing the tow ship noise, which often dominates the background noise levels.

The output of the beamformer is shown in Figure 8. This figure is a so-called "multi-beam LOFAR". For each of the eight beams a tf-plot is shown with frequency on the horizontal axis and time on the vertical axis. The axes are rotated to make the display look like a more standard LOFAR gram (waterfall) as used in passive sonars.

The tow ship noise is clearly visible in the Northern direction with several loud tonals that also leak into the other directions. Figure 8 depicts a sub-set of the single hydrophone data shown in Figure 7. The dolphin sweep is clearly visible in the Southern direction. This sweep is also weakly visible in the other directions (leakage through the sidelobes) together with some low frequency rumbles.

Figure 8 shows the intermediate result after pre-processing and beamforming. The next step is the normalization of the beamformed data and application of the power-law/Page's test detector for automated detection and extraction of the signals. The result after normalization is shown in Figure 9. This figure has the same set-up as Figure 8, *i.e.* a multi-beam LOFAR. The difference is the normalization, which equalizes the stationary background noise. Signals (fluctuations in the background) are now clearly visible. The dolphin sweep in the southern direction has been particularly clarified.
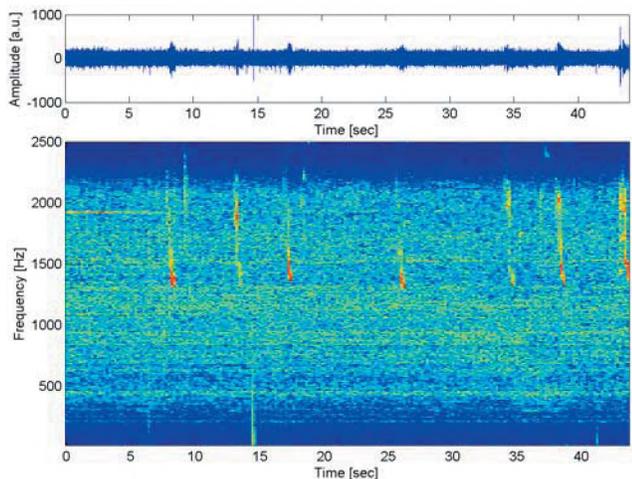
Figure 7: Time-series and tf-plot of single hydrophone data with several dolphin sweeps.
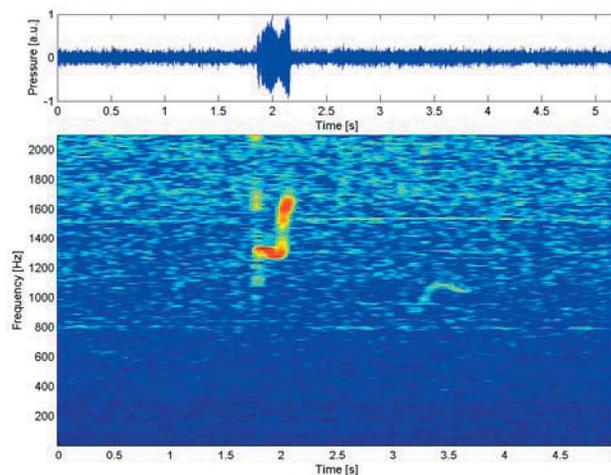
Figure 10: Time-series and "high-resolution" tf-plot of the detected dolphin sweep.
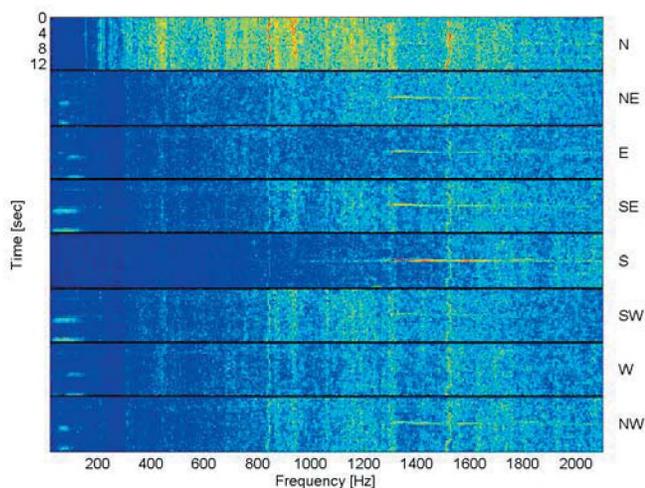
Figure 8: Multi-beam LOFAR display. In each of the eight formed beams 12 seconds of data are depicted with frequency on the horizontal axis and time on the vertical axis. The beam directions (N, NE, etc.) are listed on the right side. A dolphin sweep is visible in the Southern direction, but has also leaked into other directions via the beamformer's sidelobes.
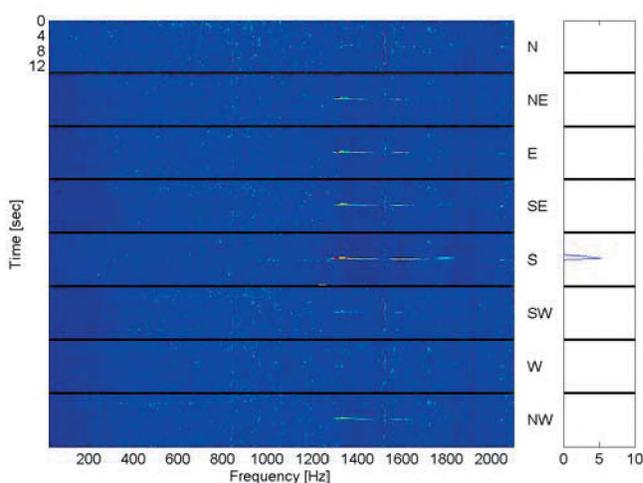
Figure 9: The CAPTAS marine mammal detection display is a multi-beam LOFAR of the normalised data (left) and corresponding Page's test output (right). The blue line on the right side depicts the Page's test output and marks a short signal detection in the Southern (aft) direction.

The right side of Figure 9 shows the output of the power-law/Page's test detector. This detector performs a summation over all frequency bins for each time step. Whenever this summation exceeds the detection onset threshold (h_0), a signal is detected. The thresholds are set using trial and error so that the detector is not sensitive for small noise bursts but still detects the low amplitude transients. In this case, the dolphin sweeps in the southern direction are detected.

After the detection of a transient, the start and stop times of this transient are known and the transient can be stored. This isolation of the signal is very helpful for further analysis (classification and localisation).

Figure 10 depicts a high-resolution *tf*-plot of the dolphin sweep detected in *Figure* 9. This *tf*-plot was made using the stored beamformed data. Compared to the *tf*-plot before beamforming, in Figure 7, it is readily apparent that the signal to noise ratio has been significantly increased. This increase in SNR makes it easier to classify the transient and to detect at longer ranges.

2.3.2 Detection of low frequency clicks

For the second example, we have used recordings made during another LFAS trial conducted in the Mediterranean Sea in the autumn of 2003. During this trial a prototype of the described detector was used for monitoring of the underwater environment and the

detection of transients. During one of the passive experiments, several low frequency transients were detected. An example of recorded, single hydrophone, data is shown in Figure 11. In this figure, a time-series and tf-plot for a 45-second snapshot are shown. The low frequency transients are present in this data but not clearly visible. Around the 10, 17 and 35 second time stamps some very weak and low frequency transients are visible. They are masked by both tow ship noise and acoustic transmissions from an active low frequency source, which was used for performance evaluation tasks during the passive experiment. These acoustic transmissions consisted of broadband noise in the 1000 to 2000 Hz frequency band and three tonals at 1000, 1100 and 1200 Hz. However, both noise sources are only present in the forward sector.

After beamforming the signal to noise ratio of the transients has improved dramatically as can be seen in Figure 12. This figure depicts the beamformer output for a snapshot around the 17 second time mark in Figure 11. The spatial filtering of the beamforming seems highly effective. Several clicks are now visible, especially around the Eastern and Western direction. For these low frequency transients Port-Starboard discrimination cannot be obtained with an array designed for 1-2 kHz. However, the transients in the Eastern direction are slightly stronger than the transients in the Western direction. Other dominating features are the three transmitted tonals, which leak into all directions due to their relative high transmission levels.

Figure 11: Time-series and tf-plot of a raw hydrophone signal with several very low frequency and barely visible transients.
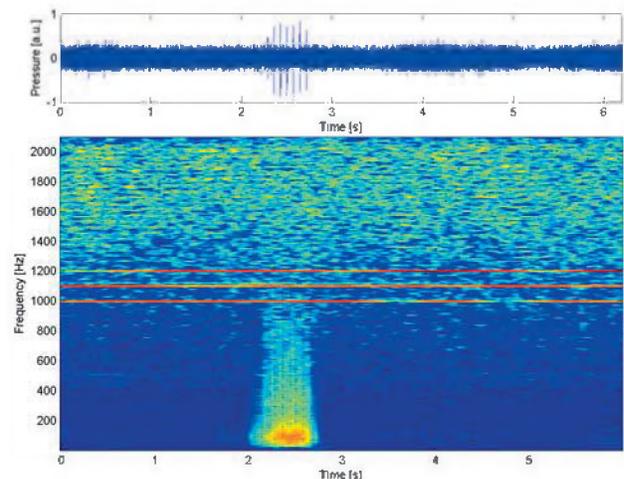
Figure 14: Time-series and high resolution tf-plot of the detected low frequency clicks.

Figure 12: Multi-beam LOFAR plots for the eight formed beams. Low frequency clicks are visible in all directions but clearest in the Eastern (and Western) direction.
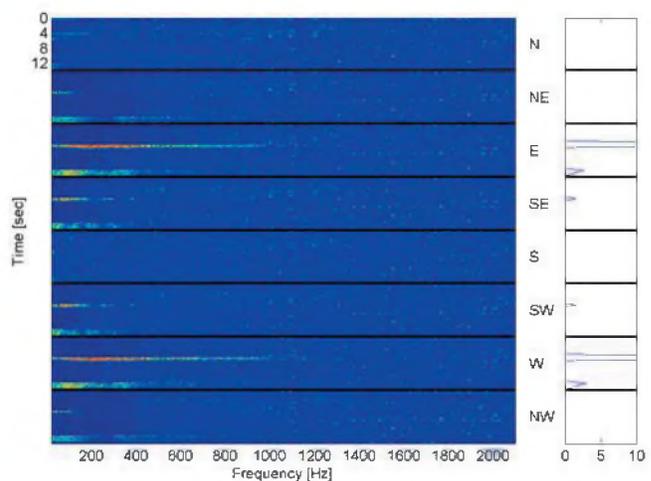
Figure 13: CAPTAS marine mammal detection display. The Page's test output marks several short signal detections especially in the Eastern and Western direction.

The corresponding normalised beamformed data and the power-law/Page's test output are shown in *Figure* 13. After normalisation the clicks are even more visible than previously in Figure 12. The power-law/Page's test detector was triggered by the strongest clicks. Lower settings for the different thresholds can make it possible to

detect the lower level transients. However, this will go at the expense of an increased false alarm rate. The system then becomes more sensitive to tow speed changes and other noise bursts. A high resolution *tf*-plot and time-series of a detected click is shown in Figure 14. Compared to Figure 11, where the clicks were barely visible, the signal to noise ratio has increased. This makes it easier to classify the transient. Nevertheless, this is still a daunting task for acousticians. Our best guess for now is that it is a large whale (maybe a fin whale or sperm whale).

3.   Classification of marine mammal transients

Once a transient is detected, it is important to know whether it is man-made or biological. In the latter case, it is interesting to classify it in more detail. Mitigation measures for large baleen whales are less severe than for small toothed whales, like harbour porpoises.

Two types of classification methods are popular for transients:

- Statistical analysis of time series (higher-order spectra),
- Pattern recognition in *tf*-plot.

We propose the latter method, as the former was unsuccessful in earlier studies. In the following we will work out an example (of harbour porpoise clicks) to demonstrate our prototype classifier, which was trained on several marine mammal recordings that were downloaded from the Internet.
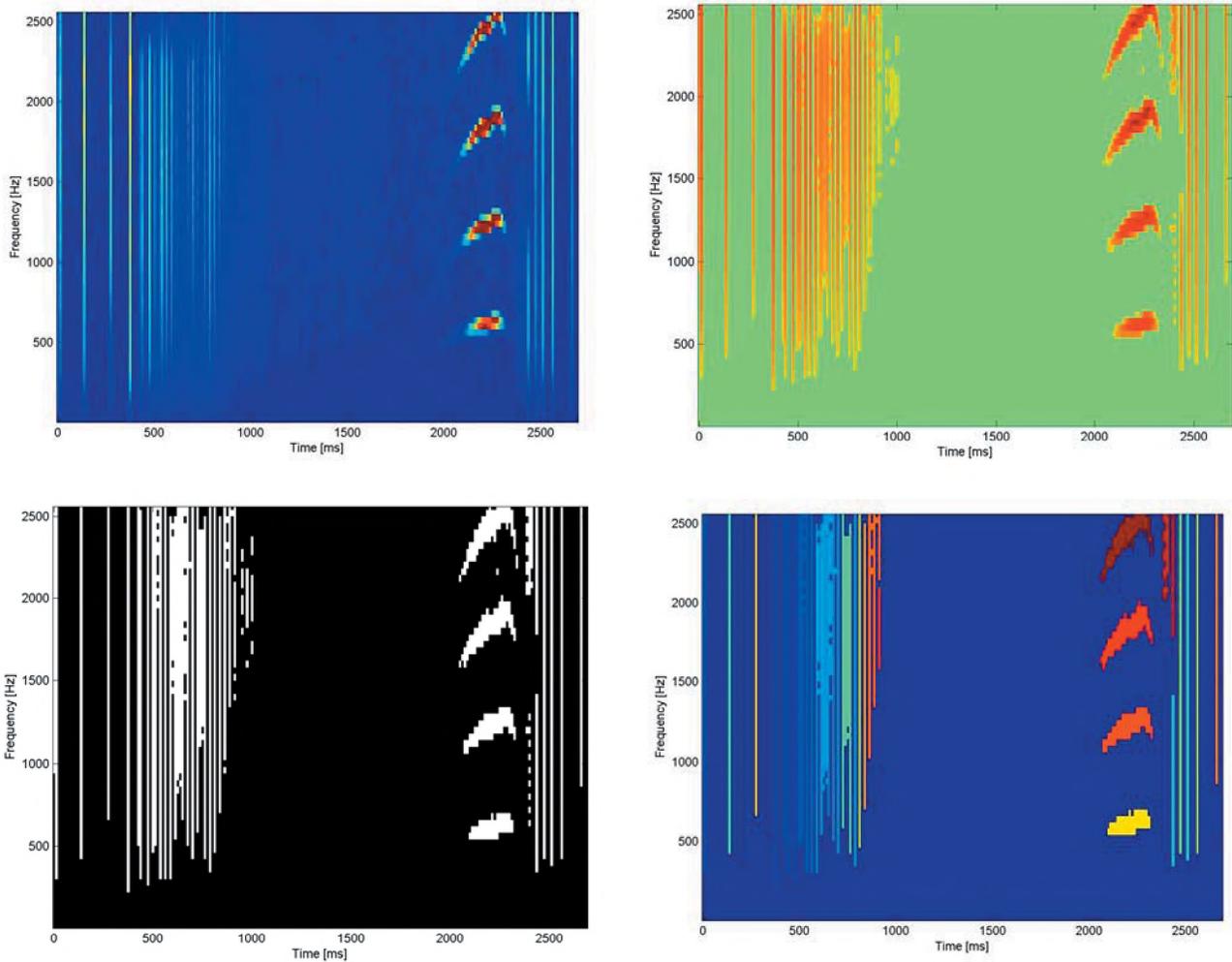
Figure 15: Processing scheme of the tf-plots to enable pattern recognition techniques.
On top left: a raw tf-plot of a porpoise clicks with varying pulse repetition frequency; on top right: a normalised tf-plot; below left: threshold crossings in the normalised tf-plot; below right: clustering of threshold crossings colour-coded by lowest frequency.

3.1 Time-frequency-plots (grams)

Different techniques are often used to compute the time-frequency distribution. The most common used techniques are summarised below:

Short-time FFT processing (STFT) is commonly used to make *tf*-plots. The time-series of the transient is cut into short segments, which are analysed spectrally by means of an FFT. Overlapping (50%) is often used. Sequential spectra are plotted in the gram. The FFT length is an important parameter. It determines the inevitable trade-off between time resolution and frequency resolution. For classification of clicks FFT lengths of 128 (corresponding to 0.025 s) seems suited, but for baleen whale calls longer integration (by a factor 4) is better.

Apart from the standard STFT technique several other methods are worth mentioning. *Cochlea processing* is a technique based on the human–ear. The cochlea in the human inner-ear acts as a logarithmic frequency filter. The technique is very suitable for the identification of human speech and seems suitable for application to other biological sounds as on marine mammal transients. *Wavelet processing* is often mentioned as being optimal for the analysis of transient signals. However, applications of this technique in sonar systems are still pending. This also holds for *Wigner-Ville processing* and other more exotic processing methods.

These innovative techniques all have their own speciality, but none of them proved to be robust for the wide variety of marine mammal sounds (ranging from long, low frequency calls to very short wideband clicks). Therefore we opted to use the simple and robust STFT processing in our prototype classifier until something better comes up. We realise that examples can be found where other processing methods perform better, *e.g.* for low frequency signals of inter animal communication from large baleen whales, the cochleagram more clearly separates the harmonic structures and appears to be the preferred time-frequency distribution. However, the major concern for tactical LFAS systems is for small and medium size odontoces (which are the hunting type of mammals), like harbour porpoises and Cuvier's beaked whales. These animals often produce clicks, which are wideband signals that STFT processing can reasonably deal with. Therefore, it was decided to proceed in this study with ordinary STFT, in which both clicks and calls can be classified.

3.2 Normalisation, thresholding and clustering

Before we can apply pattern recognition techniques, the structures in the *tf*-plot have to be isolated. To achieve this we propose the following processing scheme. See also *Figure* 15 for illustrations, where some porpoise clicks are depicted.

3.2.1 Normalisation

This is an important step. The background energy in the *tf*-plot is often distributed rather than uniformly homogeneous. Therefore "whitening" should be applied before structures can be isolated through thresholding. Both temporal and spectral effects cause inhomogeneity: Background noise levels are higher at lower frequencies. Sometimes pre-whitening is already applied in the recording system, but not always. The whitening can be theoretically compensated (-17 log f for sea noise spectrum according to Knudsen [18]). Adaptive methods that measure the actual spectral background (for instance the method used in Page's test) are used in this study. Apart from spectral variation there is also a temporal variation of the background, which is compensated for by automated level control.

In the upper right panel in *Figure* 15 a normalised *tf*-plot is shown. Compared to the raw *tf*-plot (upper left) the clicks are more clearly separated, mainly due to the temporal normalisation, which compensates for the higher background between 500 and 1000 ms.

3.2.2 Thresholding

After spectral and temporal normalisation the median is subtracted from the data (such that the noise level is 0) and the data is divided by the maximum in the *tf*-plot (such that the maximum signal level is 1). After this, a threshold can be set. Depending on the data quality its value is on the order of 0.05. (The analysed recordings in our training set were downloaded from Internet and differ in recording quality). In the lower left panel in Figure 15 a *tf*-plot after thresholding is shown. Threshold crossings are groups of vertical lines (clicks) in the first 1000 ms and after 2500 ms. Furthermore four 'islands' are visible around 2150 ms. This harmonic structure is caused by rapidly repeating clicks, for which the repetition time is (much) shorter than the integration time. (In reference [19] an elaborate study on harbour porpoise click trains is presented.)

3.2.3 Clustering

The threshold crossings are clearly grouped. In the clustering procedure all connected points are recognised as a single cluster. This procedure is a standard Matlab® function in the image processing toolbox. We removed all small clusters; signals have either duration or bandwidth, so small clusters are often just noise.

In the lower right panel in *Figure* 15 the remaining clusters (51 strongest from a total of 121) in the *tf*-plot are shown. The clusters are numbered starting with the lowest frequency. This means that colour (from blue to red) indicates the lowest frequency in the cluster.

3.3 Pattern recognition

Now that we made clusters, we are left with patterns that need to be recognised in order to classify the signal. The strength of this classification method is that although marine mammal sounds vary a lot in frequency, duration and level, they do not have a lot of different typical patterns. Basically only four typical sounds are produced:

- Clicks
- Moans
- Whistles
- Sweeps

All four of these have easy recognisable (LOFAR) patterns. Clicks are vertical lines. Moans are blobs and always have a harmonic structure. Whistles are thin lines mainly horizontal, and have (weak) harmonics. Sweeps are thin lines with more vertical structure (bandwidth) and sometimes lack harmonics.

For all four typical sounds "recognisers" are developed. These recognisers are built up in similar way. First clusters are reshaped in an automated way by standard image processing techniques. This reshaping is necessary as for instance clicks (vertical lines) are often broken down in several fragments, which can be reconnected by filling techniques. On the other hand moans (islands) tend to be connected by narrow bridges and have to be separated.

Next from the reshaped clusters *features* are determined. These features are elementary properties of the clusters like: length, height, centre of mass, standard deviation, energy content, etc.

Finally these features (or combinations) are compared to standards that are representative for the patterns of the four standard sounds. But before we start, a large false alarm reduction is achieved by recognising air-gun transmission, which is the main source for false detections in the ocean.

3.3.1 Airgun removal

No less than 75% of all detected transients are air-gun transmissions [20]. Air-guns are numerous and both powerful and with low-frequency content such that propagation is favourable for them. Air-guns are easily recognisable in a *tf*-plot, see Figure 16 for an example. They are short and band-limited transients, which are manifested as triangles on the floor of the *tf*-plot. All transients that are classified as air-guns are automatically removed.

3.3.2 Reshaping clusters (Erode-Dilate)

Our strategy is to determine features from the clusters, and compare these to standards for the four classes of signals above. Before we start to determine the features, the clusters are reshaped by means of "erode-dilate" techniques, see [21]. The number of erode-dilate steps is different for each of the four recognisers.
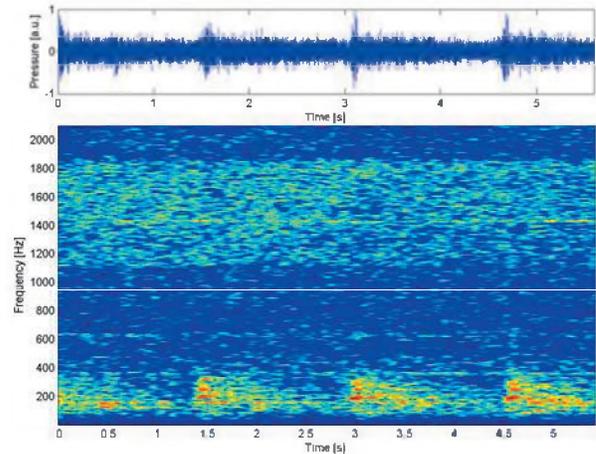
Figure 16: Tf-plot of air-gun transmissions.

For *clicks* the sequence starts with dilate steps in vertical direction followed by erode steps. This will fill the gaps between segments of a broken line. Furthermore we erode in horizontal direction to remove reverberation that tends to attach consecutive clicks. For *whistles* a similar procedure is followed, but horizontal and vertical are interchanged. For *moans* it is important to get rid of artificial vertical connections (due to imperfect FFT filtering) to separate the islands. Here erode steps in both direction are useful. *Sweeps* that have a 2-dimensional structure are best left alone. The number of erode-dilate steps is an important tuning parameter. It depends on the quality of the data and on the pre-processing. In general it can be remarked that it is better to use a low threshold and apply many erode steps, than to use high thresholds and dilate steps.

3.3.3 Classification

In order to classify a detected sound the measured features of a cluster are compared to those of the standard sounds. Below an abbreviated procedure is given:

1. Clicks; we demand the cluster to have small aspect and considerable height. Apart from single cluster features, we also check whether the cluster is repetitive, *i.e.* we check if a group of lines is present.
2. Moans; we demand the cluster to be compact (medium aspect) and repetitive in frequency.
3. Whistles; we demand the aspect to be large and the filling to be poor. There is a check for harmonics, which concludes whether the whistle is biological or man-made.
4. Sweeps; as for whistles, but with a special demand for the third moments for obtain skewness.

When a structure is not recognised as an air-gun, or any of our list the transient is unclassified.
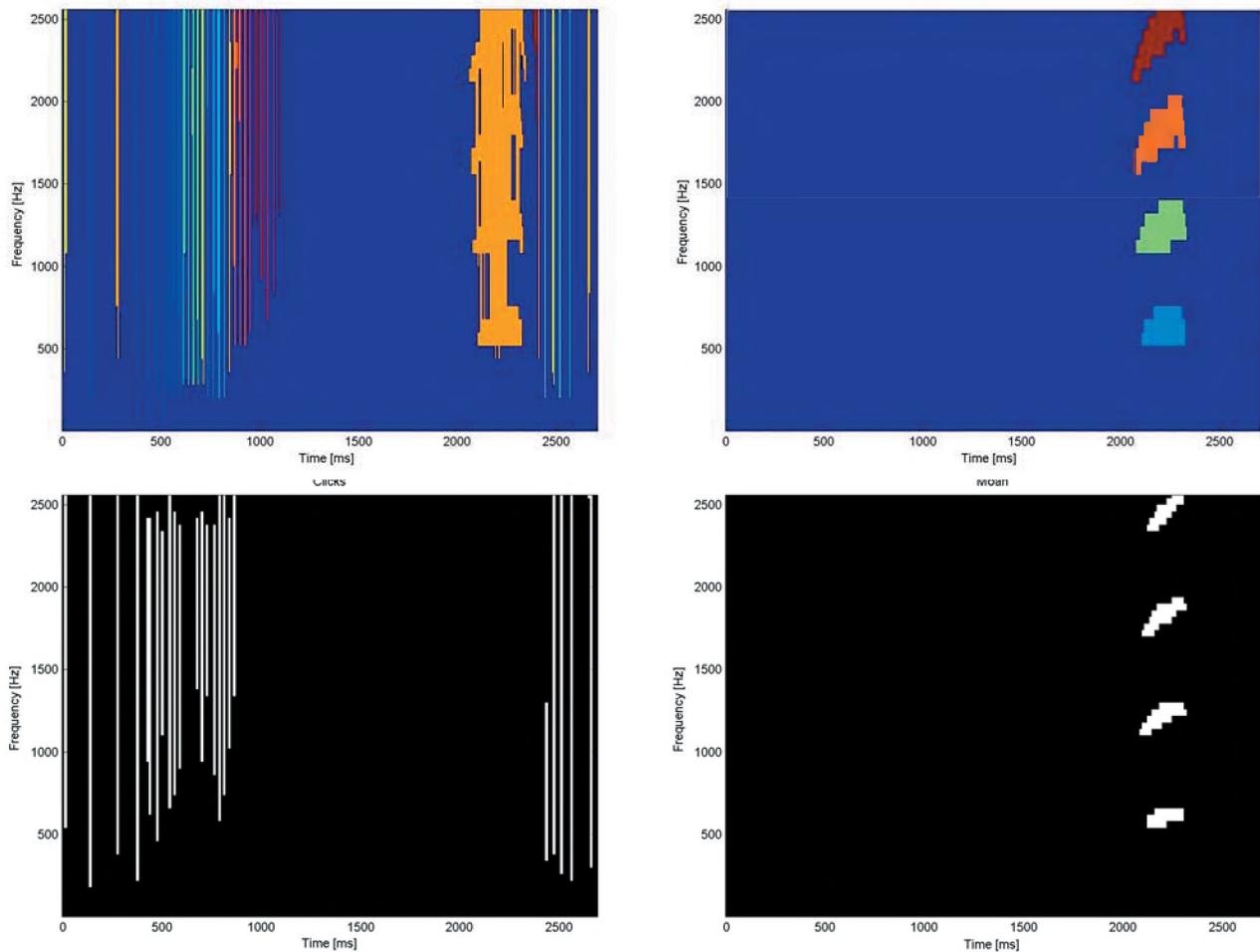
Figure 17: Example of erode-dilate processing on clicks (left) and harmonics (right). Input is the final plot of Figure 15 with clusters. For clicks erode-dilate aims at filling the gaps between the fragments. Here DDDEEE in vertical direction is applied. For the harmonics the moan detector is used. Here the bridges between the islands should be broken. Here ED in vertical and in horizontal is applied. The results are shown in the upper panels. Below pattern recognition is applied, which recognises the clicks on the left and the harmonics on the right.

4   CONCLUSIONS

The proposed transient or marine mammal detector can be separated into two basic steps. The first step is pre-processing and beamforming. This step is used to improve the signal to noise ratio and to obtain direction information on the detection. For this purpose a new type of beamformer is developed, with a constant beamwidth in the full frequency band.

The second step is automatic detection of the wide variety of marine mammal vocalisations. This is achieved by the combined use of a power-law and Page's test algorithms. The power-law integrator is robust against varying signal bandwidth while the Page's test detector is a robust detector for signals with an unknown duration.

This combination of sector beamforming and power-law/Page's test detector seems to be very promising in detecting marine mammal vocalisations; see also [17]. In

an application to an LFAS array it proved possible to detect high frequency dolphin sweeps as well as low frequency clicks from a large whale during sea trials.

The classifier is still under development. Algorithms are implemented, but tests of the classifier on recorded transients are pending. The amount of useful CAPTAS data are still limited. Some marine mammal transients from the Internet were gathered and the score on those was fair (especially for clicks and moans), but not exceptional. However, the quality of these recordings differs a lot (in noise level, filters, etc.). The algorithms are sensitive to the exact settings of the detector and therefore tuning of classification parameters for arbitrary WAV-files is cumbersome. If sufficient CAPTAS data is available a well-trained classifier could be developed as the proposed algorithms seem quite robust. A final step would be the inclusion of a localiser. The animal's range

is an essential parameter in mitigation measures. Ideas for this are being developed [22].

REFERENCES

[1] *R.W Richardson et al*, Marine mammal and noise, *Academic Press, San Diego, 1995.*

[2] *National Research Council (2000)*, Marine mammals and low frequency sound (progress since 1994), *National Academic Press, Washington DC.*

[3] *K.M. Fristrup, L.T. Hatch and C.W. Clark*, Variation in humpback whale (Megaptera novaeangliae) song length in relation to low-frequency sound broadcasts, *JASA 113(6), pp341-3424, June 2003.*

[4] *D.M.F. Chapman and D.D. Ellis*, The elusive decibel: thoughts on sonars and marine mammals, *Canadian Acoustics 26(2), pp29-31, 1998.*

[5] *C. Erbe and D.M. Farmer*, A software model to estimate zones of impact on marine mammals around anthropogenic noise, *JASA 108(3), pp1327-1331, Sept 2000.*

[6] *R.D. McCauley, J Fewtrell & A.N. Popper, (2003)*, High intensity anthropogenic sound damages fish ears, *JASA 113 (1).*

[7] *F.P.A. Benders, S.P. Beerens & W. Verboom*, SAKAMATA: the ideas and algorithms behind it, *Proceedings of the ECUA 2004 conference, July 2004.*

[8] *G.H. Hotho*, Adaptation of a submarine transient detector to the detection of whale sounds, *Proeedings of UDT Europe 1999.*

[9] *D.A. Abraham*, Passive acoustic detection of marine mammals, *Saclantcen Memorandum SM-351, April 2000.*

[10] *D.A. Abraham, R.J. Stahl*, Rapid detection of signals with unknown frequency content using Page's test, *Proceedings of 1996 Conference on information Sciences and Systems, March 1996, pp. 809-814.*

[11] *A.H. Nuttall*, Detection performance of power-law processors for random signals of unknown location, structure, extent, and strength, *Tech. Rep. 10751, Naval Underwater Systems Center, September 1994.*

[12] *E.S. Page*, Continuous inspection schemes, *Biometrika, vol. 41, pp 100-114, 1954.*

[13] *G.W.M. van Mierlo, S.P. Beerens, R. Been, Y. Doisy and E. Trouve*, Port-Starboard discrimination on hydrophone triplets in active and passive towed arrays, *Proceedings of UDT Europe 1997, pp. 176-181, 1997.*

[14] *S.P. Beerens, R. Been, J Groen, Y. Doisy and E. Noutary*, Adaptive Port-Starboard beamforming for triplet arrays, *Proceedings of UDT Pacific, 63-68, 2000.*

[15] *R.O. Nielsen*, Sonar signal processing, *Artech house 1991.*

[16] *L.J. Karam., and J H. McClellan*, Complex Chebyshev Approximation for FIR Filter Design,. *IEEE Trans. on Circuits and Systems II, March 1995, pp. 207-216.*

[17] *R. Laterveer, S.P. Beerens, S.P. van IJsselmuide, M. Snyder & J Newcomb*, A high frequency transient detector suitable for monitoring marine mammal activity, *Proceedings of UDT Europe, 2004.*

[18] *R.J Urick*, Principles of underwater sound, *Peninsula publishing, 0-932146-62-7, 1983.*

[19] *W.C. Verboom, R.A. Kastelein*, Structure of harbour porpoise (Phocoena phocoena) click train signals, *De Spil Publishers, ISBN 90-72743-07-5, 1997.*

[20] *NOAA:* *http://www.noaa.gov/whales/*

[21] *Robert M. Haralick and Linda G. Shapiro*, Computer and Robot Vision, vol. I, *Addison-Wesley, 1992, pp. 158-205.*

[22] *S.P. Beerens, S.P. van IJsselmuide & M.K. Robert*, Localisation of Marine Mammals using an LFAS array , *Proceedings of the ECUA 2004 conference, July 2004.*

APPENDIX A: Flow chart of described power-law/Page's test detector.

Set parameters:
b_0, b_1, h_0, h_1, p, m, λ and X

and initialise:
$W = 0$, $t = 1$, $i_1 = 1$, $i_0 = 1$

Normalisation and power law

$$Y_j = \frac{X_j}{\lambda_j} \quad \text{and} \quad Z = \left[\frac{1}{m} \sum_{j=1}^{m} Y_j^p \right]^{\frac{1}{p}}$$

$t = t + 1$

$W < h_0$

$W \geq h_0$

$W = \max\{0, W + Z - b_0\}$

$W = \min\{h_0 + h_1, W + Z - b_1\}$

$W = 0$

$W \geq h_0$

$0 < W < h_0$

$W \leq h_0$

$W = h_0 + h_1$

No signal
Update
background
estimate (λ)

No signal,
don't update
background

Start of signal
$W = h_0 + h_1$
$i_1 = t$

End of signal
$W = 0$
$i_0 = t$

signal still
detected
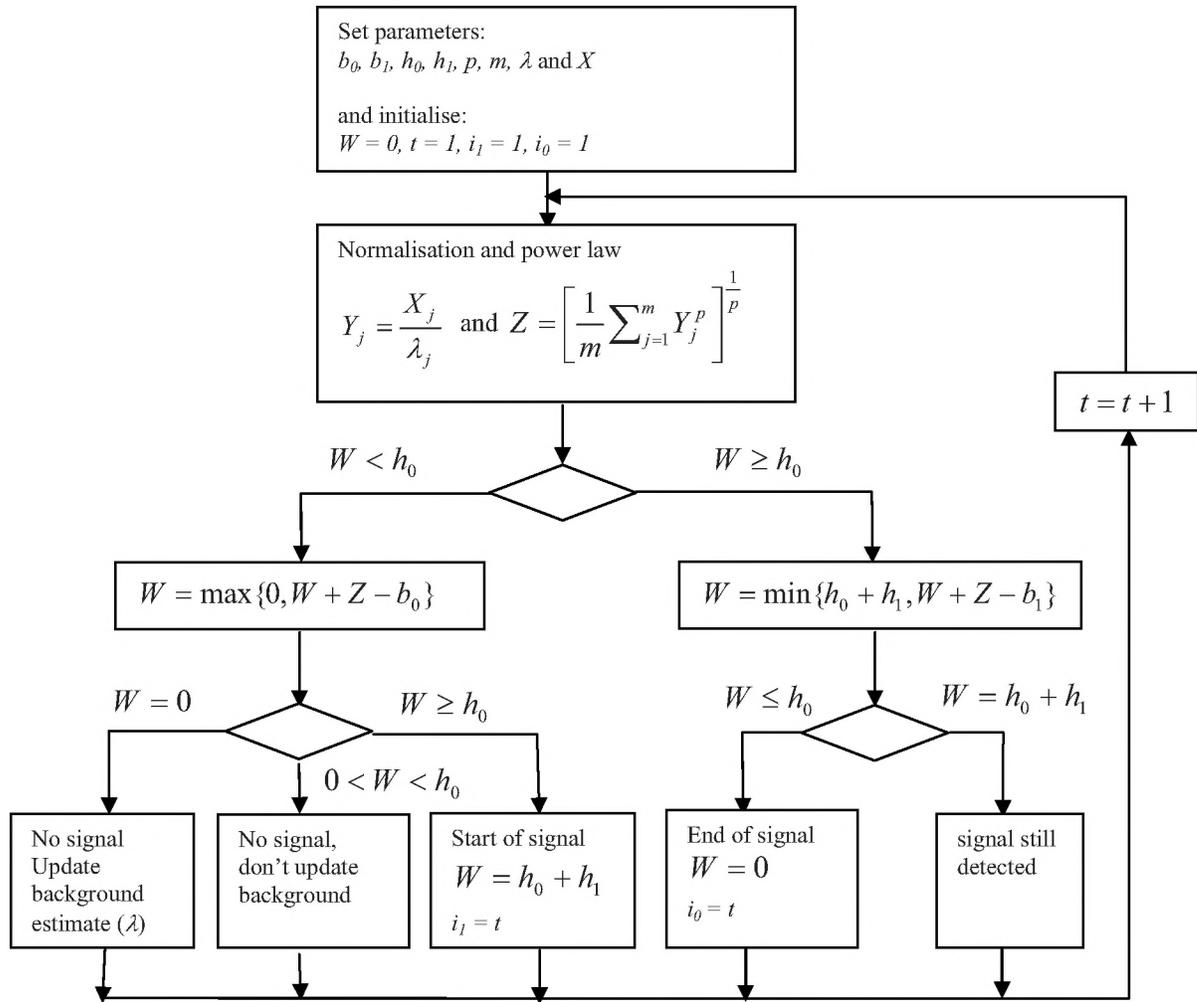
Figure A1: Block diagram of the Page's test detector.

Description of used variables

p — power law ($p \geq 1$)
h_0 — threshold for start of signal detection
b_0 — Page's test bias for start of signal detection
h_1 — threshold for end of signal detection
b_1 — Page's test bias for end of signal detection
α — time constant for exponential averaging of power spectra ($0 < \alpha < 1$)
Y — normalised power spectrum
Z — power-law output
W — Page's test statistic (is 0 if no signal and $h_0 + h_1$ if signal detected)
i_1 — start index of signal detected
i_0 — stop index of signal detected
j — frequency bin index number

ACOUSTIC DETECTION AND LOCALIZATION OF WHALES IN BAY OF FUNDY AND ST. LAWRENCE ESTUARY CRITICAL HABITATS

Yvan Simard[1,2], **Mohammed Bahoura**[3] and **Nathalie Roy**[2]

[1]ISMER, Université du Québec à Rimouski, P.O. Box 3300, Rimouski, Québec, Canada G5L 3A1
[2]Maurice Lamontagne Institute, Fisheries & Oceans Canada, P.O. Box 1000, Mont-Joli, Québec, Canada G5H 3Z4
[3]Département de Mathématiques, Informatique et Génie, Université du Québec à Rimouski,
P.O. Box 3300, Rimouski, Québec, Canada G5L 3A1

ABSTRACT

The detection and localization of marine mammals using passive acoustics is explored for two critical habitats in Eastern Canada. Two-dimensional hyperbolic localization is performed on time differences of arrivals of specific calls on grids of coarsely spaced autonomous recorders and on a shore-linked coastal array of closely spaced hydrophones. Delays are computed from cross-correlation and spectrogram cross-coincidence on signals enhanced with high-frequency emphasis and noise spectral suppression techniques. The outcomes and relative performance of the two delay estimation methods are compared. The difficulties encountered under the particular conditions of these two environments are discussed for the point of view of automated localisation for monitoring whales.

RÉSUMÉ

La détection et la localisation de mammifères marins à l'aide de l'acoustique passive est explorée pour deux habitats critiques dans l'est du Canada. La technique de localisation par hyperboles en deux dimensions est utilisée à partir des différences de temps d'arrivée à des réseaux de systèmes d'enregistrements autonomes largement espacés, ainsi qu'à un réseau serré d'hydrophones reliés à la côte. Les délais d'arrivée sont calculés par inter-corrélation ainsi que par inter-coincidence des spectrogrammes des signaux rehaussés par des techniques de rehaussement des hautes fréquences et de soustraction spectrale du bruit. Les résultats et la performance relative des deux méthodes sont comparés. Les difficultés rencontrées dans le contexte des conditions particulières de ces deux environnements sont discutées par rapport à l'automatisation de la localisation pour le monitorage des baleines.

1. INTRODUCTION

The localisation of living sound sources in the marine environment from the time difference of arrivals (TDoAs) at a series of receivers is several decades old (Watkins and Schevill 1972, Cummings and Holliday 1985). The most common localization method from large aperture arrays is hyperbolic fixing (Spiesberger and Fristrup 1990, Spiesberger 1999, 2001), though other simple (Cato 1998) or more elaborated model-based methods could be used (e.g. Tiemann and Porter 2003). With the fast development of electronic and computer technology, the setting up of such passive acoustic systems for non-intrusively monitoring whales in their environment is becoming increasingly available and spreading rapidly. This approach proved useful to gather information on the annual migrations of baleen whales over large oceanic basins (e.g.

Watkins et al. 2000). It is now sought for monitoring time-space use of habitat in intensively frequented meso-scale hot spots, eventually in real time, with the aim of improving their protection. Population density indices can also be estimated from such listening arrays (McDonald and Fox 1999), and used to follow its growth or displacement. Though the theory is well documented, its application in the field must be tuned to the particular characteristics of the local environment. This is especially important for implementing automated detection and localization algorithms. This paper is a preliminary exploration of the performance of simple techniques adapted to the conditions encountered in two critical habitats intensively visited by several species of whales during summer in eastern Canada, the Bay of Fundy and the Saguenay—St. Lawrence Marine Park.
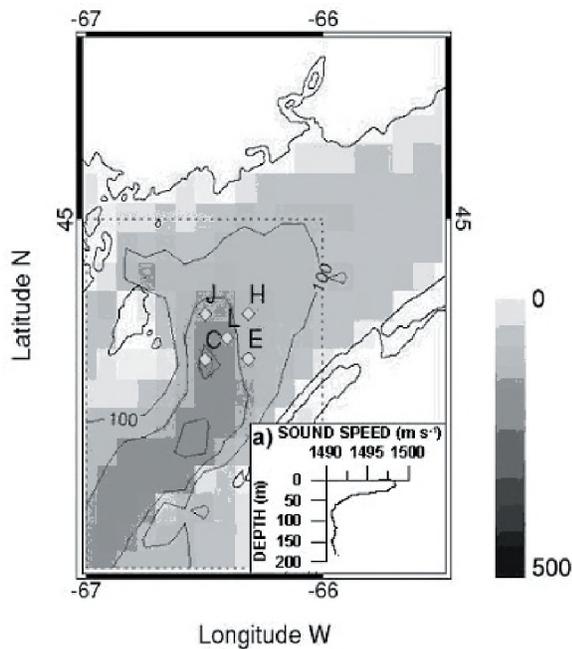
Figure 1. Bay of Fundy study area, with the bathymetry, the location of the 5 OBHs, and a typical sound speed profile.

2. MATERIAL AND METHODS

Data collection

The Bay of Fundy data set was collected in September 2002, with 5 ocean bottom hydrophones (OBHs), deployed in a centred square configuration with sides of 14.26 km, at
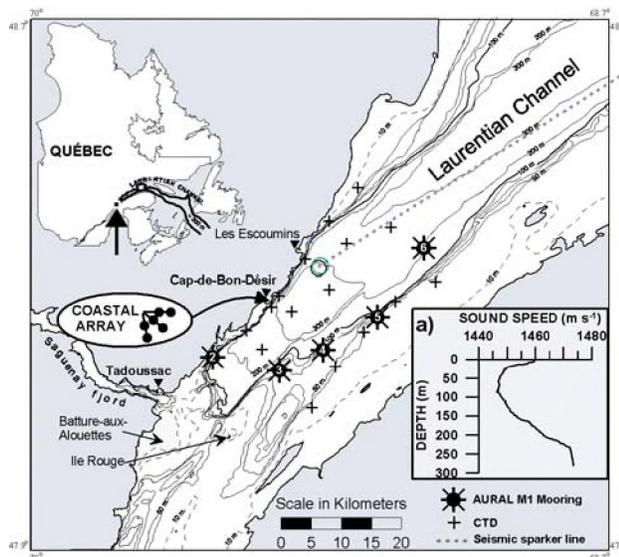
Figure 2. Study area in Saguenay—St. Lawrence Marine Park, with bathymetry, locations of the 6-hydrophone coastal array and the 5 AURAL M1 autonomous hydrophones, CTD stations and the track of a seismic-sparker RV (dotted line), with a typical sound speed profile.

the head of the ~200 m deep channel (Fig. 1). The OBH depths varied from 123 m to 210 m. The omnidirectional hydrophones (OAS model E-2SD, flat receiving sensitivity (RS) from 50 to 700 Hz) were 0.9 m from the bottom. The OBH positions were cross checked by interrogating their acoustic pinger and were accurate to 2 to 13 m. The clock drift over the 9-day deployment was negligible (<1 ms to 34 ms). The data were digitized with a 12-bit A/D converter sampling the 800 Hz low-pass signal at 1200 Hz. The OBH J RS was ~20 dB lower than the others. Temperature (XBTs) and conductivity (CTD) profiles (e.g. Fig. 1a) were performed during the experiment. A second data set was collected in August 2000 with 4 OBHs and a sampling frequency of 5000 Hz. A "calibration signal" representing right whale calls was then transmitted (source level of 155-160 dB re 1 μPa) from a rhib boat.

The St. Lawrence data sets were collected in August-September 2003 on the whale feeding ground at the head of the Laurentian channel (c.f. Simard et al. 2002), with a 6-hydrophone coastal array and a series of 5 autonomous hydrophones (AURAL M1, Multi-Electronics, Rimouski, QC, Canada) (Fig. 2). All hydrophones were omnidirectional HTI 96 MIN (flat RS from ~4 Hz to 30 kHz). The coastal array (Fig. 2) was deployed along Cap-de-Bon-Désir with three 600-m cables, each with 2 hydrophones, plunging into the sound channel (Fig. 2a). These hydrophones were ~5 m above the bottom. The array aperture was 657 m. The data were acquired without
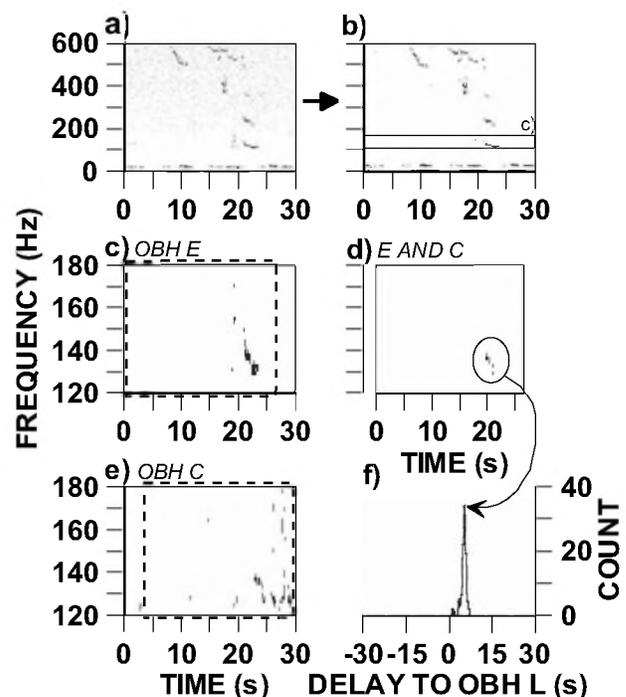
Figure 3. Example of computation of TDoA from spectrogram image cross-coincidence for Bay of Fundy low-frequency call S131-13 (see text).

interruption with fully-synchronous 16-bit A/Ds and DSPs mounted on a ChicoPlus data acquisition card (Innovative Integration, Ca, USA), sampling at 20 kHz. The exact hydrophone positions were determined from acoustic pulses transmitted from the R/V Coriolis II at a grid of stations off the array, where CTD profiles were also made
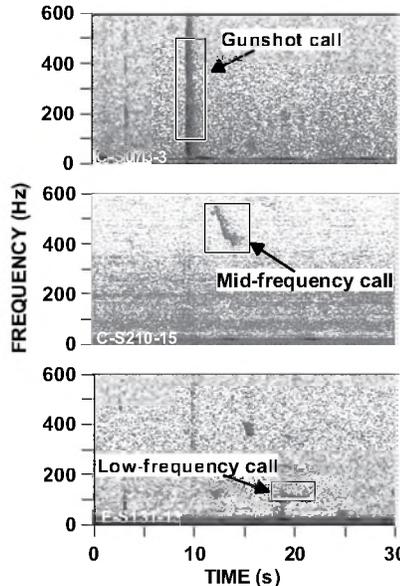
Figure 4. Spectrograms of OBH records showing the three types of northern right whale calls looked for in the Bay of Fundy data set.

for sound speed measurements. The AURAL M1 autonomous hydrophones were deployed in the sound channel (~50-60 m) on standard oceanographic moorings using sub-surface buoys. They were deployed 8-14 km apart along the border of the channel, in an arc facing the coastal array. Their position, as determined with DGPS, was precise to better than ~10 m, from crosschecks of the mooring echoes on the R/V scientific echosounders. The AURALs M1 record the depth and the ambient temperature besides the acoustic data. These 16-bit acoustic data were acquired at the 2000 Hz optional sampling rate of the AURALs M1, which includes a corresponding anti-aliasing (low-pass) filter. The internal clocks were synchronised to the microsecond with the PPS (pulse per second) signal of the GPS at the start of the recordings. The relative clock drifts were measured by synchronising all units at the recovery on a simultaneously recorded sound. CTD profiles were made at a grid of stations covering the study area at the beginning and the end of the recording period (Fig. 2).

Data analysis

The localization process from the TDoAs at the hydrophones proceeded in three steps. First, the frequency band of the selected whale call or anthropogenic sound was determined by visual inspection of the spectrogram (e.g. Fig. 3). Second, the signal was conditioned for TDoA finding algorithms, by high frequency pre-emphasis and noise spectral subtraction (Martin 2001) as follows (c.f. Fig. 4).

Pre-emphasis filter:

$$y_p(i) = y(i) - a\, y(i-1), \quad where\ 0.96 < a < 0.99 \qquad (1)$$
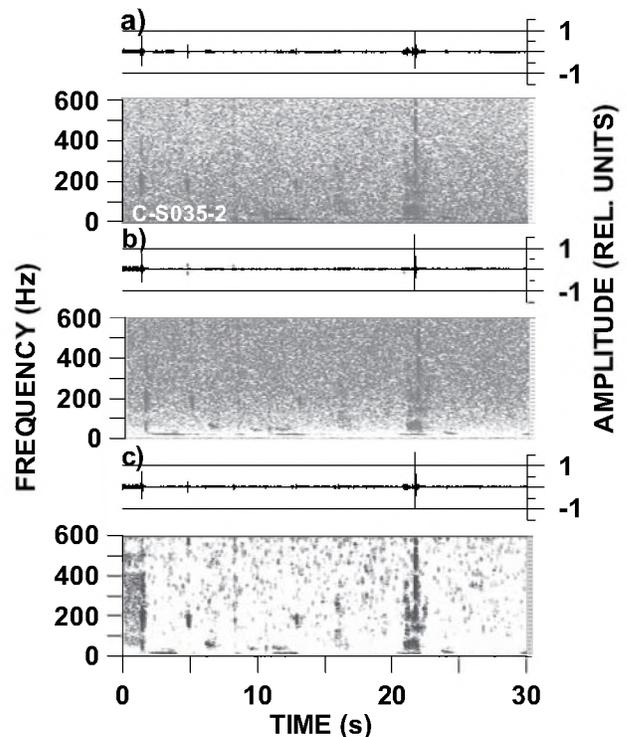
Figure 5. Waveforms and spectrograms of a OBH record containing a gunshot call, raw (a), after high-frequency pre-emphasis (b) followed by noise spectral subtraction (c).

Noise suppression:

$$y_p(t) = n(t) + x(t) \qquad (2)$$
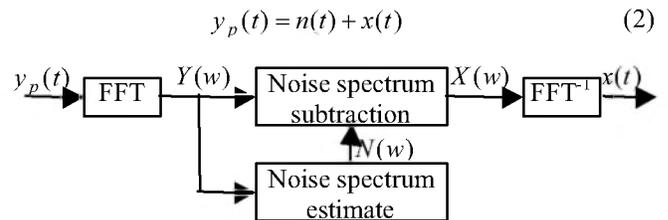
where $Y(w)$ and $N(w)$ are smoothed over window lengths chosen to maximize the difference between $x(t)$ and $n(t)$.

Third, the TDoAs between the hydrophones were computed on the waveform using cross-correlation. Data were first normalised to a 0-1 scale and then filtered (4th order high-pass or band-pass Butterworth) to keep only data in the selected call band. The absolute value of the cross-correlation series was low-pass filtered (2nd order Butterworth) to remove spikes hindering precise TDoA detection close to the maximum. The TDoAs were also computed from spectrogram "cross-coincidence" (Tiemann et al. 2001). The spectrogram of $y_p(t)$ or $x(t)$ is transformed to a binary image using a threshold value corresponding to the 95th or 99th percentile of the cumulative frequency distribution (cfd) of the spectrum values (Fig. 3a-b). The spectrograms are computed with a FFT window of 256 or
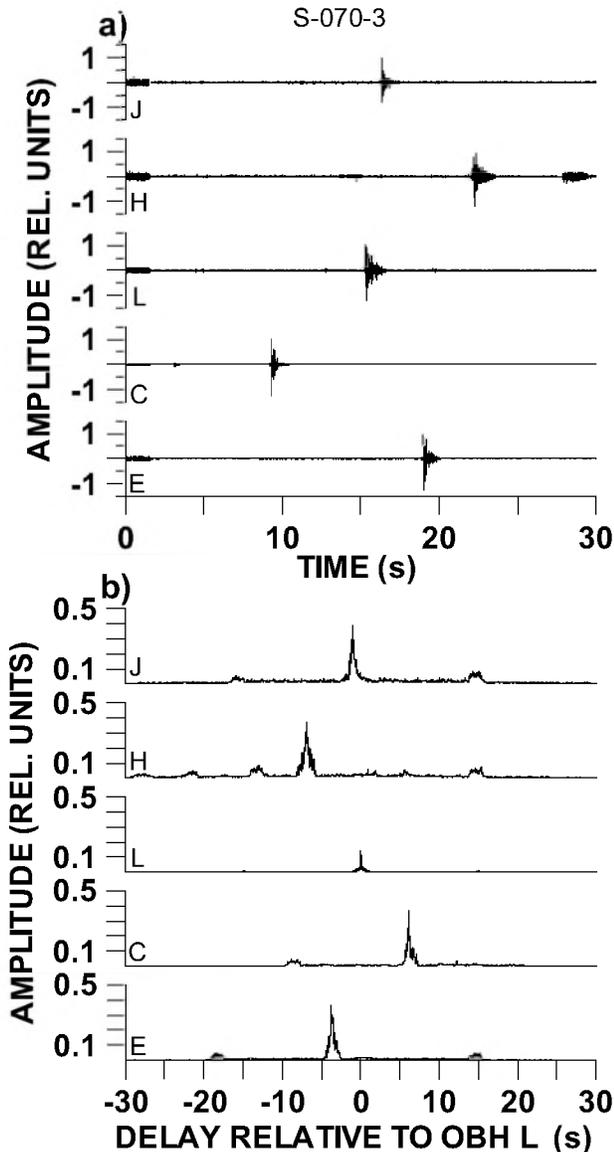
Figure 6. Conditioned and filtered OBH records (a) with their corresponding filtered cross-correlation series (b), for one gunshot call in the Bay of Fundy.

Figure 7. 2D hyperbolic localization of the gunshot call from Fig. 5 TDoAs. Position is: 44.6028° N, 66.5522° W. Rms error of the fixing was 197 m.

512 points, with 60% overlap. The frequency band of interest is extracted (Fig. 3c and e), and a logical *AND* is computed between the binary images of the hydrophone pair, for each time lag (Fig. 3c and e boxes). The resulting image for a given lag (Fig. 3d) has pixel values of 1 only when two positive pixels coincide on both images. The sum of these pixels represents the level of coincidence between the two spectrograms for the corresponding lag. A cross-coincidence series is obtained by expanding to all lags (Fig. 3f) for TDoA detection.

A constant sound speed of 1491 m/s, corresponding to the lower part (>50 m) of the water column (Fig. 1a), was used for the Bay of Fundy. For the St. Lawrence, it was

1450 m/s, which is the average speed in the sound channel where the hydrophones were deployed (Fig. 2a). Coordinates were transformed to (and from) Cartesian units using a Lambert projection. The 2D hyperbolic localization used the LocateDelays.m Matlab script (Dave Mellinger web site). This algorithm rejects delays that are larger than the maximum travel time between the hydrophone pairs given the constant sound speed. TDoAs that do not fit to this model are thus ignored for hyperbolic fixing. The predicted $TDoAs_c$ from the travel time differences between the estimated source location and the hydrophones are computed for the n valid hydrophone pairs, and the rms error relative to the observed $TDoAs_o$ is estimated as follows:

$$\sqrt{\sum_n (TDoA_o - TDoA_c)^2 / n - 2} \qquad (3)$$

The hyperbolic fixing uncertainty is obtained by converting this time error into distance error by multiplying by the sound speed.

3. RESULTS

The Bay of Fundy test data files provided for the workshop were separated into three types of North Atlantic right whale calls: gunshots, low-frequency and mid-frequency calls (Fig. 4). The selected frequency bands for these calls were respectively: 100 to 500 Hz, 100 to 180 Hz and 350 to 500 Hz for cross-coincidence, and 50 to 600 Hz, 100 to 300 Hz, and 400 to 600 Hz for cross-correlation. An example of the pre-conditioning of the signal is shown in Fig. 5. The TDoA estimation from cross-correlation is depicted in Fig. 6 for one gunshot sound. Resulting 2D hyperbolic fixing for that sound is shown in Fig. 7.
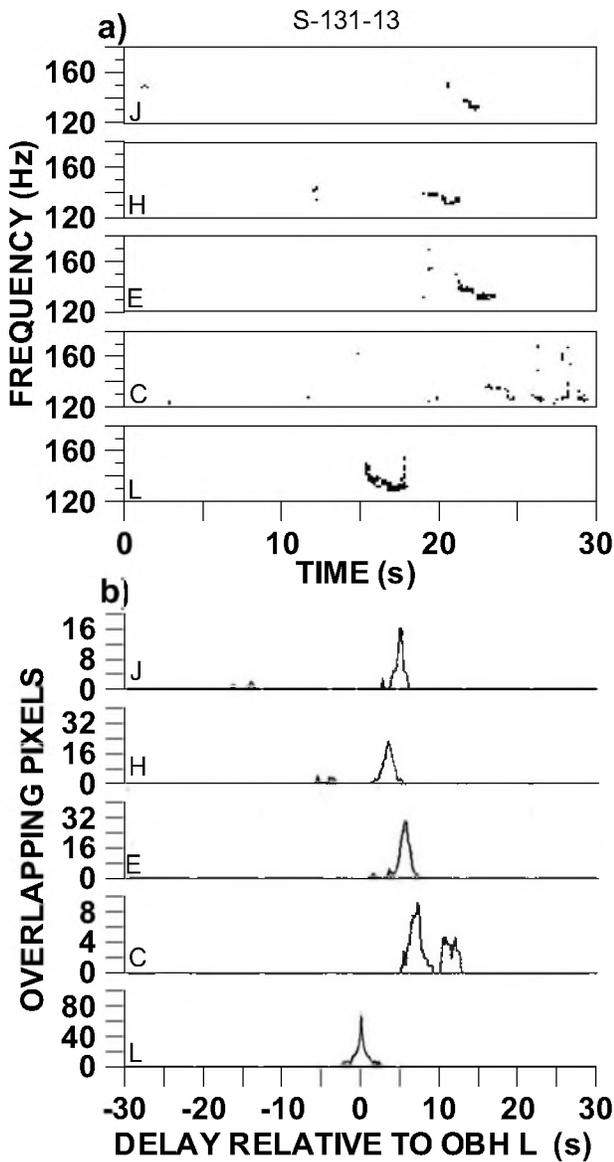
Figure 8. Binary images of the spectrograms of S-131-13 low-frequency call in Bay of Fundy for the five OBHs (a), and the corresponding cross-coincidence relative to OBH L (b).

Figure 9. 2D hyperbolic localization of the low-frequency call from Fig. 8 TDoAs. Position is: 44.6856° N, 66.3879° W. Rms error of the fixing was 381 m.
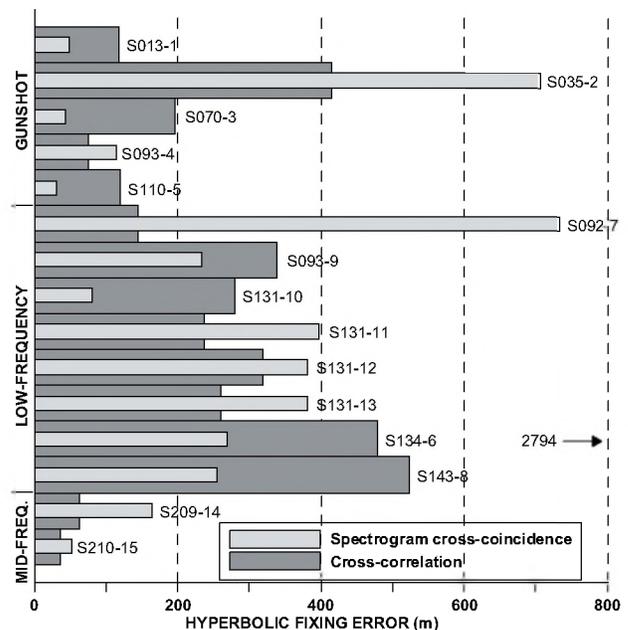
Figure 10. Comparison of hyperbolic fixing rms error obtained with the two TDoA estimation methods for Bay of Fundy call data set. Crosscorrelation for call S134-6 was done on $y_p(t)$ instead of $x(t)$, the latter producing an error of 2794 m.

TDoAs estimation with $y_p(t)$ spectrogram cross-coincidence is shown in Fig. 8 for a North Atlantic right whale low-frequency call recorded in Bay of Fundy. The localisation of the call is presented in Fig. 9. For the 15 North Atlantic right whale calls of the Bay of Fundy data set, the two TDoA estimation methods generally produced similar hyperbolic fixings (Fig. 10-11, Table 1). The differences between the two methods is generally less than 450 m, except for the distant mid-frequency calls, located more than 25 km from the nearest OBH (Table 1, Fig. 11). However, the fixing error (Table 1, Fig. 10) showed that the spectrogram cross-coincidence method had difficulties with

two calls and the cross correlation method with one call (see Discussion).

The binary images of the $x(t)$ spectrograms of a 30-s long low frequency beluga phrase, detected on the 5 AURAL M1 moorings in the St. Lawrence, is presented in
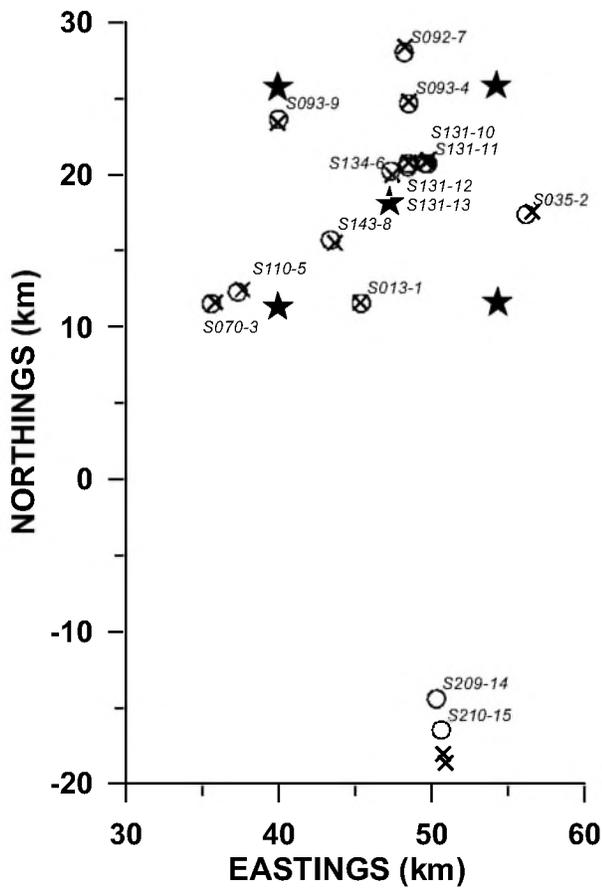
Figure 11. Localization of the 15 northern right whale calls of the Bay of Fundy data set, for TDoAs from spectrogram cross-coincidence (✕) and cross-correlation (○), with the positions of the OBHs (★).

Figure 12. Binary images of the spectrograms of a beluga low-frequency call phrase from the St. Lawrence Estuary five AURALs M1 moorings (a), and the corresponding cross-coincidence (b). The minke whale calls (dashed line box) were removed for computing the AURAL M1 #2 cross-coincidence series.

Fig. 12a. Its intensity is much higher on instruments #3 and #4. The record from instrument #2 has additional strong vocalisations, likely from close-by minke whales (Fig. 12a, dashed-line box). In computing the TDoAs for this call using spectrogram cross-coincidence, these minke whale calls had to be masked to get the right TDoA for the instrument #2, so that it corresponds to the TDoA estimate from manually inspecting the spectrograms. The localisation obtained that way is presented in Fig. 13. The one from the TDoAs obtained by manually inspecting the spectrograms differs from only 159 m from that position. The hyperbolic fixing rms error was large (2.5 km) in both cases. Figure 14a illustrates another example of a cluck clearly recorded on the AURAL M1 moorings, except for the instrument # 5 where it was severely masked by flow noise. The TDoAs estimated from cross-correlation of the $y_p(t)$ series were the same as those obtained from manually inspecting the spectrograms. The hyperbolic fixing used only a few of them though (Fig. 14b), the other ones were exceeding the expected maximum delays from the assumed sound speed

and declared invalid. The pings of a towed seismic sparker echosounder were used to localise a R/V working in the area from the AURAL M1 recordings (Fig. 2). All methods failed to find the TDoAs. A closer look at the spectrograms showed that some pings at the start of the sounding line were missing on two instruments in the narrow bandwidth (1 kHz) of the observations (source peak was ~ 2.2 kHz from the coastal array). When corrected for these missing pings, the TDoAs obtained by manually inspecting the spectrograms successfully localised the R/V at the start of its sounding line (Fig. 2, circles). The error with the true DGPS position of the 50-m R/V was 233 m, which is very
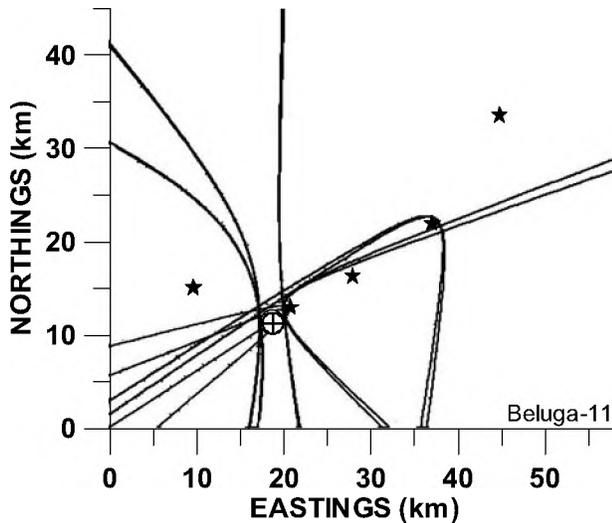
Figure 13. 2D hyperbolic localization of the low-frequency beluga call from Fig. 12 TDoAs. Position is: 48.1553° N, 69.4697° W.

small given that the distance between the DGPS antenna and the towed sparker was larger than 70 m.

A 1600-2600 Hz beluga whistle (Figs.15) from the 6-hydrophone coastal array was localized using TDoAs from spectrogram cross-coincidence. It was recorded 12 min before the beluga-11 call (Figs. 12-13) and localised in the same part of the observed area, 5 km off the array and 7.8

km away from the beluga-11 call (Fig. 16).

4. DISCUSSION

This exercise of localizing whale calls using passive acoustics in two critical habitats in eastern Canada gives an example of the performance of simple techniques in actual conditions at sea. The accuracy of source localization depends on precision of measurements of TDoAs, hydrophone positions, sound velocity and the geometry of the hydrophone network (Wahlberg et al. 2001). Precise estimation of TDoAs is critical for accurate localization. This relies on both the acquisition and the processing of the data. Substantial effort has been dedicated to precise 3D localization of the hydrophones and accurate synchronization of all recording clocks in both study areas. Though the error due to the equipment may be minimized, it is not zero because of the difficulty of accurate x y z positioning of the receivers at sea, fluctuating sound speed structures and water depth with tides, and tilting of the mooring line or displacement of bottom mounted instruments with strong currents. The level of precision required for the 3D position of the hydrophones is particularly high for the coastal array, because of the close spacing of the hydrophones and the very small TDoAs of the calls.

Table 1. 2D hyperbolic localization of Bay of Fundy northern right whale calls using TDoAs computed with spectrogram cross coincidence and cross-correlation.

File	Type	Spectrogram cross-coincidence						Cross-correlation					
		Band (Hz)	FFT (pt)	Cfd cut-off	Lat. N	Long. W	error (m)	error (m)	Lat. N	Long. W	Band (Hz)	X-correlation Low-pass filter cut-off (Hz)	Localization differences (m)
S013-1	Gunshot	100-500	256	0.99	44.6027°	66.4289°	49	118	44.6025°	66.4284°	50 - 600	36	45
S035-2	Gunshot	100-500	256	0.99	44.6559°	66.2865°	707	415	44.6541°	66.2916°	50 - 600	36	451
S070-3	Gunshot	100-500	256	0.99	44.6036°	66.5493°	42	197	44.6028°	66.5522°	50 - 600	36	247
S093-4	Gunshot	100-500	256	0.95[1]	44.7216°	66.3876°	115	75	44.7203°	66.3880°	50 - 600	36	148
S110-5	Gunshot	100-500	256	0.95[2]	44.6112°	66.5264°	30	120	44.6096°	66.5303°	50 - 600	36	357
S092-7	Low-frequency call	100-180	512	0.99	44.7538°	66.3908°	734	144	44.7506°	66.3914°	100 - 300	12	359
S093-9	Low-frequency call	100-180	512	0.99	44.7095°	66.4969°	234	339	44.7117°	66.4958°	100 - 300	12	260
S131-10	Low-frequency call	100-180	512	0.99	44.6858°	66.3741°	80	279	44.6850°	66.3753°	100 - 300	12	130
S131-11	Low-frequency call	100-180	512	0.99	44.6867°	66.3728°	397	237	44.6846°	66.3727°	100 - 300	12	233
S131-12	Low-frequency call	100-180	512	0.99	44.6856°	66.3879°	381	319	44.6850°	66.3887°	100 - 300	12	92
S131-13	Low-frequency call	100-180	512	0.99	44.6856°	66.3879°	381	260	44.6831°	66.3891°	100 - 300	12	294
S134-6	Low-frequency call	100-180	512	0.99	44.6785°	66.4017°	269	479[3]	44.6806°	66.4032°	100 - 300	12	262
S143-8	Low-frequency call	100-180	512	0.99	44.6382°	66.4503°	255	523	44.6402°	66.4534°	100 - 300	12	331
S209-14	Mid-frequency call	350-500	512	0.99	44.3357°	66.3641°	164	62	44.3684°	66.3688°	400 - 600	6	3653
S210-15	Mid-frequency call	350-500	512	0.99	44.3303°	66.3619°	53	37	44.3500°	66.3650°	400 - 600	6	2203
S282	Calibration call	420-480	512	0.99	44.6945°	66.3801°	430						
S288	Calibration call	525-580	512	0.99	44.6945°	66.3802°	271						
S289	Calibration call	525-580	512	0.99	44.6943°	66.3807°	354						

[1] Failed with a cut-off of 0.99; [2] Less precise with a cut-off of 0.99; [3] Without noise spectral subtraction.
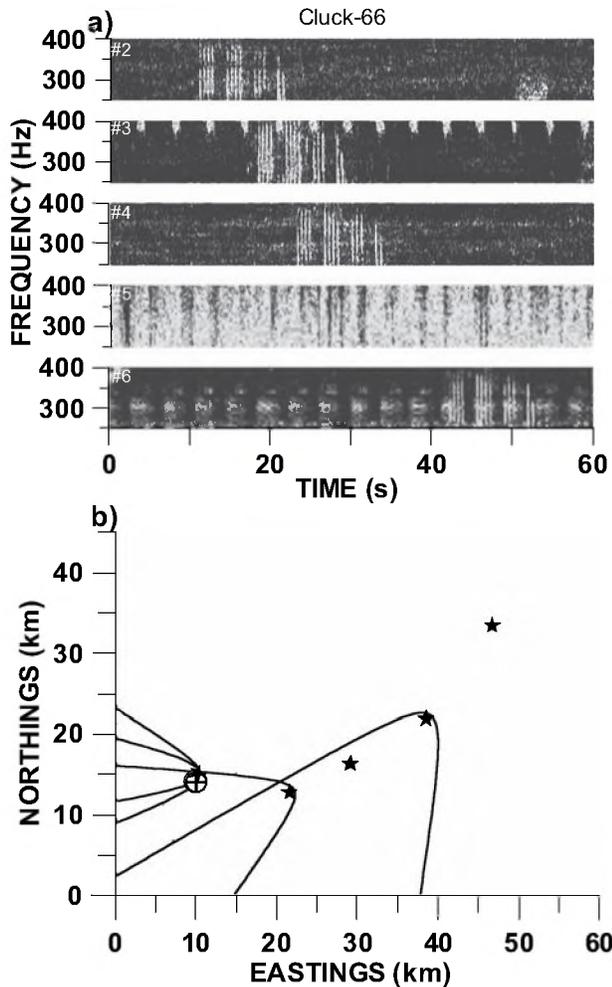
Figure 14. Spectrogram of cluck for the 5 AURAL M1 moorings in the St. Lawrence (a) and its hyperbolic fixing using TDoAs from cross-correlation (b).

Precise TDoAs also relies on signal strength relative to noise (SNR) at each hydrophone of the localization network. The three data sets showed that this is very variable and not only depends on propagation effects and travel distances, but also on masking noise (from shipping, flow, etc.). The low sensitivity of OBH J was however involved in some cases. The conditioning of the data for optimal TDoA detection helped to cancel out some of these effects. The two signal-processing steps we used to increase the SNR before computing the TDoAs proved useful to handle most calls with the same algorithm. Exceptions were encountered where the noise spectral suppression also removed the faint signals (e.g. cluck call of Fig. 14, and S134-6 call, Table 1). A step should therefore be added here to decide when noise suppression should be employed, and which parameters are best suited to the type of call considered. The spectrogram cross-coincidence method required noise spectral suppression only in very low SNR conditions, such as when shipping noise was high at some hydrophones, which was
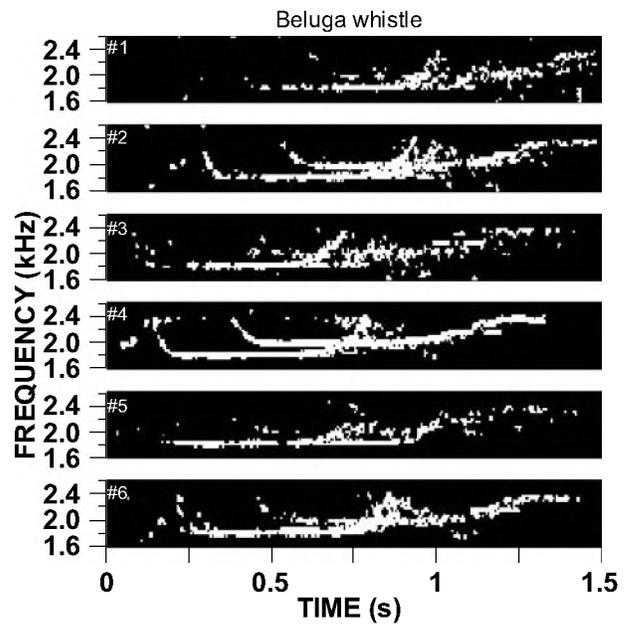
Figure 15. Binary images of the spectrograms of a beluga whistle from the St. Lawrence Estuary 6-hydrophone coastal array.

the case for the beluga call of Fig. 12. For the transformation of the spectrogram into a binary image, low SNR sometimes forced the lowering the cumulative frequency distribution cutoff from 0.99 to 0.95 (e.g. gunshot calls S093-4 and S110-5 of Table 1). Very low SNRs for OBH J and C are at the origin of the two large fixing errors for calls S035-2 and S092-7 with the spectrogram cross-coincidence method (Table 1). In this case, it would be better to drop the low SNR OBH and perform the hyperbolic fixing with only four instruments. For an unsupervised automatic fixing algorithm, another decision step should be added to reject too low SNR recordings. The filtration of the series, to remove the spikes that often occur close to the maxima before the peak detection, also appeared necessary for more
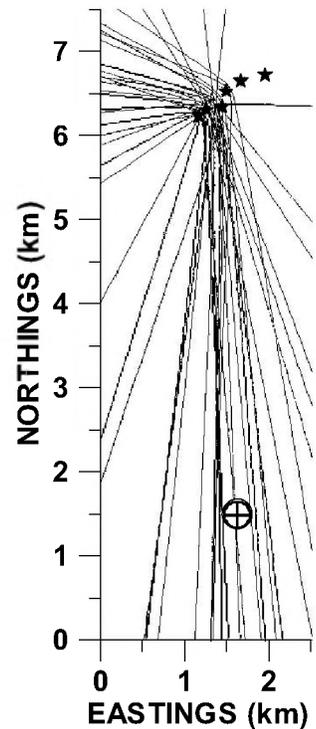
Figure 16. Hyperbolic fixing of the beluga call of Fig. 15 using TDoAs from spectrogram cross-coincidence.

robust TDoA detection with the cross-correlation method. A supervised decision was necessary to get the TDoA for the beluga call on the AURAL M1 #2 when close minke whale calls prevented its accurate estimation. This is likely to occur in critical habitats that are frequented by several whales, such as the Saguenay—St. Lawrence Marine Park. The masking of concurrent calls is then necessary and could be accomplished by connecting the TDoA finding algorithm with the prior step of call detection and classification.

The geometry of the hydrophone network is of course another important aspect affecting the precision of the localization. The centered square configuration of the Bay of Fundy OBHs, with a relatively small total width (14.26 km) insured close enough spacing (10.36 km) between all hydrophones to receive the call with a good SNR on all instruments in most cases. The arc shape of the St. Lawrence AURAL M1 configuration (which resulted from the loss of one instrument in a planned U-configuration) is less effective because of the solution for the left-right ambiguity is dependent on a single instrument, and because of the large distances (> 20 km) between the distant hydrophones. Propagation effects then become important, and the conditions are far from the linearity assumption of hyperbolic fixing (Spiesberger and Wahlberg 2002). The arrival times were increasingly late, as a function of the travel time, compared to the assumed direct path at a constant sound speed. The vertical sound speed gradient in the St. Lawrence is about three times larger than in the Bay of Fundy (c.f. Figs. 1a and 2a). This resulted in the dropping of those TDoAs exceeding the expected maximum delays between the instruments, and the localization with only a few instruments (e.g. Fig. 14). The sound speed should be allowed to change with travel time, as proposed by Spiesberger and Wahlberg (2002). A multipath propagation model (e.g. Tiemann et al. 2001) should therefore be used for proper source localization for ranges larger than the few kilometers where the direct path assumption is valid in these shallow environments. Another relevant aspect of receiver geomerty is the vertical localization of the hydrophone. The Bay of Fundy and the St. Lawrence coastal array hydrophones were placed close to the bottom and therefore subject to shadow zones and interference with bottom reflections. These latter were likely contributing to errors in TDoA detection. For the St. Lawrence coastal array, the delay error could be proportionally large because of the close spacing of the hydrophones. This could make localizing the source difficult, as we observed. The coastal array was placed along a cape in the St. Lawrence. This localization facilitated the deployment to rapidly access the sound channel. However, the proximity of the shore and cape wall gave rise to strong reflections and multipaths, which can sometimes hinder precise detection of the TDoAs. The St. Lawrence AURALs M1 were placed in the sound channel to maximize the reception range. Some

instruments were however moored on the southern border of the deep channel, which unfortunately placed them within the St. Lawrence outflow (Saucier and Chassé 2000). They were therefore subject to flow noise, which often masked the calls. Both critical habitats considered here are high-energy environments with strong tidal forcing (e.g. Saucier and Chassé 2000). It is therefore inaccurate to assume a constant propagation medium in space and time. The changes of the characteristics of the propagation medium must therefore be incorporated in the localization process to minimize the error. This can be accomplished with repeated visits of a grid of stations for CTD profiling, or the use of a ground-truthed 3D tidal circulation model. Frequent checks of the performance of the localization algorithm with transmitted sounds from known locations are likely to be essential for accurate monitoring with passive acoustics. The deployment of fixed acoustic pingers regularly transmitting a sound in the study area during the observation period should help to monitor the localization performance and take into account the main components of its variability.

5. ACKNOWLEDGEMENTS

6. REFERENCES

Cato, D.H. 1998. Simple methods of estimating source levels and locations of marine animal sounds. J. Acoust. Soc. Am. 104: 1667-1678.

Cummings, W.C., Holliday, D.V., 1985. Passive acoustic location of bowhead whales in a population census off Point Barrow, Alaska. J. Acoust. Soc. Am. 78: 1163-1169.

Martin, R. 2001. Noise power spectral density estimation based on optimal smoothing and minimum statistics. Speech and Audio Processing, IEEE Transactions 9: 504-512.

McDonald, M.A., and Fox, C.G. 1999. Passive acoustic methods applied to fin whale population density estimation. J. Acoust. Soc. Am. 15: 2643-2651.

Saucier, F.J., and Chassé, J. 2000. Tidal circulation and buoyancy effects in the St. Lawrence estuary. Atmosphere-Ocean 38: 505-556.

Simard, Y., Lavoie, D., and Saucier, F.J. 2002. From plankton to whales: Oceanography of a traditional whale feeding ground and marine park in the St. Lawrence estuary. ICES CM 2002/N:14.

Spiesberger, J.L. 1999. Locating animals from their sounds and tomography of the atmosphere: Experimental demonstration. J. Acoust. Soc. Am. 106: 837-846.

Spiesberger, J.L. 2001. Hyperbolic location errors due to insufficient numbers of receivers. J. Acoust. Soc. Am. 109: 3076-3079.

Spiesberger, J.L., and Fristrup, K.M. 1990. Passive localization of calling animals and sensing of their acoustic environment using acoustic tomography. Am. Nat. 135: 107-153.

Spiesberger, J.L., and Wahlberg, M. 2002. Probability density functions for hyperbolic and isodiachronic locations. J. Acoust. Soc. Am. 112: 3046-3052.

Tiemann, C.O., Porter, M.B., and Fraser, L.N. 2001. Automated model-based localization of marine mammals near Hawaii. In Conference Proceedings of Oceans 2001 MTS/IEEE, pp.1395-1400.

Tiemann, C.O., and Porter, M.B. 2003. A comparison of model-based and hyperbolic localization techniques as applied to marine mammal calls. J. Acoust. Soc. Am. 114: 2406.

Wahlberg, M., Mohl, B., and Madsen P.T. 2001. Estimating source position accuracy of a large-aperture hydrophone array for bioacoustics. J. Acoust. Soc. Am. 109: 397-406.

Watkins, W. A., and Schevill, W.E. 1972. Sound source location by arrival-times on a non-rigid three-dimensional hydrophone array. Deep-Sea Res. 19: 691-706.

Watkins, W.A., Daher, M.A., Reppucci, G.M., George, J.E., Martin, D.L., DiMarzio, N.A., and Gannon, D.P. 2000. Seasonality and distribution of whale calls in the North Pacific. Oceanography 13: 62-67.

A Two-Stage Method For Determining The Position And Corresponding Precision Of Marine Mammal Sounds

Dick G. Simons[a,b], Camiel van Moll[a], Mirjam Snellen[b]

[a]Underwater Acoustics Group, TNO Physics and Electronics Laboratory, Oude Waalsdorperweg 63, P.O.Box 96864, 2509 JG, The Hague, The Netherlands, e-mail: Simons@fel.tno.nl
[b]Mathematical Geodesy and Positioning, Faculty of Aerospace Engineering, Delft University of Technology Kluyverweg 1, 2629 HS, Delft, The Netherlands

ABSTRACT

Today there is a concern that man-made sounds, such as that from sonar experiments, seismic operations and oil rigs, affect marine mammals. Detection and localisation of marine mammals will definitely support measures to reduce the possible detrimental effects. This paper presents a two-stage localisation method, which is applied to data collected with an array of five hydrophones moored to the seabed in the Bay of Fundy, Canada. The array forms a 14 by 14 km square with one hydrophone in the centre. The method makes use of the relative travel times of the mammal's sound to the four hydrophones at the square vertices with respect to the travel time to the central hydrophone. First, a good initial position is obtained using hyperbolic fixing. In the second step the solution is improved in an iterative process, where each iteration determines the least-squares solution of the set of four linearized equations for the measured relative travel times. Calculating the error ellipse from the covariance matrix of the solution provides the localisation accuracy. There are several parameters that affect the source position accuracy. These include the uncertainties in arrival times, sound speed and receiver positions. Their effect on the localisation accuracy is assessed. [Work supported by Royal Netherlands Navy]

RÉSUMÉ

Aujourd'hui, une question préocupante est de savoir si des sons d'origine artificielle tels que ceux générés par les systèmes sonars, les opérations sismiques ou les installations pétrolières peuvent affecter les mamifères marins. La détection et la localisation des mamifères marins est un atou indéniable afin de réduire d'éventuels effets indésirables. Cet article présente une méthode de localisation en deux étapes appliquée à des données collectées par une antenne de cinq hydrophones amarrés au fond marin dans la baie de Fundy au Canada. L' antenne forme un carré de 14 par 14 kilomètres avec un hydrophone placé au centre. La méthode utilise la différence relative de temps de parcours des cris des mamifères marins entre les hydrophones situés aux sommets du carré et le temps de parcours jusqu'à l'hydrophone situé au centre. Premièrement, une bonne estimation initiale de la position est obtenue grâce à une correction hyperbolique. Dans la deuxième étape, cette solution est améliorée grâce à un processus itératif où chaque itération donne la solution au sens des moindres carrés d'un ensemble de quatre équations linéarisées obtenues grâce aux temps de parcours relatifs mesurés. Le calcul de l'ellipse d'erreur à partir de la matrice de covariance de la solution donne la précision de la localisation. Plusieurs paramètres affectent la précision de la position de la source. Ceux-ci incluent l'incertitude sur les temps d'arrivé, la vitesse du son et la position des récepteurs. Leurs effets sur la localisation sont évalués. [Travail subventionné par la Marine Royale Hollandaise]

1. INTRODUCTION

Marine mammals rely on their vocalizations for orientation, communication and hunting. There is an increasing concern that man-made acoustic signals are harmful to these animals. This has resulted in an increased research effort on passive acoustic techniques for detection and localisation of the mammals. This article focuses on accurate localisation.

To promote research on this topic, a workshop on detection and localisation of marine mammals using passive acoustics was held at Dartmouth, Nova Scotia, Canada, 19-21 November 2003. A dataset of marine mammal vocalizations was provided.

This data set has been used to test the localisation procedure presented here. This procedure relies on basic concepts of geodesy:

- application of the elementary adjustment principles of least squares to combine redundant measurements in an optimal way (with unbiased minimum variance)

- confidence intervals to quantify the uncertainty of calculated positions.

The procedure requires a proper starting solution, which is obtained through hyperbolic fixing. Both concepts are applied assuming free propagation in an unbounded homogeneous medium, i.e., spherical propagation where the sound is assumed to propagate along straight paths. The method is applied to data collected on a symmetrical array of five omni-directional hydrophones moored to the seabed in the Bay of Fundy, Canada. The array forms a 14 by 14 km square with one hydrophone in the center. The localisation method uses all four available travel time differences, estimated with respect to the central hydrophone L.

Section 2 describes the approach taken for determining the relative travel times and their uncertainties. Section 3 presents the localisation procedure. Section 4 presents the results. Also, in Section 4, the effect of the uncertainty in the arrival times, the uncertainty in the sound speed and the uncertainty in the receiver positions on the localisation accuracy are assessed. It is followed by the conclusions in Section 5.

2. TRAVEL TIME ESTIMATION

Since the moment at which the sound was emitted is unknown, only relative arrival times are available for the localisation. The data provided for the workshop consist of three distinct types of sound: "gunshots", mid-frequency calls and low-frequency calls. The approach selected for determining the relative travel times depends on the characteristics of the signal. For this paper we will consider both the gunshots and the mid-frequency calls.

The gunshots are broadband high-amplitude transients of short duration, making it possible to estimate by eye the arrival time from the spectrogram. Also the uncertainty of the arrival time is estimated from the spectrogram. For the mid-frequency calls accurate arrival time estimation from the spectrogram is not possible due to the nature of the signal. Here a clip of the signal from the spectrogram of one hydrophone is used as a template to be matched with the spectrograms of the other hydrophones. Successively using the received signals for each of the hydrophones as the template, a set of travel time differences is obtained. Taking a weighted average provides the estimated travel time differences. Although this matched filtering gives

considerable pulse compression, the estimated uncertainty in the arrival times of the mid-frequency calls is much larger than that for the gunshots. Table I and Table II present the estimated arrival times and their corresponding uncertainties.

Table I: Estimated arrival times and standard errors (σ_t) [s] for the five hydrophones (L,C,E,H,J) and the five gunshots (S013-1, S035-2, S070-3, S093-4, S110-5).

	S013-1	S035-2	S070-3	S093-4	S110-5
L	15.09	15.17	15.27	14.78	15.20
C	14.20	21.64	9.20	21.02	9.42
E	16.50	13.38	18.88	19.98	18.95
H	21.80	15.06	22.12	14.22	22.03
J	20.70	21.61	16.31	16.24	16.70
σ_t	0.03	0.03	0.01	0.02	0.02

Table II: Estimated arrival times and standard errors (σ_t) [s] for the five hydrophones (L,C,E,H,J) and the two mid-frequency calls (S209-14, S210-15).

	S209-14	S210-15
L	0.00	0.00
C	-3.36	-3.38
E	-4.31	-4.29
H	5.33	5.20
J	5.94	5.92
σ_t	0.15	0.15

3. THE LOCALISATION METHOD

The following notation will be used:

- Lower case bold: column vector;
- Upper case bold: matrix;
- Lower/upper case italics: scalar

The unknown position of the source is denoted by $\mathbf{x} = (x_1, x_2, x_3)$, where the third co-ordinate indicates depth in the water. The positions of the hydrophones are denoted by $X_{i,j}$ (with $i = 1,2,3$ (co-ordinate index) and $j = 1,2,3,4$ (corresponding to the hydrophones denoted C,E,H,J, respectively)). The position of hydrophone L is taken as the origin, i.e., at (0,0,0). $\mathbf{x}$ should be solved from the following 4 equations ($j = 1,2,3,4$)

$$\sqrt{(x_1 - X_{1,j})^2 + (x_2 - X_{2,j})^2 + (x_3 - X_{3,j})^2}$$
$$- \sqrt{x_1^2 + x_2^2 + x_3^2} = y_j = R_j - R_0 = \overline{c}(t_j - t_0) \qquad (1)$$

Written shortly as

$$\mathbf{y} = F(\mathbf{x}) \qquad (2)$$

where

- $\mathbf{y}$ is the measurement vector containing the ranges R_j of the source to the jth hydrophone minus the range R_0 of the source to the central hydrophone L;
- $\bar{c}$ is the mean sound speed;
- t_j is the travel time of the sound from the source at $\mathbf{x}$ to the jth hydrophone.

3.1. Least squares solution

Determining three unknown position co-ordinates from four relative travel times gives an inconsistent set of equations. The best estimate for the unknown co-ordinates can be found by application of least-squares adjustment. Consider a linear relation between observations (containing the relative travel times) and unknowns (the position co-ordinates)

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e} \qquad (3)$$

with $\mathbf{A}$ the (m x n) design matrix, $\mathbf{y}$ the column vector of measurements (length m), $\mathbf{x}$ the column vector containing the parameters to be determined (length n) and $\mathbf{e}$ the column vector containing the measurement errors (length m). For the situation considered $m = 4$ and $n = 3$, i.e., $m > n$, an over-determined system.

The least squares solution to this problem is:

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{Q}_y^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Q}_y^{-1} \mathbf{y} \qquad (4)$$

with $\mathbf{Q}_y$ the covariance matrix of the measurement vector $\mathbf{y}$. The covariance matrix of the solution reads

$$\mathbf{Q}_{\hat{\mathbf{x}}} = (\mathbf{A}^T \mathbf{Q}_y^{-1} \mathbf{A})^{-1} \qquad (5)$$

which provides the precision of the solution, accounting for the uncertainties of the observations ($\mathbf{y}$) through $\mathbf{Q}_y$. The method of least-squares adjustment is based on minimizing the discrepancy between $\mathbf{y}$ and $\mathbf{A}\hat{\mathbf{x}}$.

3.2. Linearization of the problem

In our problem there is no linear relation between measurements $\mathbf{y}$ and unknowns $\mathbf{x}$. Therefore, the expression has to be linearized. A first order approximation of the nonlinear relation can be derived using a Taylor series approximation around $\mathbf{x}^{(0)}$

$$\mathbf{y} = F(\mathbf{x}^{(0)}) + \mathbf{J}\,\Delta\mathbf{x} \qquad (6)$$

with $\mathbf{J}$ the (m x n) Jacobian matrix

$$J_{j,i} = \frac{\partial F_j}{\partial x_i} = \frac{x_i - X_{i,j}}{\sqrt{(x_1 - X_{1,j})^2 + (x_2 - X_{2,j})^2 + (x_3 - X_{3,j})^2}}$$
$$- \frac{x_i}{\sqrt{x_1^2 + x_2^2 + x_3^2}} = \frac{x_i - X_{i,j}}{R_j} - \frac{x_i}{R_0}$$
$$(7)$$

evaluated at $\mathbf{x} = \mathbf{x}^{(0)}$. Further $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}^{(0)}$.

3.3. The solution to the linearized problem

With $\Delta\mathbf{y} = \mathbf{y} - F(\mathbf{x}^{(0)})$ we can write

$$\Delta\mathbf{y} = \mathbf{J}\,\Delta\mathbf{x} + \mathbf{e} \qquad (8)$$

To this equation we can apply the theory given in Section 3.1. The solution is

$$\Delta\hat{\mathbf{x}} = (\mathbf{J}^T \mathbf{Q}_y^{-1} \mathbf{J})^{-1} \mathbf{J}^T \mathbf{Q}_y^{-1} \Delta\mathbf{y} \qquad (9)$$

$\mathbf{Q}_y$ is still the covariance matrix of the measurement vector $\mathbf{y}$, since errors on $\mathbf{y}$ are equal to those on $\Delta\mathbf{y}$.

The solution for the source position $\mathbf{x}$ is given by

$$\hat{\mathbf{x}} = \mathbf{x}^{(0)} + \Delta\hat{\mathbf{x}} \qquad (10)$$

which is used in an iteration process, i.e., this solution $\hat{\mathbf{x}}$ is used as the initial value (here $\mathbf{x}^{(0)}$) for the next iteration

$$\hat{\mathbf{x}}^{(k+1)} = \hat{\mathbf{x}}^{(k)} + (\mathbf{J}^T \mathbf{Q}_y^{-1} \mathbf{J})^{-1} \mathbf{J}^T \mathbf{Q}_y^{-1} (\mathbf{y} - F(\hat{\mathbf{x}}^{(k)})) \qquad (11)$$

with the Jacobian $\mathbf{J}$ evaluated at $\mathbf{x} = \hat{\mathbf{x}}^{(k)}$.

The covariance matrix of the kth solution is

$$\mathbf{Q}_{\mathbf{x}^{(k)}} = (\mathbf{J}^T \mathbf{Q}_y^{-1} \mathbf{J})^{-1} \qquad (12)$$

The process ends once the difference between successive solutions is negligible. For this problem typically five iterations suffice.

3.4. Calculating $\mathbf{Q}_y$

The precision of the measurements $\mathbf{y}$ is contained in the covariance matrix $\mathbf{Q}_y$ which is determined as follows. Recall

that $y_j = R_j - R_0$, and denote the standard deviation of the measurement error on R_j ($j = 0,1,2,3,4$) as σ, i.e., $\overline{R_j^2} - \overline{R_j}^2 = \sigma^2$. Further, assume that the errors on R_j and R_k are uncorrelated, i.e.,

$$\overline{(R_j - \overline{R_j})(R_k - \overline{R_k})} = 0, \ j \neq k \tag{13}$$

Then it can easily be shown that the diagonal elements of $\mathbf{Q}_y$ are

$$\sigma_{y_j}{}^2 = \overline{y_j^2} - \overline{y}_j^2 = 2\sigma^2 \tag{14}$$

and the off-diagonal elements

$$\overline{(y_j - \overline{y}_j)(y_k - \overline{y}_k)} = \sigma^2 \tag{15}$$

Hence

$$\mathbf{Q}_y = \begin{pmatrix} 2\sigma^2 & \sigma^2 & \sigma^2 & \sigma^2 \\ \sigma^2 & 2\sigma^2 & \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 & 2\sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 & \sigma^2 & 2\sigma^2 \end{pmatrix} \tag{16}$$

3.5. Properties of $\mathbf{Q}_x$

The covariance matrix $\mathbf{Q}_x$ describes the precision of the obtained position $\hat{\mathbf{x}}$ of the sound source. Sometimes only the diagonal elements of $\mathbf{Q}_x$ are considered. These diagonal elements describe the variances of the unknown parameters. However, in this way a possible correlation between the unknown co-ordinates is not accounted for. Consequently presenting errors for each coordinate separately is insufficient. The correlation between the errors in the position coordinates is relevant information and should be presented as output too. Therefore, we use the so-called confidence region, which gives the area in which the estimated position is likely to be. This confidence region is calculated from the covariance matrix. In three dimensions the confidence region in an ellipsoid. In two dimensions the confidence region is an ellipse.

Now consider the two-dimensional (2D) situation: $\mathbf{x} = (x_1, x_2)$, i.e., sound source and all five receivers are assumed to be in the same horizontal plane. The two eigenvalues of $\mathbf{Q}_x$, denoted $\lambda_{\min}$ and $\lambda_{\max}$, determine the length of the semi-axes a and b of the error ellipse according to

$$a = \sqrt{\lambda_{\max}} \ \text{ and } \ b = \sqrt{\lambda_{\min}} \tag{17}$$

The orientation of the ellipse is given by the direction of the eigenvector corresponding to $\lambda_{\max}$. The ellipse is centered at the least-squares estimate $\hat{\mathbf{x}}$. The probability that the true position lies within the error ellipse is equal to 39 % (for the 2D situation). The 95% confidence region is obtained by multiplying the length of the semi-axes by 2.45 (assuming a Gaussian distribution).

Figure 1 illustrates for the given receiver configuration (using hydrophone L as reference) the error ellipses corresponding to a series of source positions, thereby demonstrating the localisation performance of the receiving network. For this simulation σ_t was taken as 0.1 s. Note that the size and orientation of the 95% error ellipse is determined by the receiver geometry and the source position, where source position inside the square show much smaller error ellipses than positions outside the square.
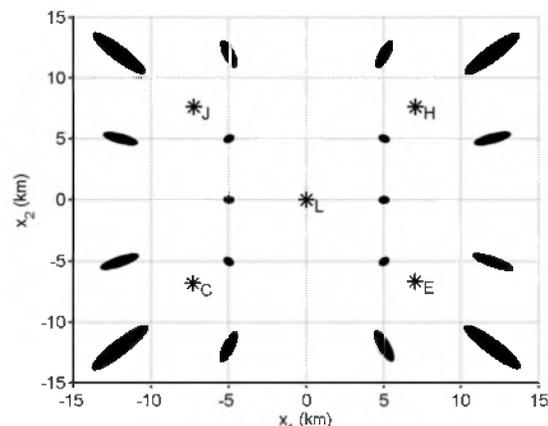
Figure 1: Simulation results: the centers of the error ellipses are the simulated source positions. The stars indicate the five hydrophone locations.

3.6. Estimating a starting solution

To assure convergence of the iterative least squares approach, an acceptable starting solution $\mathbf{x}^{(0)}$ is required. For this we assume the 2D situation with both the source and the hydrophones in the same horizontal plane.

Consider the geometry of the system of five hydrophones, forming a square with hydrophone L in its centre. Each perpendicular bisector of the line connecting two hydrophones, say J and L, defines the set of positions with equal distance to these two hydrophones, with the relative travel time to J with respect to L, t_{JL} ($= t_J - t_L$), equal to zero. All points with a positive value of t_{JL} will be at the L-side of the bisector, while a point with a negative t_{JL} will be at the J-side.

Pair-wise combination of the five hydrophones results in a subdivision of the horizontal plane in 24 sub-sectors. To prevent a too small subdivision, only 16 sub-sectors are defined as shown in Figure 2.
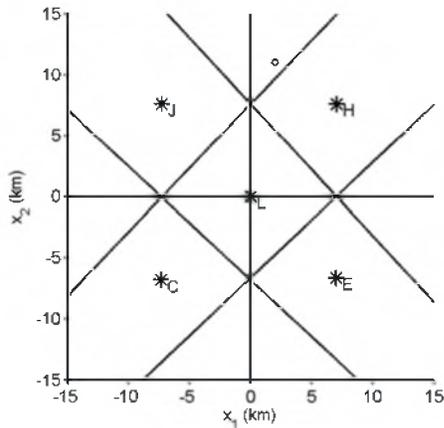
Figure 2: The 16 sub-sectors, divided by the six lines indicated. The hydrophone positions are denoted by stars, whereas the circle indicates an example initial position with $t_{CL} > t_{EL}$ and $t_{JL} < 0$.

Figure 3: The 7 source positions (.) as determined by the two-stage method. The positions of the hydrophones are indicated by a star (*).

Table III: x_1 and x_2 estimates for the 7 source positions.

file	x_1 [km]	x_2 [km]	longitude	latitude
S013-1	-1.88	-6.57	-66.428	44.603
S035-2	9.45	-0.66	-66.285	44.656
S070-3	-11.66	-6.63	-66.552	44.602
S093-4	1.40	6.46	-66.387	44.720
S110-5	-9.85	-5.72	-66.529	44.611
S209-14	2.87	-34.2	-66.368	44.354
S210-15	2.95	-34.2	-66.367	44.355

Based on the sign of the measured relative travel times, the appropriate sub sector can be selected. As an example, the combination of $t_{CL} > t_{EL}$ and $t_{JL} < 0$ gives a position as the one indicated by the 'o' in Figure 2.

After selection of the appropriate sub sector, hyperbolic fixing is applied (Spiesberger 2001) yielding a position for each pair of relative travel times. With respect to L there are four relative travel times. With six different pair wise combinations of them and each combination leading to at most two solutions, 12 solutions result. Some of these solutions can easily be removed; the complex ones and solutions outside the selected sub sector. In this way about eight potential solutions remain. Next, ambiguous positions that are relatively far away from the main cluster of positions are removed. Taking the average of the co-ordinates of the remaining positions proves to give a good starting solution for the iterative least squares approach, wherein optimal use is made of all available travel times.

4. RESULTS

The single path assumption and the given shallow water geometry cause the relative arrival times to depend little on the source depth x_3, prohibiting determination of x_3. Henceforth we assume x_3 fixed, reducing the number of unknowns from three to two. We choose $x_3 = 0$, i.e. the sound source is at the depth of hydrophone L, but another choice would have been equally good. As a second step all hydrophones are assumed to be at the depth of hydrophone L. This assumption has proven to have a negligible effect on the estimated source position. As a consequence, the problem will be assumed 2D, i.e., $x_3 = X_{3,j} = 0$, in the remainder of this paper.

The positions (x_1, x_2) corresponding to the five gunshots and the two mid-frequency calls, found with the procedure described in Section 3, are shown in Figure 3 above. The two mid-frequency calls are seen to almost coincide. Table III lists the x_1- and x_2-positions.

These estimates for the source position should not be used without an assessment of their accuracy. For example, if the uncertainty in the gunshot positions is of order several kilometers, the sound of gunshots S070-3 and S110-5 might just as well have been transmitted at the same position.

There are several contributions to the source position uncertainty (Wahlberg 2001). They stem, among other things, from uncertainty in arrival times, sound speed and hydrophone positions. These are subsequently discussed in the following sections.

4.1. Effect of uncertainty in arrival time

In Table I and II the uncertainty on the arrival times σ_t is presented. Using mean sound speed values as derived from the measured sound speed profiles, resulting uncertainties in the measurement vector $\mathbf{y}$ ($y_j = \bar{c}(t_j - t_0)$) can be estimated and the 95% confidence regions (or error ellipses), i.e., the area in which the true position is likely to fall, are calculated. The figures below show the error ellipses corresponding to the solutions presented in Table III. Figure

4 shows the results for the gunshots, whereas in Figure 5 the results for the mid-frequency calls are presented, for which the uncertainty in the arrival times is much larger than for the gunshots.
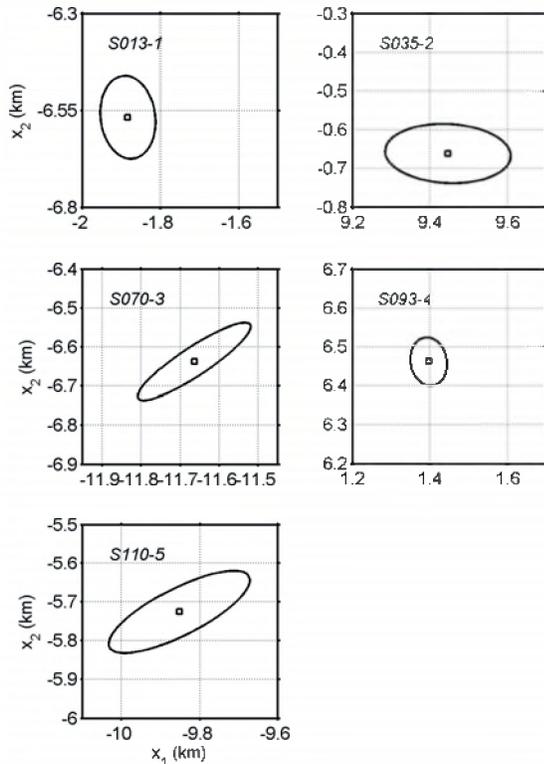
Figure 4: Error ellipses, accounting for the inaccuracies in arrival times for the gunshots. The source position estimates are indicated by squares.
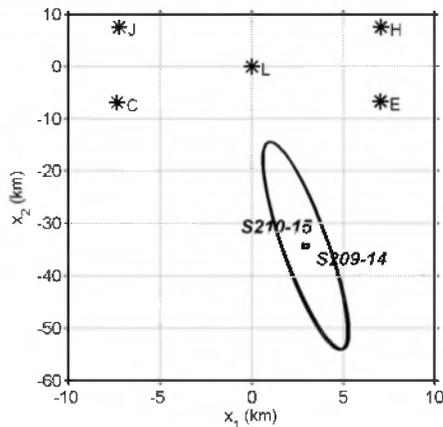
Figure 5: Error ellipses, accounting for the inaccuracies in arrival times for the mid-frequency calls. The hydrophone positions are indicated by stars. The source position estimates are indicated by squares. The positions of the two mid-frequency calls almost coincide.
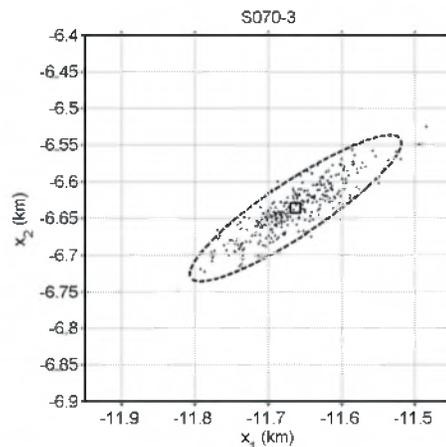
Figure 6: Results of the simulations for the gunshot S070-3. The dashed line is the theoretical error ellipse.

An alternative way for obtaining the confidence interval around the estimated source position is by means of Monte Carlo simulation. These simulations can be used for comparison with the calculated error ellipses and are also used in Section 4.3 for assessing the effects of receiver position uncertainty. The four arrival time differences are selected randomly from distributions that are assumed to be Gaussian, with means and standard deviations as given in Tables I and II. The starting position is, as previously, obtained through hyperbolic fixing. Figure 6 shows the results for gunshot S070-3. Almost all simulation results (dots) lay within the 95% confidence region as calculated according to Section 3.

4.2. Effect of uncertainty in sound speed

The effect of the uncertainty in the sound speed on the localisation accuracy is investigated as follows.

The sound speed profiles roughly show an upper layer of approximately 40 m with a sound speed of about 1499 m/s. Below the thermocline the sound speed drops to about 1489 m/s. For the calculations of Section 4.1 a typical sound speed of 1492 m/s has been used.

Here calculations are carried out for the highest and the lowest sound speed encountered, i.e., 1489 m/s and 1499 m/s. In Table IV the deviations of the thus estimated source position relative to those listed in Table III are presented. Employing these two extremes in sound speed is seen to result in deviations that are much smaller than the size of the 95% error ellipses given in Section 4.1. Hence the uncertainty in arrival times is dominating over uncertainty in sound speed.

Table IV: Shifts in estimated source position when using values for the mean sound speed of 1489 m/s and 1499 m/s, relative to the source position that is estimated using a typical sound speed of 1492 m/s.

file	1489 m/s Δx_1 [m]	1489 m/s Δx_2 [m]	1499 m/s Δx_1 [m]	1499 m/s Δx_2 [m]
S013-1	2.5	11	-5.9	-26
S035-2	-37	5.5	38	-5.6
S070-3	43	40	-95	-89
S093-4	-2.4	-8.2	5.6	19
S110-5	34	29	-79	-69
S209-14	-50	516	119	-1234
S210-15	-56	551	133	-1324

4.3. Effect of uncertainty in receiver position

Figure 7 shows the results of Monte Carlo simulations that account for the uncertainty in the receiver positions. Now the hydrophone positions $X_{i,j}$ are selected randomly from Gaussian distributions with standard deviations as given in Table V.
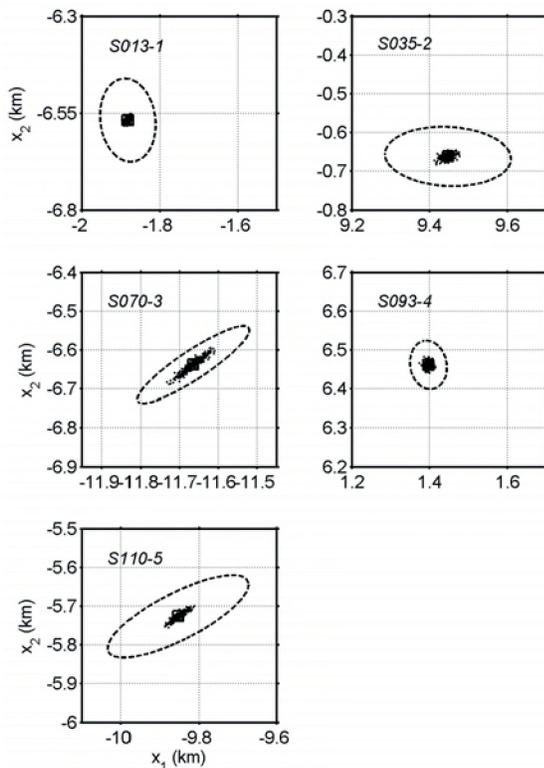
Figure 7: Results of simulations accounting for the error in the receiver positions for the five gunshots. The theoretical error ellipse, indicating the uncertainty due to uncertainty in arrival times, is plotted as a dashed line for comparison.
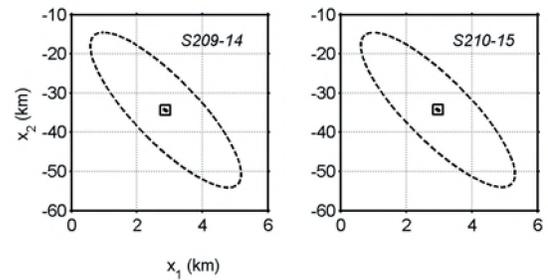
Figure 8: Results of simulations accounting for the error in the receiver positions for the two mid-frequency calls. The theoretical error ellipse, indicating the uncertainty due to uncertainty in arrival times, is plotted as a dashed line for comparison.

For comparison we have included the 95% error ellipse due to arrival time uncertainty. It is clear that the effect of hydrophone position uncertainty is small compared to the effect of the uncertainty in the arrival times.

Table V: Hydrophone position uncertainty.

Hydrophone	σ_{X_1} [m]	σ_{X_2} [m]
C	2.15	6.06
E	5.13	4.16
L	3.14	2.08
H	12.55	11.47
J	0.42	9.72

5. SUMMARY AND CONCLUSIONS

In this paper an accurate method for localizing the sound made by marine mammals is described. The method is applied to experimental data collected in the Bay of Fundy, Canada. The receiving system consists of five hydrophones, moored on the bottom. The five hydrophones form a square of 14 by 14 km with one hydrophone in the middle. Both gunshot type of signals and mid-frequency calls have been considered.

The method uses travel times of the received signals. Since the moment at which the sound was emitted is unknown, only relative travel times could be used. The middle hydrophone is taken as reference hydrophone.

The localisation procedure assumes straight path propagation in an unbounded medium and consists of two steps. The first step provides through hyperbolic fixing a first estimate for the source position, which is used as input for the second step. In this second step the solution is improved in an iterative process, where each iteration determines the least-squares solution of the set of the four linearized equations for the measured relative travel times. This so-called adjustment theory combines redundant uncertain measurements in an optimal way, by weighting the

observations with a measure of confidence and by determining the least squares solution. The method also gives the precision of the estimated source positions by presenting the 95% confidence region giving the area in which the estimated position is likely to be.

The gunshot signals have been accurately localized, whereas the mid-frequency calls show a much larger position uncertainty due to the larger uncertainty in relative travel times and the larger distance to the hydrophone array. The precision of the gunshot positions, i.e., the size of the error ellipse, is of order 200 m, whereas it amounts to several tens of kilometers for the mid-frequency calls. It should be emphasized that the correlation between the errors in the position co-ordinates is important and is therefore presented too by the orientation of the error ellipse.

An important item addressed in this paper is an assessment of the position accuracy due to uncertainties in arrival times, sound speed and receiver position. It is found that for the experiment considered, the uncertainty in the arrival times is dominant.

LITERATURE

Spiesberger, J.L. (2001). "Hyperbolic location errors due to insufficient numbers of receivers, " J. Acoust. Soc. Am. 109, 3076-3079

Wahlberg, M., Mohl, B. and Teglberg Madsen, P. (2001). "Estimating source position accuracy of a large-aperture hydrophone array for bioacoustics, " J. Acoust. Soc. Am. 109, 397-406

COMPARING A LINEAR WITH A NON-LINEAR METHOD FOR ACOUSTIC LOCALIZATION

Magnus Wahlberg

Department of Zoophysiology, Aarhus University, C. F. Moellers Alle Building 131, DK-8000 Aarhus C, Denmark

ABSTRACT

The performance of two different acoustic localization techniques is evaluated with signals from right whales in the Bay of Fundy. The methods are compared to the GPS localization error (114-273 m, N=3) through the use of played back whale calls. The linear approach underestimates the source location error (22 m, N=3), whereas the non-linear approach exaggerates the error (462-1166 m, N=3). The linear approach may render unrealistic error bounds because of the inherent non-linear properties of the localization problem. The non-linear approach may exaggerate error bounds by choosing the wrong cross-correlation peak for the time-of-arrival difference measurements. Whereas the GPS localization error was always contained within the non-linear error bounds it was never contained within the linear error localization bounds. This indicates that the non-linear approach can give more realistic error estimates, especially in situations where the sound path geometry is unknown. [Work supported by the Office of Naval Research and the Oticon Foundation.]

SOMMAIRE

La performance de deux méthodes différentes de localisation acoustique est évaluée à partir de la localisation acoustique des baleines franches dans la Baie de Fundy. Les méthodes sont comparées à l'erreur de localisation GPS (114-273 m) à partir de vocalisations de baleines franches préenregistrées. L'approche linéaire sous-estime l'erreur de localisation de la source sonore (22 m), alors que l'approche non-linéaire surestime l'erreur (462-1166 m). L'approche linéaire rend irréaliste la marge d'erreur possible à cause des propriétés non-linéaires du problème de localisation. L'approche non linéaire exagère la marge d'erreur, ce qui est expliqué par le choix du mauvais maximum de corrélation croisée des mesures de différences de temps d'arrivée. Toutefois, l'erreur de localisation GPS était toujours contenue à l'intérieur d'une marge d'erreur non-linéaire et n'était jamais contenue à l'intérieur d'une marge d'erreur linéaire de localisation. Ceci indique que l'approche non-linéaire peut donner des erreurs d'estimation plus justes, spécifiquement dans les situations où la trajectoire du son est inconnue. [Travail supporté par l'Office of Naval Research et la Oticon Foundation.]

1. INTRODUCTION

In bioacoustics it is often relevant to determine the location of a calling animal. In studies ranging from acoustic census of animal populations to behavioural studies the knowledge of animal location greatly extends the types of problems that can be addressed and broadens the analytical techniques that can be applied.

Acoustic localization is performed using a receiver array to locate a vociferous animal by measuring the arrival times of corresponding signals at different receivers (Wahlberg *et al.* 2001; Spiesberger and Fristrup 1990). Methods for acoustic localization are currently in rapid development, in terms of both recording and analysis techniques (see Møhl *et al.* 2001 and the papers in the present volume of this journal).

How many receivers are needed for specific localization tasks, assuming all receivers to have omni-directional receiving characteristics? We may view acoustic localization as a mathematical transformation from measured time-of-arrival differences (TOADs) to the source coordinates. With an array of N receivers one can measure $N-1$ independent TOADs. (The word *independent* is here used with the meaning of two variables not being 100% correlated. The $N-1$ TOADs mentioned in the text which are not linear combinations of one another, are not completely uncorrelated with each other. E.g. the TOAD between the signal at receiver 1 and 2 and the TOAD between receiver 1 and 3 both contain a measurement of the time-of-arrival at receiver 1. See Wahlberg et al. (2001) for details.) By convenience the TOADs are measured between each receiver and one reference receiver, denoted receiver 1, defined as being in the origin of the coordinate system. Any other definable TOAD from an N-receiver array could be expressed as a linear combination of the other TOADs: for example, the TOAD measured between receiver 2 and 3 is the same as the TOAD between receiver 3 and 1 minus the TOAD between receiver 2 and 1.

From mathematical analysis we know that $N-1$ TOADs may be transformed into a maximum of $N-1$ source coordinates. To track an animal in two dimensions

two source coordinates are required. To achieve this one needs to measure at least two independent TOADs, and the array must consist of at least three receivers. For three-dimensional tracking requiring three coordinates a minimal number of four receivers is needed. Following the notation of Wahlberg *et al.* (2001), we call such an array a *Minimum receiver number array* (MINNA). Arrays with more receivers than MINNAs are called ODAs (*over-determined arrays*; Wahlberg *et al.* 2001). For ODAs more TOADs are available than minimally required for calculating the source coordinates. In this case the source coordinates can be calculated through some kind of averaging technique, such as least squares.

The number of TOADs necessary to solve a certain localization task may also be determined from geometric considerations. Each TOAD restricts the source location either to a hyperbolic curve (in a 2-D source-array geometry) or to a hyperboloid surface (in 3-D). Two intersecting curves (corresponding to two TOADs) are sufficient for localizing animals in 2-D (Fig. 1). In 3-D, three intersecting surfaces are needed to localize the animal.
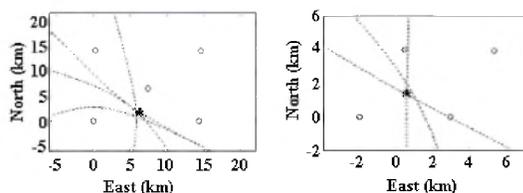
Figure 1. Hyperbola plot of (left) a right whale vocalization recorded in 2002 (file nr. S093-9) and (right) a played back right whale call recorded in 2000 (file no. S282). The circles indicate receiver locations. The least-squares estimate of the source location is denoted with an asterisk ('*').

The simplest acoustic location equations usable for MINNAs are quadratic (Wahlberg *et al.* 2001), potentially rendering two source solutions instead of one for each set of TOADs. Geometrically this corresponds to two hyperbolas or three hyperboloids intersecting in two rather than only in one point. In such cases an extra receiver is needed to resolve the source location ambiguity (Spiesberger 2001). The introduction of an extra receiver renders an extra TOAD and therefore an extra hyperbola/ hyperboloid. This array is now an ODA.

Assuming that the sound speed of the medium is constant, the source location can be derived from MINNAs using analytical equations. These algorithms are invertible, except for locations where source locations are ambiguous. A problem with MINNA arrays is that there is no implicit information available on the accuracy of source coordinate estimates. The investigator has to rely upon error propagation analysis to evaluate the magnitude of localization errors (Wahlberg *et al.* 2001). For ODA systems, the redundant TOADs may be used to either assess the error in source location through regression techniques, or through an analysis of error propagation (Spiesberger and Fristrup 1990; Wahlberg *et al.* 2001). The transformation from TOADs to source locations is in

general not invertible for ODA systems, as the calculation involves some type of data smoothing.

Both for MINNA and ODA analysis, the simplest form of an error analysis is to linearize the location problem and its error components (Spiesberger and Fristrup 1990; Wahlberg *et al.* 2001). Linearization has the advantages of yielding fast computations with well-defined procedures for the error estimation. The magnitude of various error sources, such as variations in the sound speed, inaccuracies in TOAD measurements, and drifting receiver locations, can be studied and modelled separately. Also, one may rapidly evaluate how any covariance between the input variables affects derived source locations (Wahlberg *et al.* 2001).

However, linearizing the localization equations introduces several problems. If the hyperbolas are not crossing, there will be no source location (The source location coordinates will in this case be a complex number) in a MINNA system, even if the shape of the hyperbolas indicates that the source ought to be restricted to a certain area. Even if a source location is obtained, there is no possibility of verifying that the input TOADs were measured to a stated accuracy. These problems are alleviated through the introduction of another receiver. However, it should be recalled that the localization equations of both MINNAs and ODAs are inherently non-linear, and the degree of non-linearity is spatially variant within and around the array. Therefore, linearization may work acceptably in some cases but not in others (Spiesberger and Wahlberg 2002).

Such problems call for the development of non-linear localization methods. Spiesberger and Wahlberg (2002) developed a non-linear form of acoustic location error analysis based on computer simulations of permuted subsets of MINNA receiver constellations (Fig. 2). Using synthesized data, this numerical form of error analysis seemed to give more realistic error estimates than linear analysis. The authors noted the need for application to real data before the method's performance could be fully evaluated (Spiesberger and Wahlberg 2002).

In November 2003 a workshop was organized in Halifax on passive acoustic localization of marine mammals (Anon. 2003). Before the workshop, the organizers supplied the participants with right whale *(Eubalena glacialis)* recordings as a training dataset for investigating alternative localization and detection routines. The dataset also included playback recordings, where signals were broadcasted from known locations. The dataset provided an opportunity for comparing the performance of the linear and the non-linear localization methods outlined above.

2. METHODS

2.1 Data material

Sound recordings were obtained from the organizers of *the Workshop on detection and localization of marine mammals using passive acoustics, 19. – 21. Nov. 2003* (Anon. 2003). Data was collected with 5 Ocean Bottom Hydrophones (OBHs, *Defense Research & Development*

Canada, Halifax) moored in the Bay of Fundy in September 2002 in an area where foraging right whales (*Eubalaena glacialis*) are regularly observed during summertime. Recordings were made with a sampling rate of 1200 Hz. The dataset is described in detail in Anon. (2003). The bottom depth varied between 123-210 m at the site of the hydrophones. The locations of the receivers were determined both by GPS and by recording playback signals at known locations. The total error in receiver coordinates was acoustically assessed to vary between 4 and 18 m. TOAD measurements were prone to errors arising from differential clock drift in the OBH recording units (Anon. 2003). The clock drift was measured both before and after the fieldwork and varied between 65 and 174 µs per hour. Signal time-of-arrivals at each OBH recording were compensated assuming that the clock drift was linear throughout the recording period. The workshop organizers supplied 16 sound files containing right whale vocalizations, which had been classified as being either 'low-frequency', 'mid-frequency', or 'gunshot' calls (Anon. 2003). In addition, sound speed data was derived from 6 conductivity-temperature-depth (CTD) profiles measured closely in time to the sound recordings.

From data gathered during a previous array deployment in August 2000 in the same area the performance of the location system was assessed. Right whale sounds were transmitted from a small boat at a known location (Anon. 2003). This data was used to investigate the precision of acoustic location and anticipated error estimates. During these recordings, the array consisted of four (instead of five) OBH's, moored at 131-190 m depth, the source being placed at 20 m depth. The workshop organizers supplied four sound files from the playback sessions. In addition, data was made available from 3 CTD profiles obtained in the area at the time of the recordings. There was no acoustic calibration of the receiver locations during the 2000 recordings. The field recordings were made during such a short time interval that compensation for the buoy clock drift (see above) was considered unnecessary (Anon. 2003).

2.2 Analysis

Data were extracted with sound-analysis software (*Cool Edit*, Syntrillium), and measurements were made with scripts written in *Matlab 6.5* (Mathworks, Inc.). TOADs were measured by cross-correlating signals recorded at different receivers. The TOAD measurements included compensation for the buoy clock drift in the right whale recordings (see above), assuming the clock drift rate to be constant during the recordings. The TOADs, the sound speed, and the receiver locations were used to calculate the location of the source as well as the associated error (*sensu* Wahlberg *et al.* 2001 and Spiesberger and Wahlberg 2002). A linear and a non-linear method method were compared in the localization process.

The vertical aperture of the array (the differences in bottom depths between the receivers) was much smaller than (less than 1%) the horizontal distance between the receivers. Thus, the array is situated approximately in the horizontal plane. The water depth was considered

insignificant (about 1-2 %) compared to the horizontal extent of the array. It was not expected to be possible to locate sound sources in the vertical plane with a resolution better than the depth of the water column. Therefore acoustic location was made with 2-D versions of the algorithms presented in Wahlberg *et al.* (2001) and Spiesberger and Wahlberg (2002). These algorithms assume both the source and the receiver array being situated in the same horizontal plane.

In the *linear error estimation* approach, the source location was assessed through the least squares technique described by Spiesberger and Fristrup (1991) and Wahlberg *et al.* (2001). A homogenous sound speed was assumed. Error estimates of the source location coordinates were achieved using a linear error propagation model (LEP, Wahlberg *et al.* 2001) applied to the input variables. As input to the calculations, the input variables and their errors were required, as well as the covariances between the variables (Wahlberg *et al.* 2001).

The *non-linear error estimation* approach is described in detail in Spiesberger and Wahlberg (2002) and is depicted in Fig. 2. The array is divided into a number of MINNA subunits, each containing 3 or 4 receivers (for 2-D and 3-D positioning, respectively). Each subunit has a set of input variables: the receiver locations, the TOADs, the sound speed and error estimates. The corresponding source location is calculated through the MINNA localization formula given in Wahlberg *et al.* (2001). If the sound speed varies between the source and the different receivers, a set of quadratic equations, called *isodiachron equations* (Spiesberger and Wahlberg 2002) can be used alternatively. Therefore the non-linear analysis is not restricted to the assumption of a homogenous sound speed. The input variables and their estimated errors are used to randomly shift the sound speed, receiver locations, and TOADs of the MINNA system. For each shift in input variables a new source position is derived. We assessed 1000 locations for each MINNA sub-array. This generates a cloud of possible source locations. For any TOAD that generates a doublet location an additional receiver is chosen to solve the ambiguity. The procedure is repeated for all the MINNA sub-array constellations, or (if there are too many constellations) for a Monte Carlo Subset of these constellations. Each constellation generates a cloud of possible source locations. The location of the source is defined as being the surface (in 2-D) or volume (in 3-D) where all the generated clouds intersect. This intersection is the only possible region in space where the source must be situated, provided the input variables are given with adequate error intervals and the sound speed is constant between the source and each receiver (see Spiesberger and Wahlberg 2002 for details).

2.3 Error assessment

The error analysis of both the linear and non-linear approach demands proper assessment of the accuracy of all input variables. The input variables are the sound speed, the TOADs, and the receiver coordinates. The error

assessment is considered in some detail for the data used for the workshop (Anon. 2003).
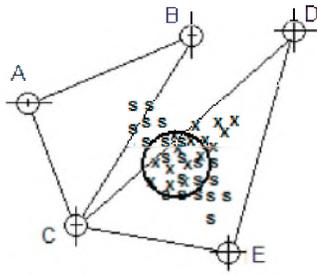
Figure 2. The principle of the non-linear method for acoustic localization, adapted after Spiesberger and Wahlberg (2002). Circles are receiver locations with error bars. Two subset MINNA receiver constellations are shown, ABC and CDE. The source locations derived from varying the errors in the input variables of the ABC subset are denoted with 's', and the corresponding source locations from the CDE subset are denoted with 'x'. The source is defined as being within the region enclosed by the source solution of the two subsets, marked with a circle.

Sound speed

The linear analysis assumes that the medium has a constant sound speed. The non-linear approach allows the sound speed to vary with the direction of source to receiver, but not with range. In the real ocean sound speed normally varies both horizontally and vertically, the most pronounced gradients usually being vertical. Propagating sound waves are refracted in a sound speed gradient according to Snell's law (Urick 1983). The acoustic path from the source to the receiver may pass through a range of sound speeds. Consequent ray bending can be studied using ray-tracing. If detailed algorithms are not applied, Spiesberger and Fristrup (1990) have derived an alternative approximation to quantify the effect of ray bending on TOAD measurements.

In the present study, sound speed profiles were averaged both spatially (vertically) and temporally, and the standard deviation calculated. The mean and the magnitude of one standard deviation were used as inputs to the error assessment of source localization. In 2002 the sound speed was 1494 ± 3 m/s; in 2000 it was 1494 ± 6 m/s.

Time-of Arrival Differences (TOADs)

There are many available methods for measuring TOADs for corresponding signals recorded on different receiver channels. The most widely used technique is cross-correlation. In general cross-correlation performs well if the signal has a large time-bandwidth product (Spiesberger and Fristrup 1990). The width of the peak of the cross-correlation envelope function is given by Spiesberger and Fristrup (1990):

$$\delta t \approx 1 / [2\pi W_{rms} d]$$

δt is defined as the time-of-arrival measurement inaccuracy, W_{rms} is the rms bandwidth of the signal, while d is the linear signal-to-noise ratio of the cross correlator. The cross correlator signal-to-noise ratio can be derived from the signal-to-noise ratio of the recordings (defined as the rms intensity of the right whale signals divided by the rms intensity of the noise in the frequency band of the signal), which was measured to be 13-22 dB, and from the number of samples in the digitized signal (see Spiesberger and Fristrup 1990). The number of samples in each right whale signal is about 1500-2000. The TOAD of right whale calls (having a rms bandwidth of 9-10 Hz) is estimated to be measured with an accuracy δt of about 4 to 29 μs. A value of 30 μs was used in the error assessment of acoustic localization presented below. The sampling frequency of 1200 Hz corresponds to a sample time resolution of 833 μs, which is larger than the 30 μs time resolution. Therefore, the cross correlation function was interpolated ten times to resolve arrival times at a scale dictated by the calculated timing accuracy. A typical cross correlation from the playback localizations is shown in Fig. 3. There is one well-defined peak of the cross correlation function, but also there is a whole series of peaks, probably caused by multiple paths from the source to the receiver. As will be discussed below, the precision of 30 μs is only valid if we assume that the correct cross-correlation peak has actually been measured. (An even better TOAD resolution may be obtained from using the peak of the cross correlation function, rather than the peak of the envelope of the cross correlation function. However, for the present cross correlation signals it was often difficult to assess which peak in the cross-correlation function should be chosen, whereas the envelope function usually rendered an unambiguous peak (c.f. Fig. 3).]

Time (s)
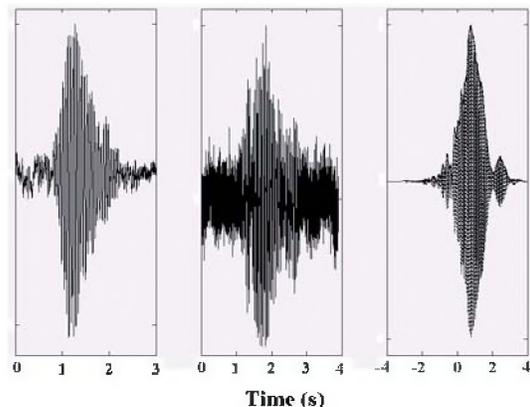
Figure 3. The cross correlation of a call recorded on two of the receivers of the array during 2003. Left: 'Gunshot call' recorded on the buoy C, Middle: the same 'gunshot call' recorded on buoy H. Right: The cross-correlation (stipled) and its envelope (solid line) of the two signals in (a) and (b).

Receiver coordinate errors

Receiver coordinate errors were assessed using the pinger recordings in 2002. The estimated error in the north-south direction was 2-12 m, and in the east-west direction it was 0.5-13 m (Anon. 2003). For the playback

recordings from 2000 no measurements of receiver coordinate errors were available. For the analysis of the playback signals it was assumed that the north and east mean values in receiver location errors were equal in 2002 and 2000. This assumption may be too liberal but was used due to the lack of better data.

3. RESULTS

3.1 Error maps

Once the magnitude of the errors in the input variables have been defined, the LEP model may be used to derive error contours for the array. If it is assumed that the errors in sound speed, TOADs and receiver coordinates are uncorrelated, then the error maps can be split into the contributions from each error source (Wahlberg *et al.* 2001). This procedure is useful for evaluating the localization precision of various source-to-array geometries, and also to pinpoint which input variable errors have the largest effect upon the localization error. Fig. 4 shows an example of error maps so derived. The source location error seems mainly to be caused by errors in sound speed and receiver locations, rather than in the precision of the TOAD measurements.
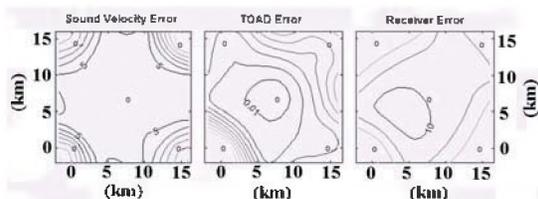
Figure 4. Error maps of the array used in 2002, calculated with a linear error propagation model (see text for details). Receiver locations are indicated with circles. The effect on the localization error is shown separately for the sound speed error (left, 5 m error contours, dc= 3 m/s), the TOAD errors (middle, 1 cm error contours, dt=30 μs), and errors in receiver coordinates (right, 10 m error contours, receiver errors from Anon. 2003).

3.2 Choices of error distributions and error estimates for source coordinates

The assessed variables and their errors are fed into the linear and non-linear acoustic localization analyses. For the linear approach we may choose between studying the residual error of the least-squares fitting, or to use the linear error propagation model to assess the magnitude of the errors. With all other quantities held constant these two error calculations should render comparable results. If not the case, it is usually a sign of a problematic localization task, e.g. that one of the TOADs has been erroneously interpreted and measured.

For the non-linear approach it is necessary to define the shape of the error distribution about the variable's mean value. When performing error analysis it is usually assumed that the errors are normally distributed and can be modelled as measured in terms of standard deviations or standard errors. However, the normal distribution may

not be the best way to define all error limits. For example, the tails of the normal distribution does not fall to exactly zero. Therefore, if one assumes normally distributed receiver locations there is always a small chance that the receiver is at an arbitrarily large range from the other receivers, which for physical reasons cannot be true. A better approach is to use truncated normal distributions, uniform distributions or other distributions with well-defined limits. In the work presented here we choose a uniform distribution for the nonlinear analysis. This renders comparable results for the linear model if we assume that the standard deviations of the linear model represents the range of a uniform distribution rather than a normal one. It is believed that the discrepancy between the two error distributions is minor and does not significantly influence the comparison between the linear and non-linear error estimations.

3.3 Acoustic localization of playback signals

Three out of the four playback files from 2000 contained events where the sound source could be located. The dropped file (S-289) was apparently dominated by source multi-path making definite cross correlations impossible.

In Table 1 the discrepancies between logged position of the play back vessel and the acoustic location of the signal are compared to the linear and non-linear source location error estimates. Fig. 1 (right) shows a sample hyperbola plot from the acoustic localization of a play back signal.

Table 1. The error in acoustic localization of play back signals ('ODA-GPS') compared to the localization error assessed with the linear ('ODA-LEP') and the non-linear techniques (see text). The linear and non-linear errors are given in meters, and in % relative the ODA-GPS error.

Seq.	Difference ODA - GPS	ODA LEP	Non-linear Error
S-282A	114 m	22 m (19%)	1166 (1023%)
S-282B	273 m	22 m (8%)	462 (169%)
S-288	174 m	22 m(13%)	1151 (661%)

3.4 Acoustic localization of right whale calls

Four out of 16 files from 2002 contained right whale signals which yielded cross-correlation functions usable for sound localization. The remaining files were problematic, either due to signal overload (the signals were digitally clipped in 4 of the files), interference between calls from several whales (in 4 files), or poor signal-to-noise ratio (in the remaining 4 files). An example of a successful acoustic location is depicted in Fig. 5. The mean and 1 s.d. error estimates from the linear error propagation model are depicted as a black cross. The corresponding nonlinear error limits are depicted as a rectangle. In Fig 5 (top) a 30 μs TOAD error estimate is used.

The linear error in Fig. 5 (top) is so small so that the error cross cannot be observed on th emap. To evaluate the risk of choosing the wrong cross-correlation peak, in Fig. 5 (bottom) the TOAD error is increased to 1 second which

approximates the maximum distance between the peaks in the cross-correlation envelope function.

4. DISCUSSION

This first trial compares the performance of the linear and non-linear error methods from Wahlberg *et al.* (2001) and Spiesberger and Wahlberg (2002) using real data. The linear error model underestimates errors whereas the non-linear model over-estimates them (Table 1). The GPS localization error (defined as the difference between the GPS position - with a 10 m error of its own – and the ODA acoustic location) is never contained within the linear error estimates, whereas it is always contained within the non-linear error estimates (Table 1). One may therefore claim that the non-linear approach renders the most realistic error estimates.
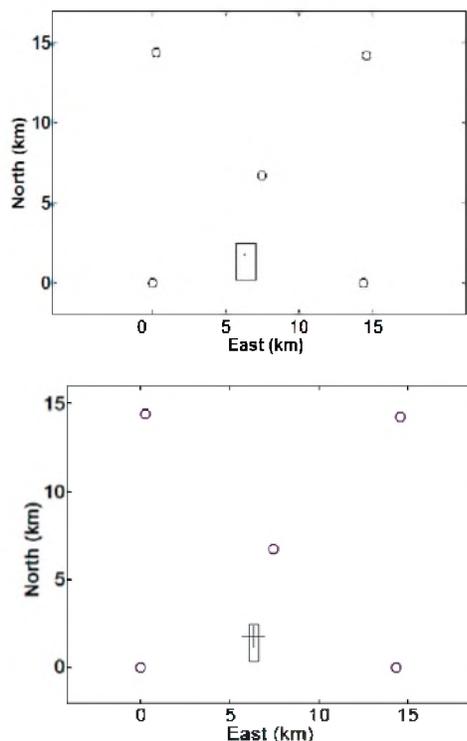
Figure 5. Acoustic localization of right whale number S093-09 (from Anon. 2003). The localization error estimated from linear error propagation is shown as a cross (in the top figure the cross is so small that it looks like a dot). The error estimated from the non-linear approach is depicted as a rectangle. The TOAD error is set to 30 µs in top, and to 1 s in bottom. See text for details.

A problem with the linear model is that the both the localization and its associated error analysis is achieved with linearizing techniques, despite the fact that the acoustic localization problem is non-linear in nature. This problem compounds by additional unrealistic approximations: the sound speed is set constant, and one assumes that the correct cross-correlation peak is chosen in the presence of multi-path. The huge discrepancy between the linear and non-linear error estimates (Fig. 5 top) assuming a TOAD uncertainty of 30 µs is alleviated

by increasing the TOAD uncertainty to 1 s (Fig. 5 bottom). This latter TOAD error better approximates reality considering that one may select the incorrect cross-correlation peak when measuring the time-of-arrival differences.

The non-linear approach yields more realistic error estimates as it presents the true range of possible source locations, given correct input data. This approach, while not requiring any linear assumptions, still assumes that the sound speed is constant (even though the algorithm may be modified to contain cases where the sound speed is variable between different source – receiver paths, see Spiesberger and Wahlberg 2002). Although the true source location error is always found within the non-linear error estimates (Table 1) the non-linear errors frequently appear to be almost an order of magnitude too large. The reason for this is not clear. It may indicate that the model is wrong: we may not have chosen the right cross-correlation peak, the sound speed profile causes ray bending so that we are not detecting the direct path but surface and bottom reflections, etc. All these reasons should have affected the derived source location. However it is not clear why they only affect the non-linear error estimation.

The non-linear approach has the advantage of treating each array as a constellation of several MINNA arrays. Each MINNA location is a reversible transformation from the input variables to the source coordinates (with the removal of ambiguous locations using an additional receiver; see above). Therefore, variations in the input variables within the assessed error limits are directly transformed into variations in the source coordinates that are reversibly related to the original input variables. In other words, for each MINNA system, the source location has to be exactly where it is calculated to be. Therefore the cloud of locations derived from each MINNA system directly reflects the only possible limits of the source location coordinates. Furthermore, the input variables may have different error distributions, and such effects can be directly observed upon the shape of the location cloud. When combining all the MINNA subsystems comprising the array, and always assuming that the input variables are accurately describing the real recording situation, the source location must lie inside the space defined by the intersection of all the location clouds.

Therefore, while the linear approach gives a possibly faster approximate source location, the non-linear approach inherently generates error bounds that reflect the only possible location of the source, given that the input variables and their error ranges are realistic.

The problem of computing time for numerical models has diminished with faster computers. The non-linear calculations made here can be accomplished on a standard laptop within a few seconds. For larger array systems, longer calculation times are expected, so the feasibility of the non-linear method decreases - especially for online applications.

Acoustic localization is a non-linear acoustic problem that can be solved either through linearization, or through non-linear techniques. The inherently non-linear nature of the localization problem suggests that only the non-linear

approach can be used for more sophisticated future models. Such non-linear techniques could also be a gateway towards inverse acoustic localization methods, such as matched-field and inverse processing (Thode *et al.* 2000; Spiesberger 1999).

ACKNOWLEDGEMENTS

REFERENCES

Anon. 2003. Handout describing the dataset for *the Workshop on detection and localization of marine mammals using passive acoustics.* Halifax, Nova Scotia, 19th-21st of November,. Document available at *www.atlantic.drdc rddc.gc.ca.*

Møhl, B., Wahlberg, M., and Heerfordt, A. 2001. A large-aperture array of nonlinked receivers for acoustic positioning of biological sound sources. Journal of the Acoustical Society of America 109(1), 434-437.

Spiesberger, J. and Wahlberg, M. 2002. Probability density functions for hyperbolic and isodiachronic location. Journal of the Acoustical Society of America 112, 3046-3052.

Spiesberger, J. L. 1999. Locating animals from their sounds and tomography of the atmosphere: experimental demonstration. Journal of the Acoustical Society of America 106(2), 837-846.

Spiesberger, J. L. 2001. Hyperbolic location errors due to insufficient numbers of receivers. Journal of the Acoustical Society of America 109(6), 3076-3079.

Spiesberger, J. L. and Fristrup, K. M. 1990. Passive localization of calling animals and sensing of their acoustic environment using acoustic tomography. American Naturalist 135(1), 107-153.

Thode, A., D'Spain, G. L., and Kuperman, W. A. 2000. Matched-field processing, geoacoustic inversion, and source signature recovery of blue whale vocalizations. Journal of the Acoustical Society of America 107(3), 1286-1300.

Urick, R..J. 1983. Principles of Underwater sound, Peninsula Publishing.

Wahlberg, M., Møhl, B., and Madsen, P. T. 2001. Estimating source position accuracy of a larger-aperture hydrophone array for bioacoustics. Journal of the Acoustical Society of America 109(1), 397-406.

LOCALISATION OF RIGHT WHALE SOUNDS IN THE WORKSHOP BAY OF FUNDY DATASET BY SPECTROGRAM CROSS-CORRELATION AND HYPERBOLIC FIXING

Marjo H. Laurinolli[1], Alex E. Hay[2]

[1]Department of Fisheries and Oceans, Bedford Institute of Oceanography, PO Box 1006, Dartmouth, NS B2Y 4A2

[2]Department of Oceanography, Dalhousie University, Halifax, NS B3H 4J1

ABSTRACT

In September 2002, five ocean-bottom hydrophones recorded acoustic data in the Bay of Fundy at 1200 Hz sampling frequency for 165.6 h. Arrival time differences for 15 right whale sounds (5 gunshots, 10 tonals) were determined by spectrogram cross-correlation of logarithmic (i.e. dB re 1 $\mu Pa^2/Hz$) spectral densities. The sound source locations were estimated from the intersections of the linearly independent equal time difference hyperbolae for different hydrophone pairs. The root-mean-square (RMS) localisation error was examined using three sound speeds. The lowest average RMS error of 0.85 km was obtained for 1485 m/s, roughly 7 m/s less than the measured average sound speed. The non-gunshot sounds had greater localisation error than the gunshot sounds by 0.4 km. The mean and maximum ranges from the centre hydrophone in the array were 10 km and 33 km respectively.

RÉSUMÉ

En septembre 2002, cinq hydrophones ancrés au fond marin dans la Baie de Fundy ont enregistré des données acoustiques échantillonnées à 1200 Hz pour une durée de 165.6 h. Des différences de temps d'arrivée pour 15 sons de baleines franches (5 « coups de feu », 10 tonals) ont été déterminés par corrélation croisée de spectrogrammes de densité spectrale logarithmique (i.e. dB re 1 $\mu Pa^2/Hz$). La localisation des sources sonores a été estimée à partir des intersections d'hyperboles linéairement indépendantes de différences temporelles égales pour différentes paires d'hydrophones. La moyenne quadratique (RMS) de l'erreur de localisation a été examinée en utilisant trois vitesses de son. L'erreur RMS moyenne la plus basse (0.85 km) a été obtenue avec 1485 m/s, soit 7 m/s de moins que la mesure moyenne de la vitesse du son. Les vocalisations avaient une plus grande erreur de localisation que les sons « coup de feu», soit 0.4 km de plus. Les distances moyennes et maximales à partir de l'hydrophone central du réseau étaient de 10 et 30 km respectivement.

1. INTRODUCTION

The North Atlantic right whale population is in serious jeopardy from mortalities related to anthropogenic activities (*e.g.* Caswell *et al.* 1999, Perry et al. 1999, Knowlton *et al.* 1992, Kraus 1990). Passive acoustics has been previously suggested for monitoring cetacean location and presence, both by the current authors and others (Laurinolli *et al.* 2003, Mellinger *et al.* 2000, Folkow and Blix 1991, Clark and Fristrup 1997). This method allows for non-obtrusive, round-the-clock observation of whale locations and behaviour, and thus provides a potential means for reduction of ship-strikes and fishing gear entanglements.

A small data set of right whale sounds was distributed to the workshop group to encourage comparison of different localisation techniques and error analysis. The techniques used in this paper, spectrogram cross-correlation and hyperbolae of equal time difference, are effective for obtaining first-order location estimates, and are meant to

provide a basis against which the other estimates in this volume can be compared.

2. METHODS

Five ocean-bottom hydrophones (OBH) were deployed in the Bay of Fundy in a face-centred square array about 14 km on a side (Figure 1, see also Desharnais *et al.*, this issue). The average spacing was 13.8 km and the maximum distance between any two OBHs was 20.2 km. Each device was equipped with an omnidirectional (OAS model E-2SD) hydrophone, a 2-GB disk drive, a temperature-regulated quartz crystal clock, and an acoustic burn-wire release. The OBH's recorded 19.41 min data files at a sampling frequency of 1200 Hz, with a 10-s gap between files. The total recording time was 165.6 h over the 167 h deployment time. The hydrophone signal was low-pass filtered (800 Hz cutoff) prior to recording. The data were digitised using a 12-bit A/D converter with ±5 V range.

The hydrophones were calibrated on the DRDC acoustic barge in Bedford Basin, and have nearly constant sensitivity over the 50 to 700 Hz frequency range. One instrument recorded levels lower than expected by approximately 20 dB: the records from later deployments of this unit in the fall of 2002 indicate an intermittent fault in the receiver electronics.

The bottom locations of the OBHs were refined by using an over-the-side hydrophone to detect timing signals from the acoustic pinger on each OBH (Desharnais *et al.*, this issue), in order to account for OBH drift during its descent to the bottom after launch at the surface. The maximum error in clock drift between OBHs was 36 ms and the maximum uncertainty in OBH positions was 12.5 m.
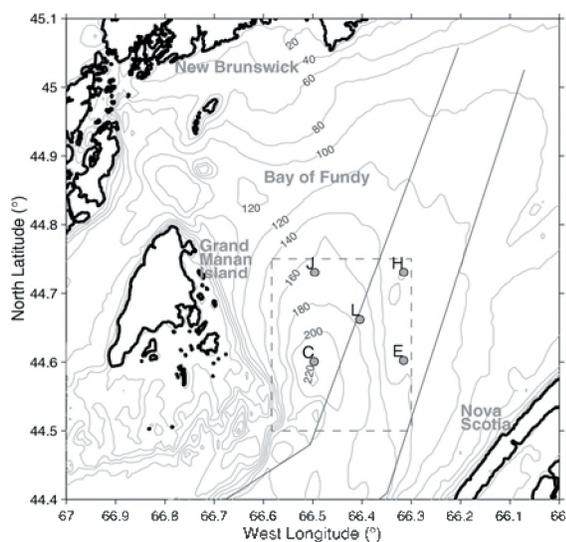
Figure 1. OBH locations in the Bay of Fundy, September 2002.

The data set analysed here consists of a selected set of 15 right whale sounds detected on each of the five hydrophones in the array. Five of these sounds are broadband gunshot-type sounds; the other 10 are tonal-type sounds. The arrival-time-difference between each pair of hydrophones was determined by spectrogram cross-correlation of the logarithmic spectral densities for each sound on the centre OBH against the same sound on the other four OBHs (*e.g.*, Altes 1980, Clark and Ellison 2000). The cross-correlation was performed on a frequency band and time duration manually selected around the sound of interest with 5-Hz frequency resolution and 6-ms time resolution. The intersections of equal-time-difference hyperbolae, for linearly independent time differences among the different hydrophone pairs, then determined the approximate locations of the sound sources. The computations were made assuming constant and spatially uniform sound speed. Sound speed varied between 1490 to 1498 m/s during the time of acoustic sampling (Fig. 2). The variation of RMS

location error with sound speed was evaluated at the three sound speed values of 1480, 1485, and 1490 m/s based on preliminary tests of which sound speeds would give more precise hyperbolae intersections.

Localisation error estimates were made using a root-mean-square (RMS) distance ε of the hyperbolae intersections from the mean, $\varepsilon^2 = \varepsilon_x^2 + \varepsilon_y^2$ where ε_x and ε_y are the standard deviations in the zonal and meridional directions respectively.

3. RESULTS

Spectrogram cross-correlation (Fig. 3) and hyperbolae of equal time difference (Fig. 4) produced localisations for all 15 right whale sounds sampled. The seven points of intersection of hyperbolae triples are within about 0.5 km of each other in the example shown (Fig. 5). The average of these points provides the approximate sound source location relative to OBH-L and the standard deviation in the points provides an estimate of the location error. Table 1 lists the relative arrival times of each sound on each OBH.
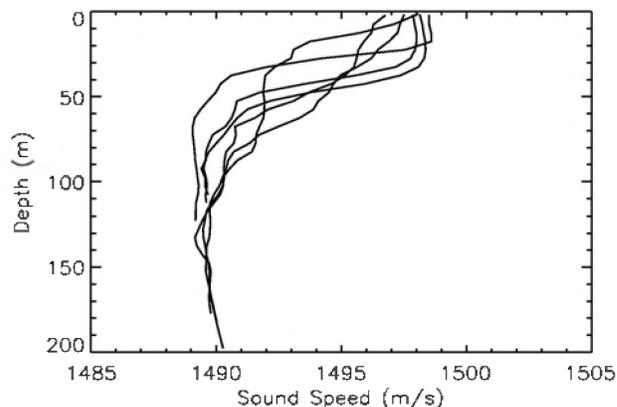
Figure 2. Sound speed profiles from six CTD casts taken near the time of the acoustic data samples.

The locations of the sounds are plotted in Figure 6, together with error bars representing plus or minus one standard deviation. Most of the sound locations are either within or relatively close to the footprint of the array. Two sounds are located about 30 km south of the array centre and, this distance being large compared to the array aperture, are associated with the largest errors.

The smallest RMS localisation error obtained with these data was 0.23 km. Variation of sound speeds for each of the whale sounds affected the localisation precision. Although the average measured sound speed was 1492 m/s, a speed of 1485 m/s gave the best results (smallest overall average RMS error) among the three sound speeds used (Fig. 7). The mean RMS error at 1485 m/s was 0.85 km as compared to 1.10 km and 1.19 km for 1480 m/s and 1490 m/s

respectively (Table 2). Speeds greater than 1490 m/s resulted in very poor localisations. The maximum RMS error at 1485 m/s was 5.6 km for one of the distant (~30km away) mid-frequency sounds. The error for the ten tonal-type sounds (1.0 km) was greater than for the five gunshot sounds (0.6 km). The mean range to the sounds from OBH-L was 10 km and the maximum range was 33 km. The locations, error and range of each localised sound are given in Table 3.
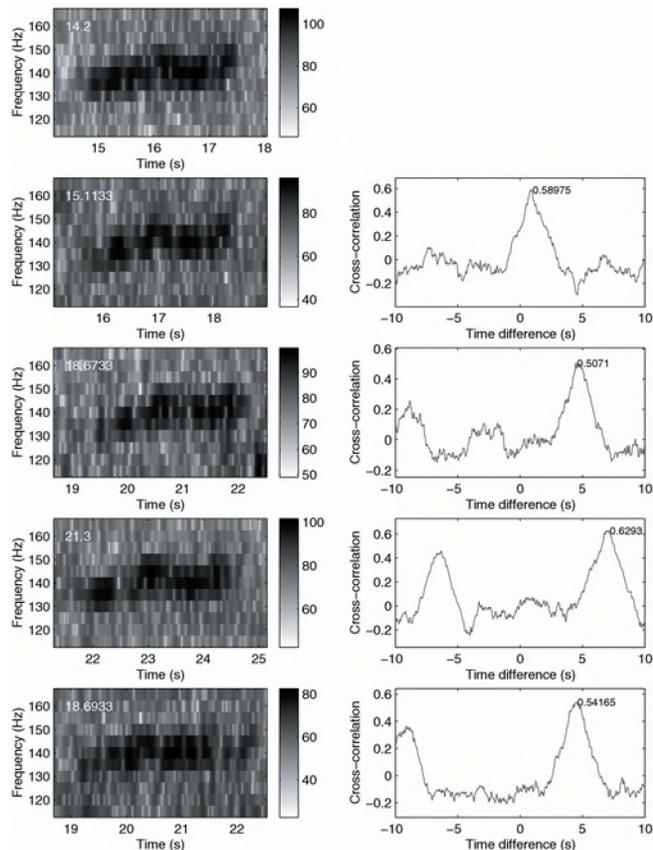
Figure 3. Spectrogram cross-correlation of a tonal sound S-143-8 on OBH-L against OBH-C, E, H, and J respectively. Greyscale in dB re 1 µPa²/Hz. The relative time difference between the first spectrogram and the other four is estimated from the time of the peak in the cross-correlation function.
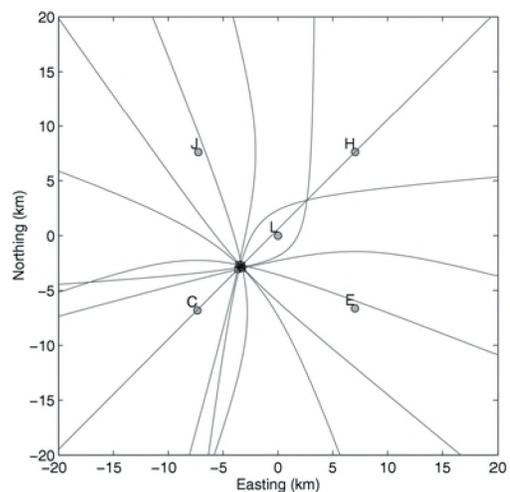
Figure 4. Hyperbolae of equal time difference for S-143-8.
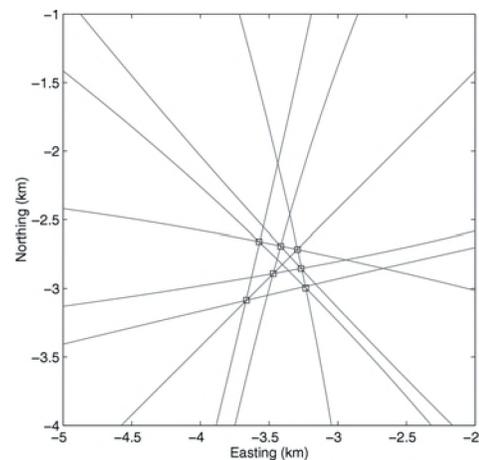
Figure 5. Close-up of localisation in Fig. 3 with squares marking intersections of hyperbolae triples.

Table 1. Relative sound arrival times on each OBH.

Filename	Time (s)				
	OBH-L	OBH-C	OBH-E	OBH-H	OBH-J
S013-1	14.600	13.720	16.000	21.273	20.220
S035-2	14.760	20.527	12.953	14.673	21.247
S070-3	14.680	8.620	18.333	21.500	15.727
S093-4	14.280	20.553	19.520	13.733	15.727
S110-5	14.600	8.813	18.373	21.467	16.107
S092-7	15.400	21.333	20.647	13.227	14.493
S093-9	14.600	17.167	21.347	18.527	10.287
S131-10	15.320	22.320	20.107	17.407	20.227
S131-11	15.400	22.647	19.880	17.487	20.640
S131-12	14.680	21.733	20.273	18.213	19.667
S131-13	14.600	21.773	20.247	17.940	19.720
S134-6	15.320	22.193	21.580	20.020	20.447
S143-8	14.200	15.113	18.673	21.300	18.693
S209-14	14.280	10.907	10.047	19.427	20.153
S210-15	14.600	11.273	10.200	19.740	20.493

Table 2. Average error for the three sound speeds and three types of sounds (gunshot, low-frequency tonal, and mid-frequency tonal). All low-frequency tonals were localised at less than 10 km and mid-frequency tonals at more than 28 km from OBH-L.

	RMS (km)		
Sound speed (m/s)	1480	1485	1490
Gunshots	0.45	0.56	0.89
Low-frequency tonals	0.55	0.39	0.43
Mid-frequency tonals	4.86	3.46	5.00
All tonals	1.41	1.00	1.34
All sounds	1.09	0.85	1.19

Table 3. Position, error, and range of sounds relative to OBH-L, for 1485 m/s sound speed.

Filename	Type	x (km)	ε_x (1 sd)	y (km)	ε_y (1 sd)	RMS (km)	Range (km)
S013-1	G	-1.84	0.03	-6.55	0.06	0.07	6.80
S035-2	G	8.95	0.76	-0.97	0.62	0.98	9.00
S070-3	G	-11.20	0.56	-6.39	0.36	0.66	12.89
S093-4	G	1.32	0.12	6.42	0.22	0.25	6.56
S110-5	G	-9.75	0.71	-5.75	0.44	0.84	11.32
S092-7	LF	0.90	0.18	9.69	0.42	0.46	9.73
S093-9	LF	-6.95	0.49	5.25	0.35	0.60	8.71
S131-10	LF	2.37	0.17	2.86	0.17	0.24	3.72
S131-11	LF	2.83	0.30	2.60	0.29	0.42	3.85
S131-12	LF	1.23	0.23	2.51	0.21	0.31	2.79
S131-13	LF	1.35	0.35	2.58	0.30	0.46	2.91
S134-6	LF	0.18	0.32	2.06	0.22	0.39	2.07
S143-8	LF	-3.42	0.16	-2.84	0.16	0.23	4.45
S209-14	MF	2.64	0.50	-29.10	5.58	5.60	29.22
S210-15	MF	3.16	0.23	-32.94	1.29	1.31	33.09
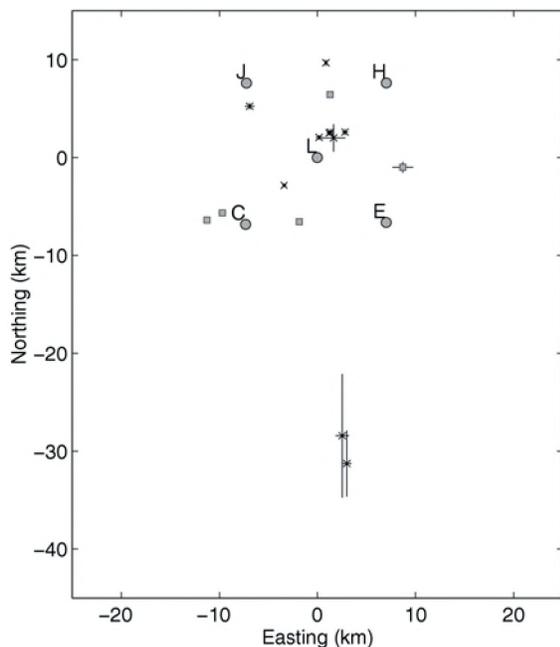
Figure 6. Localised position of each sound with 1 sd error bars. Sound speed = 1485 m/s. Non-gunshots are crosses, and gunshots are squares. Circles represent the OBHs.
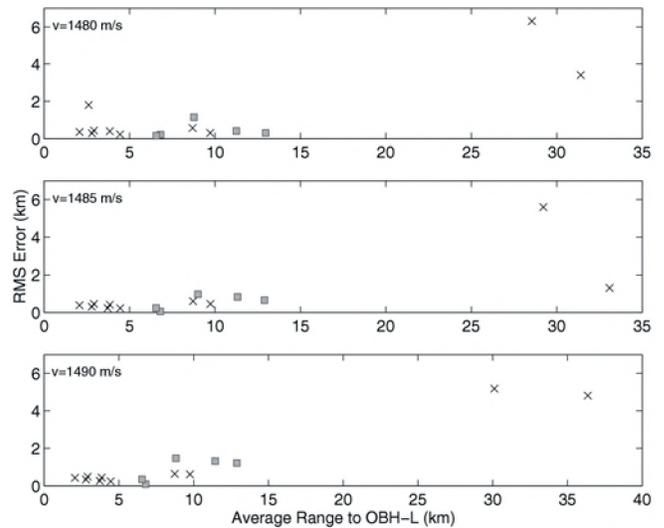
Figure 7. RMS error at three sound speeds. Non-gunshots are crosses, and gunshots are squares.

4.   DISCUSSION

The right whale sounds in the Bay of Fundy workshop dataset were localised assuming isovelocity, two-dimensional sound propagation and using spectrogram correlation to determine the arrival time differences. The method yields a relative RMS error, based on the variance among the seven linearly independent location estimates, of 1 km overall (*i.e.* averaged over all 15 sounds). The error for low-frequency tonal sounds and gunshots located at distances less than about 10 km away from the centre of the array is about 0.5 km. The mid-frequency tonal sounds resulted in a much higher error of about 5 km. There were only 2 mid-frequency sounds, both 30-km distant, and the larger error for these sounds is likely due to this distance being larger than the array aperture and not due to the sounds being of higher frequency. Range error increases with distance from the array with the hyperbolic fixing technique.

Errors in arrival-time-differences determined from spectrogram cross-correlations were unlikely to have resulted in significant localisation errors. Time resolution in the spectrograms was 6 ms, so a shift in the cross-correlation peak of 5 samples would result in 50 m error in the localisation. The errors in the OBH timing and positions could have contributed about 70 m error in the localisations. These errors were small compared to the 340 m difference between using 1485 m/s and 1490 m/s sound speed.

The location results of this study can be directly compared to those obtained on the same dataset by Desharnais *et al.* (this issue) and Simons *et al.* (this issue). As with a sonobuoy study in the same area (Laurinolli *et al.* 2003), right whales were localised as far as 30 km and the average

localisation distance was 10 km. The smaller location error obtained with a sound speed less than the mean measured speed indicates that direct-path detection is unlikely. The sounds are taking longer to get to the receivers because of multipaths and reflections not because the sound speed is lower than expected. Thus, the location error could likely be reduced by allowing the sound speed profile to vary with time and space, requiring the use of range-dependent ray or modal sound propagation models.

ACKNOWLEDGEMENTS

REFERENCES

Altes, R.A. 1980. Detection, estimation, and classification with spectrograms. *J. Acoust. Soc. Am.* 67(4):1232-1246.

Caswell, H., M. Fujiwara, and B. Solange. 1999. Declining survival probability threatens the North Atlantic right whale. *Proc. Natl. Acad. Sci. USA.* 96:308-3313.

Clark, C. and W. Ellison. 2000. Calibration and comparison of the acoustic location methods used during the spring migration of the bowhead whale, *Balaena mysticetus*, off Pt. Barrow, Alaska, 1984-1993. *J. Acoust. Soc. Am.* 107(6):3509-3517.

Clark, C. and K. Fristrup. 1997. Whales '95: A combined visual and acoustic survey of blue and fin whales off Southern California. *Rep. Int. Whal. Commn.* 47:583-600.

Desharnais, F., M. Côte, C.J. Calnan, G.R. Ebbeson, D.J. Thomson, N.E.B. Collison and C.A. Gillard. 2004. Right whale localization using a downhill simplex inversion scheme. *Canadian Acoustics, Vol 32, #2, June 2004.*

Desharnais, F., M.H. Laurinolli, D.J. Schillinger. and A.E. Hay, 2004. A descripton of the workshop datasets. *Canadian Acoustics, Vol 32, #2, June 2004.*

Folkow, L. and A.S. Blix. 1991. Norwegian whale sighting and acoustic surveys in the Atlantic Ocean during the winter of 1989/90. *Rep Int. Whal. Commn.* 41:531-538.

Knowlton, A., J. Sigurjosson, J. Ciano, and S. Kraus. 1992. Long-distance movements of North Atlantic right whales (*Eubalaena glacialis*). *Mar. Mamm. Sci.* 8(4):397-405.

Kraus, S. 1990. Rates and potential causes of mortality in North Atlantic right whales (*Eubalaena glacialis*). *Mar. Mamm. Sci.* 6(4):278-291.

Laurinolli, M.H., A.E. Hay, F. Desharnais, and C.T. Taggart. 2003. Localization of North Atlantic right whale sounds in the Bay of Fundy using a sonobuoy array. *Mar. Mamm. Sci.* 19(4):708-723.

Mellinger, D., C. Carson and C. Clark. 2000. Characteristics of minke whale (Balaenoptera acutorostrata) pulse trains recorded near Puerto Rico. *Mar. Mamm. Sci.* 16(4):739-756.

Perry, S., D. DeMaster, and G. Silber. 1999. The great whales: history and status of six species listed as endangered under the U.S. Endangered Species Act of 1973. *Mar. Fish. Rev.* 61(1):1-74.

Simons, D.G., C. van Moll, and W. Boek. 2004. A three-stage method for localization of marine mammals. *Canadian Acoustics, Vol 32, #2, June 2004.*

RIGHT WHALE LOCALISATION USING A DOWNHILL SIMPLEX INVERSION SCHEME

Francine Desharnais[1], Mathieu Côté[1], Colin J. Calnan[2], Gordon R. Ebbeson[1],
David J. Thomson[1], Nicole E.B. Collison[1] and Chris A. Gillard[3]
[1]DRDC Atlantic, P.O. Box 1012, NS, Canada, B2Y 3Z7
[2]Xwave, 36 Solutions Dr, Halifax, NS, Canada, B3S 1N2
[3]DSTO, P.O. Box 1500, Salisbury SA 5108, Australia

ABSTRACT

The downhill simplex part of a hybrid nonlinear inversion procedure combining simulated annealing with a downhill simplex algorithm [S. E. Dosso and M.J. Wilmut, "An adaptive hybrid algorithm for geoacoustic inversion," Proceedings of the Fifth European Conference on Underwater Acoustics, 185-190 (2000)] was used to localize sounds from the workshop dataset. The procedure relies on relative arrival times for the direct propagation paths from the sources to each receiver. An eigenray model [S.E. Dosso, N.E.B. Collison, "Acoustic tracking of a freely drifting sonobuoy field," J. Acoust. Soc. Am. 111, 2166-2177 (2002)] was used to estimate travel time along the direct paths. The positions of the receivers and whales were obtained by inversion of estimated relative travel delays. It was found that the direct path assumption was a problem for the distances involved with the workshop dataset. This paper will discuss the solution of using a constant sound speed instead of actual sound speed profiles for localisation, as well as the error associated with arrival time inaccuracies. The error growth for multi-source cases will also be discussed.

RÉSUMÉ

Un algorithme de descente du simplexe faisant partie d'une procédure hybride d'inversion non linéaire combinant un recuit simulé et de descente du simplexe [S. E. Dosso et M.J. Wilmut, « An adaptive hybrid algorithm for geoacoustic inversion », Proceedings of the Fifth European Conference on Underwater Acoustics, 185-190 (2000)] a servi à localiser des sons dans un ensemble de données de travail. Cette procédure se fonde sur les temps relatifs d'arrivée pour des trajets de propagation directe entre les sources et chacun des récepteurs. Un modèle à vecteurs propres [S.E. Dosso, N.E.B. Collison, « Acoustic tracking of a freely drifting sonobuoy field », J. Acoust. Soc. Am. 111, 2166-2177 (2002)] a permis d'évaluer le temps de propagation sur des trajets directs. Les positions des récepteurs et des baleines ont été obtenues par l'inversion des délais de propagation relatifs estimés. Il s'est avéré que l'hypothèse des trajets directs posait un problème aux distances comprises dans l'ensemble de données de travail. Ce document examine la solution consistant à utiliser une vitesse de son constante plutôt que des profils réels de vitesse du son pour la localisation, ainsi que l'erreur liée aux imprécisions des temps d'arrivée. Le développement de l'erreur dans les cas de sources multiples est également abordé.

1. INTRODUCTION

In this paper, the downhill simplex part of a hybrid nonlinear inversion scheme that combines simulated annealing with a downhill simplex algorithm [1] is used to localise positions of North Atlantic right whales using a sparse array of five Ocean Bottom Hydrophones (OBHs). The sounds used were extracted from the DRDC/Dalhousie dataset provided at the *Workshop on Detection and Localization of Marine Mammals Using Passive Acoustics,* held at Dartmouth, NS, Canada, 19–21 November 2003.

A frequency-based cross-correlation procedure was used to determine relative arrival times of the right whale sounds on each OBH. Synthetic arrival times based on measured sound speed profiles were calculated using an eigenray model, assuming direct path propagation. The inversion procedure derived whale positions by minimizing the mismatch between the measured and modeled relative arrival times.

This paper describes the detection and localisation procedures, and presents the whale position estimate results. The effect associated with using a constant sound speed *vs.* an actual sound speed profile is discussed, as well as the

impact of detection times on localisation.

This paper concentrates on the 2002 workshop dataset. A 2000 calibration dataset was also available to the workshop participants. The 2000 dataset was also analyzed, and the results are included in Appendix A for reference. These results were also given to the workshop organizers to allow for comparisons with other authors.

Details regarding both datasets and a description of the experimental environment (Bay of Fundy) are given in an accompanying paper in these proceedings [2].

2. PROCEDURES

2.1 Arrival times

The detection algorithm used on the DRDC/Dalhousie dataset was kept simple on purpose. Although the accuracy of the technique is expected to be less than that associated with the cross-correlation techniques typically used in marine mammal localisation work [3,4], it was nonetheless adopted to introduce diversity in the comparison of results obtained with other techniques presented at the workshop. The use of a different detection scheme permits the effect of travel time error on localization accuracy to be assessed.

The technique utilized here consisted of first displaying a sonogram of the whale sound, as shown in Fig. 1. To produce this figure, a 1024-pt FFT was used with 90% overlap in time. A time and frequency window concentrating on the desired portion of the sound is selected from the gram (shown as the rectangle in Fig. 1). In the case of vocalizations, this selection was limited to one or possibly two of the strongest harmonics. In the case of a "gunshot" event, most of the frequency band was selected. For numerical convenience, the time window was adjusted to contain a multiple of the FFT size, centered on the pre-selected time period. This corresponds to approximately 0.85 s for the year 2002 dataset, where the sampling frequency was 1200 Hz. The time and frequency window selection was based on the sounds recorded on OBH sensor "L". The arrival times at mid-sound and the frequency band values are given in Section 3.1.

The sonograms of the signals recorded on all other OBHs were produced in a similar manner, and a first estimate of arrival time for the sounds were picked from these grams. A hanning window was then applied to the time series corresponding to the selected time window of each OBH. The windowed and filtered signal of OBH "L" was then cross-correlated in the frequency domain to the windowed signals of the other OBHs, following the technique of Carter and Ferrie [5], which is a different technique then [3,4]. The cross-correlation peaks defined the final relative arrival time

delays of the sounds on all OBHs relative to OBH "L".
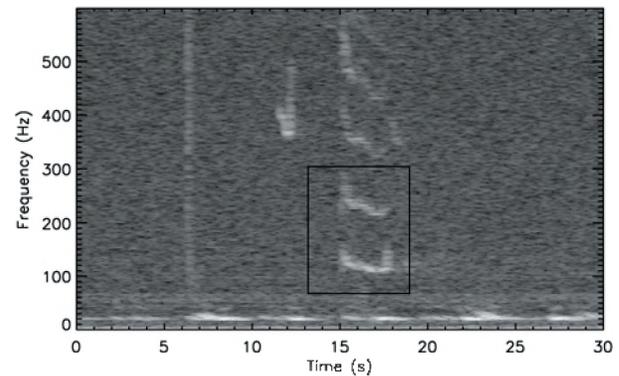
Figure 1. Sonogram of vocalization S131 as a function of time and frequency. The intensity levels are relative. The box represents the selected time and frequency band.

2.2 Localisation

Relative arrival times can be calculated from the measured sound speed profile with an eigenray model, and compared with the relative arrival times obtained from the cross-correlation techniques. Since the source position is unknown, an optimization technique is used to search the geographical space in an efficient manner. The estimated sound source position is taken as the location corresponding to the lowest error in the overall fit of the model to the data. The optimization algorithm is based on a hybrid nonlinear inversion procedure combining simulated annealing (SA) with a downhill simplex (DHS) algorithm. It should be noted that in the case of a simple geometric problem such as the inversion of one source and multiple receivers, the error surfaces are expected to present one clear minimum, and no secondary minima (except potentially in the vertical). For this reason, only the DHS part of the algorithm was used for this paper.

2.3 Eigenray model

The eigenray model described in Dosso and Collison [6] was used for this analysis. It provides expressions for the range r and arrival time t along a ray path between the source and receiver in an ocean with a sound speed profile $c(z)$, derived by applying Snell's Law to an infinite stack of infinitesimal layers. Both r and t are functions of $c(z)$ and the ray parameter $p = \cos(\theta(z))/c(z)$. In this model, the ray parameter for an eigenray connecting a source/receiver pair is determined by searching for the value of p that produces an r that equals the geometrical horizontal range (to a specified tolerance). Then this ray parameter is used to calculate the corresponding arrival time t.

2.4 Inversion engine

Inversion procedures have been used in the past for array element localization (AEL) problems, or to precisely localize elements of an acoustic array, using arrival times (if the source instants are known) or arrival time differences. For AEL, this problem has been solved either by linearizing the problem and using start up positions to iterate towards a final solution [7,8], or by searching the 3D space for solutions that will minimize the global relative arrival time error, with an algorithm such as simulated annealing [9].

The whale localization problem is similar to the AEL problem, though the receiving array in this case is very sparse, and the sources positions are completely unknown. In this paper, we chose to use an inversion procedure to search the 3D space for individual whale positions, while assuming the receiver positions to be known. The OBH receivers were located 0.9 m above the seabed and were known within 10 m in the XY plane. Changes in the receiver positions within this uncertainty resulted in negligible effects on the estimated source locations, and thus, fixed receiver positions could be used for the inversions. In this case, only the 3D position of the source was inverted for, using 4 relative arrival times, allowing the problem to remain over-determined.

The objective of the inversion algorithm is to search the model parameter space until a low mismatch value is found. The amount of mismatch between the model and data are calculated via a cost function. The cost function used here is the rms difference between the measured relative arrival times ΔT_{meas}, and the modelled relative arrival times ΔT_{model}:

$$E = \sqrt{\frac{\sum \left(\Delta T_{meas} - \Delta T_{model}(\mathbf{m})\right)^2}{N_{hyd} \cdot N_{src}}} \,, \qquad (2)$$

where N_{hyd} is the number of hydrophones, N_{src} is the number of sources, and the arrival times ΔT are relative to a reference hydrophone (OBH "L"). The model parameters $\mathbf{m} = [x_1, y_1, z_1, \ldots, x_M, y_M, z_M]$ are the 3D positions of the sources to be localized. The unit of the mismatch "E" is second [s].

The parameter space is searched with the DHS part of a hybrid Adaptive Simplex Simulated Annealing (ASSA) search technique developed by Dosso et al. [1,10]. ASSA combines the strengths of the downhill simplex method for minimizing a function locally and simulated annealing that is effective for global random search. Pure DHS minimization is based on an intuitive geometric scheme for moving downhill in a multi-dimensional space. The method operates on a simplex of $M+1$ models in an M-dimensional space. Each model is ranked according to its mismatch E and the simplex undergoes a series of reflections,

expansions and contractions to work its way downhill. After each step, the difference between the highest and lowest mismatches relative to their average is used as a convergence criterion (10^{-5} was used in this paper). The primary advantages of the DHS method are that it retains a memory of regions where the function is small, and it is effective in navigating the search down the axes of long narrow parameter space valleys that are not aligned with the search parameter axes. These valleys are normally due to correlation between parameters in the search space. Whilst more efficient methods for finding local minima exist, the method is fast and efficient enough to use in the hybrid scheme of ASSA. The primary disadvantage of the DHS method alone is that it may become trapped in a local minimum and often must be started from many different points in the search space.

In this localization exercise, however, a single source is inverted for. The error surface for each source is expected to present a single minimum and no secondary minima (except potentially in the vertical, for particular deployment geometries other than ours) for known fixed receivers positions. In such a case, there is no risk of the DHS algorithm getting trapped in a secondary minimum, and is therefore used alone for the inversion.

3. RESULTS

3.1 Arrival times of the 2002 dataset

Table 1 summarizes the arrival times on OBH "L" (2002 dataset), as well as the frequency band selected for each whale sound for the cross-correlation. The arrival times on the other sensors (relative to OBH "L") are listed in Table 2. For reference, the results for the 2000 calibration dataset are listed in Appendix A.

Table 1. Arrival times on sensor "L" of selected sounds for dataset 2002, and processing bands used for cross-correlation.

Sound #	Sound name	Freq. band [Hz]	T_L [s]
1	S013-1	126-485	14.54
2	S035-2	29-550	15.31
3	S070-3	42-560	16.84
4	S093-4	40-534	14.71
5	S110-5	76-571	16.93
6	S092-7	79-222	16.07
7	S093-9	89-172	16.59
8	S131-10	58-157	13.02
9	S131-11	86-336	17.95
10	S131-12	68-274	17.01
11	S131-13	68-303	16.25
12	S134-6	71-448	17.44
13	S143-8	86-191	16.33
14	S209-14	308-542	15.31
15	S210-15	360-568	16.33

3.2 Localisation of the 2002 dataset

The 3D receiver positions were assumed to be known exactly, and were taken from [2]. The search space for the sources was limited by the bounds listed in Table 3. The bounds for the source positions were originally set to ±20 km for all sources; further analysis showed that two of the sources were located well south of the OBH pattern. The search space for these 2 sources was shifted south, but the 40 km ranges were preserved to simplify comparisons.

Table 2. OBH arrival times (ΔT's) for sensors "C", "E", "H" and "J" relative to sensor "L" for dataset 2002 (ΔT>0 denotes arrival on sensor "L" first).

Snd #	Snd name	ΔT_C [s]	ΔT_E [s]	ΔT_H [s]	ΔT_J [s]
1	S013-1	-0.92	1.45	6.58	5.61
2	S035-2	5.77	-1.88	-0.08	6.23
3	S070-3	-6.09	3.86	6.96	0.87
4	S093-4	6.09	5.22	-0.51	0.75
5	S110-5	-5.79	3.79	6.93	1.42
6	S092-7	5.85	5.25	-2.27	-0.83
7	S093-9	2.44	6.62	3.95	-4.59
8	S131-10	6.73	4.75	2.18	5.08
9	S131-11	7.27	4.62	2.25	5.17
10	S131-12	7.01	5.69	3.55	5.11
11	S131-13	7.04	5.65	3.49	5.12
12	S134-6	6.86	6.74	4.86	5.24
13	S143-8	0.96	4.79	6.99	4.26
14	S209-14	-3.37	-4.32	5.53	5.88
15	S210-15	-3.31	-4.36	5.20	5.82

Table 3. Search bounds used for the inversion. Values for the receivers are relative to individual OBH positions [2]; values for sources are relative to OBH "L", located at (0,0) in Cartesian coordinates.

	X	Y	Z
All sources but S209-14 and S210-15	±20 km	±20 km	0 to 214 m
S209-14 and S210-15	±20 km	-40 to 0 km	0 to 214 m

Each source position was inverted individually (one source per inversion event), using only the DHS part of the inversion algorithm. The sound speed profile that was measured closest in time to the sound recording time was selected for each individual sound [2]. For each inversion run, a new seed was used in the random number generator for an arbitrary initial search of the space, providing different final solutions for each run. Ten inversions were carried out for each source to provide a qualitative characterization of the algorithm variability, via a statistical distribution of the inversion solutions. On the order of 1000 to 2000 forward models were computed for each inversion.

The inversion algorithm converged for seven of the 2002 sources in this first attempt (sources #3, 8, 9, 10, 11, 12 and 13), based on the convergence criterion described in Section 2.4. The solution with the lowest mismatch E (for each individual source) is shown in Fig. 2. The origin of the plot is centered on OBH "L". The source numbers are as listed in Table 1. The mismatch E (or overall mismatch between the modelled and measured time arrivals) had an average of 0.128 s, excluding source #3 (S070-3), which had a mismatch of 3.54 s. The average inverted depth for the sources was 124 m. The inversion code did not converge for the remaining sources (sources #1, 2, 4, 5, 6, 7, 14 and 15).
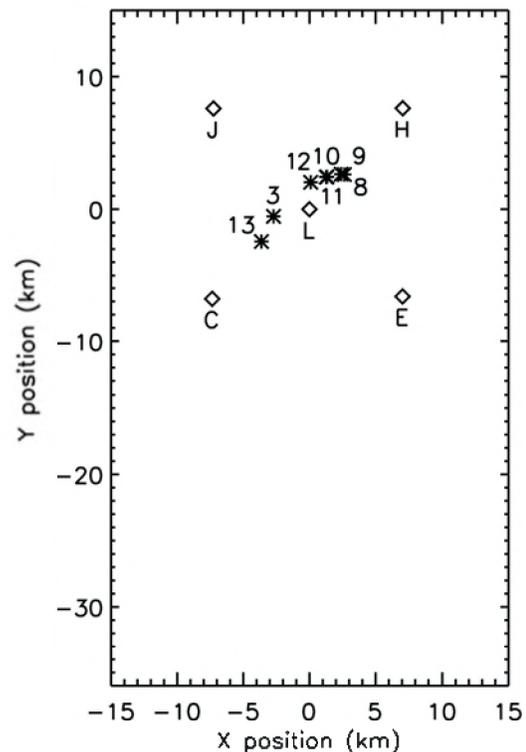
Figure 2. Localisation results using measured sound speed profiles. Diamonds represent OBHs; stars represents inverted sound positions. Sound numbers are as listed in Table 1.

In order to assess why the mismatch for source 3 was so large, and why some of the sources could not be localized, a ray plot was produced for a sound source located at 100-m depth. The left-hand side of Fig. 3 shows a simplified version of the sound speed profile from file "T7_00004.EDF", which was arbitrarily selected for this analysis; the right-hand side shows 20 rays traced at 2° intervals from 10° incidence below horizontal to 28° above horizontal.

It can be seen from Fig. 3 that a direct ray to the seabed

(where the hydrophones were located) does not exist for this source depth beyond approximately 6 km in range. If the source depth was shallower (results not shown here), the maximum propagation range would be shortened. For a source near the seabed, longer paths do exist, but refraction paths closer to the surface would lead to longer travel times.

The inversion code could not converge for sources that were located beyond the range of a direct path. For those sources where the algorithm did not converge, the average depth of 124 m corresponds to a compromise solution that minimized arrival time mismatch while allowing the rays to reach the sources.

3.3 The effect of constant sound speed

With an OBH pattern on the order of 14 km in extent, it is not surprising that direct paths cannot reach the required range in water depths of 220 m or less, unless the whales are near the middle of the pattern. As explained in [11], the multipath effect is aggravated by the position of the sensors near the seabed. The signals captured by the OBHs include multipath arrivals, and our cross-correlation technique has no way of distinguishing individual ray paths.
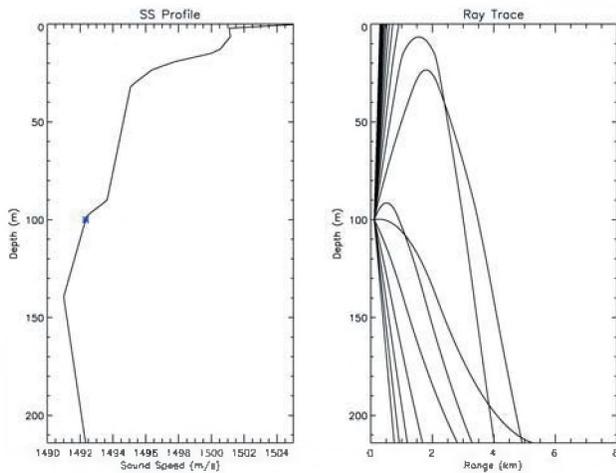
Figure 3. Sound speed profile (left panel) from "T7_00004.EDF" and ray paths (right panel) traced at 2° intervals from -10° to +28° grazing angle.

A simple method for avoiding this difficulty is to replace the measured sound speed profile with some average constant sound speed. In order to determine an effective constant sound speed, the behaviour of the mismatch E as a function of sound speed was investigated, using the sound that was arbitrarily selected from the first file (S013-1).

The sound speed was varied from 1488 m/s to 1502 m/s in increments of 1 m/s; a limit lower than the lowest sound

speed in the water column was used as indirect paths treated as direct paths would imply a speed slower than the average sound speed in the water column. As before, the inversion procedure was carried out ten times for each sound speed, and the solution with the lowest mismatch E is plotted as a function of sound speed in Fig. 4.
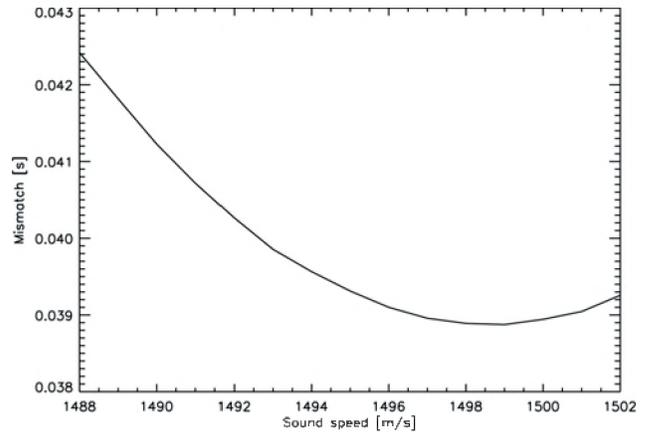
Figure 4. Mismatch _vs._ sound speed for sound S013-1.

The mismatch shows a broad minimum with a minimum value at a sound speed of 1499 m/s. This sound speed was therefore selected as an effective constant sound speed to use for each source. The inversion procedure was applied to each source using this effective constant sound speed, and the updated best results are shown in Fig. 5. Convergence was obtained for all fifteen sources.

These updated results show that several sources are now localized outside of the OBH pattern, as far as 35 km away from OBH "L", located in the middle of the pattern. The standard deviation of all ten runs was calculated as an indicator of the algorithm variability; it was on the order of 20 m inside the OBH pattern, and increased with range to reach 3 km for the two sources that are approximately 34 km south. This variability is a good indicator of the width of the minima in the mismatch surfaces for the various sources depending on their positions relative to the OBH pattern. The mean mismatch for all sources was 0.104 s, lower than the 0.128 s obtained with the actual sound speed profiles (Fig. 2). The source depths were again on the order of 120 m (from 80 to 152 m).

The difference between the results using the actual profiles (method A) and those obtained with a constant sound speed (method B) are summarized in Table 4, for the seven sources for which the algorithm converged. For these sources, the depth results were very similar for both methods, mainly within 17 m of each other. In the XY plane, the two methods localized sources within 15 m of each other, except for the third source (S070-3). The third

source had a very high mismatch when the actual sound speed profile was used (3.54 s), but it was reduced to 0.135 s with a constant speed of 1499 m/s. Fig. 5 also shows that the third source was localized outside of the OBH pattern by method B, over 11 km away from the position estimated by method A. It is believed that the new position is more accurate. Method A was able to converge on this source, but because of the direct path assumption, it could not reach the actual source location. The solution was not optimal, as demonstrated by the very high mismatch.
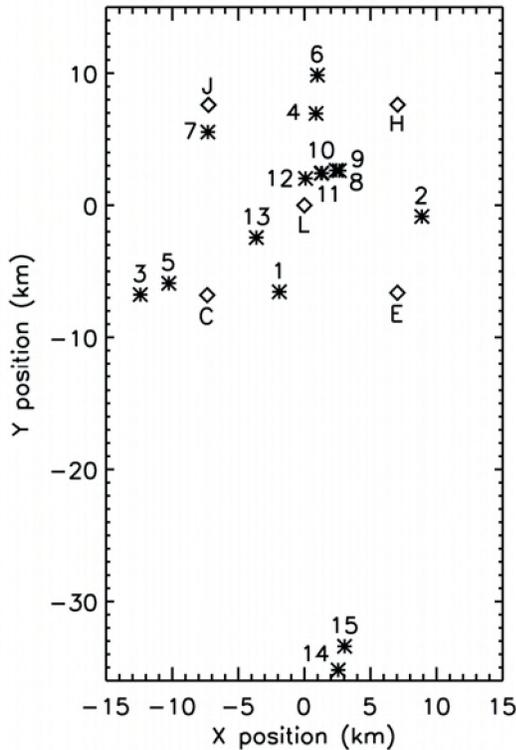
Figure 5. Localisation results using an effective constant sound speed of 1499 m/s. Diamonds represent OBHs; stars represents inverted sound positions.

Table 5 lists the best position estimates from method B. These final positions were used to compare the results of this localisation technique with those of other authors (see overall discussion on results elsewhere in these proceedings).

3.4   The impact of arrival times on localisation

As a further check on accuracy, the impact of arrival time estimates on inverted positions was investigated using arrival times from other authors. The cross-correlation scheme of Laurinolli et al. [12] was expected to lead to more accurate arrival times than the simple scheme used in this paper, as it takes full advantage of the time/frequency

structure of whale vocalizations. Two of the sources with the smallest arrival time differences, and two of the sources with the greatest differences were selected for this test. The relative arrival times of Laurinolli et al. [12] were used to invert positions with the localisation technique described in Section 2.2. Table 6 shows the test results for these sounds. We emphasize that E {Laurinolli} in Table 6 indicates inversion results using the Laurinolli et al. arrival times, as opposed to their positions results.

Table 4. Differences between inverted positions using the actual sound speed profiles (method A) and an effective constant sound speed of 1499 m/s (method B), for the seven sounds for which both methods converged. $|\Delta XY|$ is the difference in horizontal distance. Values were rounded to the nearest integer.

Sound #	Sound name	$|\Delta Z|$ [m]	$|\Delta XY|$ [m]
3	S070-3	35	11528
8	S131-10	3	7
9	S131-11	8	7
10	S131-12	14	3
11	S131-13	11	0
12	S134-6	17	7
13	S143-8	5	13

Table 5. Best positions (lowest mismatch) with a constant sound speed of 1499 m/s. Values were rounded to the nearest integer.

Sound #	Sound name	X [m]	Y [m]	Z [m]
1	S013-1	-1899	-6551	126
2	S035-2	8884	-848	137
3	S070-3	-12417	-6768	111
4	S093-4	879	6950	77
5	S110-5	-10262	-5904	142
6	S092-7	987	9857	80
7	S093-9	-7305	5545	133
8	S131-10	2403	2642	115
9	S131-11	2639	2625	148
10	S131-12	1267	2425	117
11	S131-13	1335	2442	143
12	S134-6	91	2027	112
13	S143-8	-3639	-2446	135
14	S209-14	2558	-35192	152
15	S210-15	3037	-33394	86

Table 6. Comparison of results for selected sounds with the greatest and smallest differences in arrival times between Desharnais *et al* (this paper) and Laurinolli *et al.* [12]. Table includes general location of sound relative to OBH pattern, difference in detection times (averaged over all source/receiver pairs), difference in resulting positions (range and depth), and final energies.

	Greatest differences		Smallest differences	
Sound	S093-4	S134-6	S110-5	S210-15
Sound location	inside edge	middle	outside edge	30 km away
ΔT [s]	0.24	0.19	0.04	0.06
ΔXY [m]	442	104	27	949
ΔZ [m]	-12	21	-4	16
E {Laurinolli} [s]	0.038	0.031	0.133	0.059
E {Desharnais} [s]	0.047	0.048	0.230	0.044

4. DISCUSSION

The downhill simplex part of a hybrid inversion technique combining simulated annealing with a downhill simplex algorithm was used to invert whale call positions using the sparse array of bottom-mounted sensors for the datasets provided for the *Workshop on Detection and Localization of Marine Mammals Using Passive Acoustics*, Dartmouth, NS, 19–21 November 2003.

Using the measured sound speed profile, it was found that several of the sources could not be localized because no direct path existed between the source and some of the receivers. All sources could be localized using an effective constant sound speed approximation.

It was found that low data/model mismatch values were obtained if the source positions were inverted individually using the downhill simplex part of the inversion algorithm only. Whether the full sound speed profile was used (*i.e.,* when a direct path existed between the source and all receivers), or a constant sound speed was assumed, the resulting 3D positions were usually within 15 m of one another. On one occasion, however, the constant sound speed approximation led to a change in estimated position in excess of 11 km. In this case, the high mismatch was a good indicator that the use of the actual sound speed profile led to a poor solution biased by the invalid assumption of an existing direct path.

The choice for an effective constant sound speed was based on a mismatch analysis for one of the sources. This may not have been the best choice for all of the sources, which occurred on different days and different locations. Future work will consider inverting for an optimized sound speed as well as optimized source and receiver positions.

The inverted source depth was usually on the order of 120 m, which is deeper than the presumed depth of vocalizing right whales [13]. With actual sound speed profiles, this depth was necessary to fulfill the requirement of a direct ray, since no direct ray exists at long range between a shallow source and a deep receiver. With a constant sound speed approximation, this depth optimized the data/model fit for the sound speed we selected. Either way, the depth estimates presented here are presumed to be inaccurate.

The issue of using a constant sound speed, or a direct path assumption was debated at length during the Workshop panel discussions. In this dataset, the direct path either does not exist, or is unlikely to be the arrival with the most energy, especially at long range. Our experimental arrival times are thus matched wrongly to direct path arrivals assumed by our eigenray model. The impact of this assumption is two-fold. First, the determination of relative arrival times will be affected, unless a full propagation model that includes multipaths is used. Second, the resulting XY position estimates can be biased. We believe that an optimization technique such as the one presented here, is likely to give good XY results if the source is located within a well-distributed array of receivers, since the biases on average should cancel out. If the source is located outside the receiver pattern, the errors will compound and grow with source-receiver range. The depth results will not be accurate in either case.

A simple frequency-based cross-correlation algorithm was used to determine relative arrival times. This technique does not take full advantage of the frequency/time structure of whale vocalizations. Yet, it led to consistent localization results within a 1 km diameter circle up to a range of 30 km from the positions based on arrival times determined from an alternative technique. For many purposes, this accuracy may be sufficient.

REFERENCES

[1]   S. E. Dosso and M.J. Wilmut, "An adaptive hybrid algorithm for geoacoustic inversion," Proceedings of the Fifth European Conference on Underwater Acoustics, M.E. Zakharia, P. Chevret and P. Dubail, eds., Lyon, France, pp. 185-190, 2000.

[2]   F. Desharnais, M.H. Laurinolli, D.J. Schillinger and A.E. Hay, "A description of the workshop datasets," Canadian Acoustics, June 2004.

[3] C. Clark and W. Ellison, "Calibration and comparison of the acoustic location methods used during the spring migration of the bowhead whale, Balaena mysticetus, off Pt. Barrow, Alaska," 1984-1993. J. Acoust. Soc. Am., 107(6), pp. 3509-3517, 2000.

[4] C. Clark, P. Marler, and K. Beeman, "Quantitative analysis of animal vocal phonology: an application to swamp sparrow song," Ethology, 76, 101–115, 1987.

[5] C.G. Carter & J.F. Ferrie, "A coherence and cross spectral estimation program," IEEE Programs for Digital Signal Processing, edited by the Digital Signal Processing Committee, IEEE Acoustics, Speech, and Signal Processing Society, pp. 2.3-1-18, 1979.

[6] S.E. Dosso, N.E.B. Collison, "Acoustic tracking of a freely drifting sonobuoy field," J. Acoust. Soc. Am. 111, 2166-2177 (2002).

[7] G.H. Brooke, S.J. Kilistoff, and B.J. Sotirin, "Array element localization algorithms for vertical line arrays," Proceedings of the 3rd European Conference on Underwater Acoustics, J.S. Papadakis, ed., Crete U.P., Heraklion, pp. 537-542, 1996.

[8] M. Barley, S. Dosso, and P. Schey, "Array element localization of a bottom moored hydrophone array, Canadian Acoustics 30(4), pp. 3-11, 2002.

[9] M.V. Greening, "Array element localization of rapidly deployed systems," Canadian Acoustics 28(2), pp. 7-13, 2000.

[10] S.E. Dosso, M.J. Wilmut, and A.S. Lapinski, "An adaptive-hybrid algorithm for geoacoustic inversion," IEEE J. Ocean Eng. 26, 324-336, 2001.

[11] D.M.F. Chapman, "You can't get there from here: shallow water sound propagation and whale localization," Canadian Acoustics, June 2004.

[12] M.H. Laurinolli, A.E. Hay, "Localisation of right whale sounds in the workshop Bay of Fundy dataset by sprectrogram cross-correlation and hyperbolic fixing," Canadian Acoustics, June 2004.

[13] Tyack, Pers. comm.

ACKNOWLEDGEMENTS

APPENDIX A. Results for the 2000 calibration dataset

A calibration dataset based on recordings from 2000 was made available to the participants. The results from the analysis of this dataset were given to the organizers to allow comparisons between authors and techniques. The results are presented here for future reference.

A1 Arrival times

Table A1 summarizes the arrival times determined for the 2000 dataset. For each whale sound, the frequency band selected for cross-correlation is also listed.

A2 Localisation of the calibration dataset

Our best position estimates for the 2000 calibration dataset are shown in Fig. A1 and Table A2. The positions were obtained with the measured sound speed profiles, not a constant sound speed. It should be noted that the transmissions may not have occurred at the RHIB boat positions, and therefore our positions should only be compared to other authors' positions, not to the known RHIB boat positions.

Table A1. OBH arrival times (ΔT's) relative to sensor "D" of selected sounds for dataset 2000 (ΔT>0 denotes arrival on sensor "D" first). Also shown is the processing band used for cross-correlation and the travel time to OBH "D".

Sound name	Freq. band [Hz]	T_D [s]	ΔT_B [s]	ΔT_C [s]	ΔT_E [s]
S282	300 – 570	4.69	-1.35	-1.49	-1.66
S282	300 – 570	3.35	-1.30	-1.42	-1.57
S288	50 – 250	3.20	-1.27	-1.50	-1.15
S289	50 – 250	3.10	-1.61	-1.90	-1.40
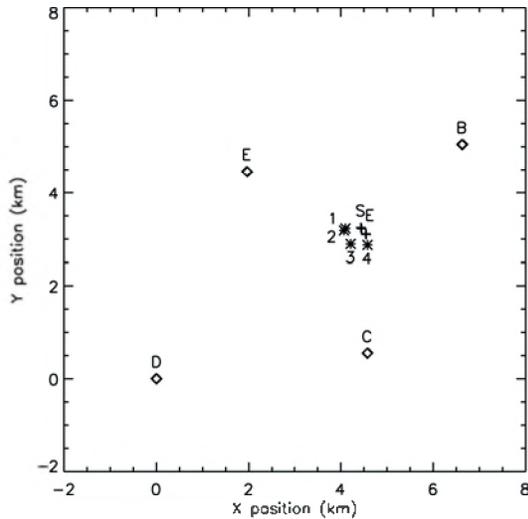
Figure A1. Positions of the sounds from the 2000 calibration dataset. "S" and "E" indicate the start and end positions of the RHIB boat where the playbacks were made from.

Table A2. Best positions (lowest mismatch) for the 2000 calibration dataset. Values were rounded to the nearest integer. Values are relative to OBH "D" located at (0,0).

Sound #	Sound name	X [m]	Y [m]	Z [m]
1	S282	4109	3229	119
2	S282	4107	3205	129
3	S288	4373	2903	103
4	S289	4619	2903	128

WAVEGUIDE PROPAGATION ALLOWS RANGE ESTIMATES FOR NORTH PACIFIC RIGHT WHALES IN THE BERING SEA

Sean M. Wiggins[1], Mark A. McDonald[2], Lisa M. Munger[1], Sue E. Moore[3], John A. Hildebrand[1]

[1]Scripps Institution of Oceanography, 9500 Gilman Drive-0205, La Jolla, CA 92093-0205, swiggins@ucsd.edu

[2]Whale Acoustics, 11430 Rist Canyon Road, Bellvue, CO 80512

[3]NOAA/National Marine Mammal Laboratory, 7600 Sand Point Way NE, Seattle, WA 98115

ABSTRACT

The shallow and uniform water depth of the eastern Bering Sea shelf results in an acoustic waveguide. Propagation within this waveguide produces waveform dispersion which is dependent upon range. We present a means for using dispersed waveforms to determine range to calling whales from a single autonomous acoustic recording instrument. The predominant North Pacific right whale (*Eubalaena japonica*) call is frequency upswept from about 90 Hz to around 160 Hz and lasts approximately 1 s. The regional bathymetry of the eastern Bering Sea middle shelf is relatively uniform and shallow (~ 70 meters deep). This geometry provides a plane-layered waveguide in which right whale upswept calls can be detected at ranges over 50 km and have multiple modal arrivals that become dispersed, displaying different propagation velocities for different frequencies. Dispersion characteristics of modal arrivals are dependent on the calling whale's depth, the receiver's depth, the water depth, the range from caller to receiver, and various environmental parameters including water and sediment density and sound velocity. A model of sound propagation for the eastern Bering Sea middle shelf is developed from right whale call dispersion recorded on sonobuoys and seafloor acoustic recording packages, using individual calls recorded at multiple instruments. After development of the model, waveform dispersion allows estimation of caller range based on single instrument recordings. Estimating range between instrument and calling whales provides a means to estimate minimum abundance for the endangered North Pacific right whale.

RÉSUMÉ

L'eau peu profonde et uniforme de la rive Est de la mer de Béring produit un excellent guide d'ondes acoustiques. Dans ce guide de propagation, la dispersion des ondes sonores est dépendante de la distance. Nous présentons ici un moyen pour utiliser la dispersion des ondes sonores pour déterminer la portée de sons émis par des baleines à partir d'un unique instrument d'enregistrement du signal acoustique. La vocalisation prédominante de la baleine franche du Pacifique Nord (*Eubalaena japonica*) est une modulation ascendante d'environ 90 à 160 Hz et d'une durée approximative de 1 s. La bathymétrie régionale de la rive Est de la mer de Béring est relativement uniforme et peu profonde (~70 m de profondeur). Cette géométrie fournit un guide d'ondes à couches horizontales ou les vocalisations modulées de baleines franches peuvent être détectées à des distances supérieures à 50 km et ont de multiples arrivées modales qui deviennent dispersées, démontrant différente vitesse de propagation à différentes fréquences. Les caractéristiques de dispersion des arrivées modales sont dépendantes de la profondeur de la baleine, la profondeur du récepteur, la profondeur de l'eau, la distance de l'émetteur et du récepteur et une variété de paramètres environnementaux incluant la densité de l'eau et des sédiments, et la vitesse du son dans ces deux media. Un modèle de la propagation du son pour la rive Est de la mer de Béring est développé à partir de la dispersion des vocalisations des baleines franches enregistrées à partir de bouées acoustiques et de systèmes acoustiques ancrés sur le fond marin, en utilisant les vocalisations individuelles enregistrées à partir de multiples instruments. Après le développement du modèle, la dispersion de l'onde sonore permet l'estimation de la distance de la vocalisation basée sur l'enregistrement d'un seul instrument. Estimer la distance entre l'instrument et les vocalisations de baleines permet d'estimer l'abondance minimale de la baleine franche menacée d'extinction dans le Pacifique Nord.

1. INTRODUCTION

The North Pacific right whale (*Eubalaena japonica*) is a critically endangered baleen whale. There is no reliable estimate for the eastern population, but it probably numbers less than 50 individuals (Clapham *et al.*, 1999). Efforts to study these whales in the eastern Bering Sea have provided visual observations of them since 1996 (Fig. 1) (Goddard and Rugh, 1998); (Moore *et al.*, 2000); (LeDuc *et al.*, 2001); (Tynan *et al.*, 2001). To complement these visual surveys, shipboard acoustic surveys have recorded North Pacific right whale calls in the eastern Bering Sea since 1999 (McDonald and Moore, 2002). In addition to providing the first descriptions of North Pacific right whale calls, the shipboard acoustic surveys provided the baseline acoustics needed to use long-term, autonomous acoustic recorders for passive monitoring of these endangered whales.

Long-term autonomous acoustic recording provides a means for monitoring whale calling activity in poor weather conditions and during periods when ship-based visual and acoustic techniques are either impossible or cost prohibitive (Wiggins, 2003). By recording sound continuously for periods of more than one year, whale seasonal occurrence and minimum population estimates can be made. To do this requires an understanding of the relationship between calls recorded and total number of whales present within a given region. Knowledge of call detection range is critical. How far a call can be detected with an acoustic instrument depends on the characteristics of the call and the acoustic environment. Environmental noise from ships, storms or other calling whales may reduce the call detection range. In addition, acoustic propagation depends upon environmental factors such as water temperature profile and bathymetry. These factors can effectively enhance or decrease call detection range, and may distort call characteristics.

Calls may be distorted by the environment in a range dependent way such that the distorted calls contain information about the caller's location. For example, multi-path arrivals are common in environments where the distance from caller to receiver is less than a few times the water depth. The first arrival of the call at the receiver is from the direct path wave. The next few arrivals are from surface and bottom reflected waves and may interfere with the first and other arrivals. The call will appear at the receiver to be a summation of these arrivals, however, knowledge of the acoustic propagation may allow for the original call to be extracted from the distorted signal. If the sound speed profile, the water depth, and the receiver location are known, then range to and depth of the caller can be calculated.
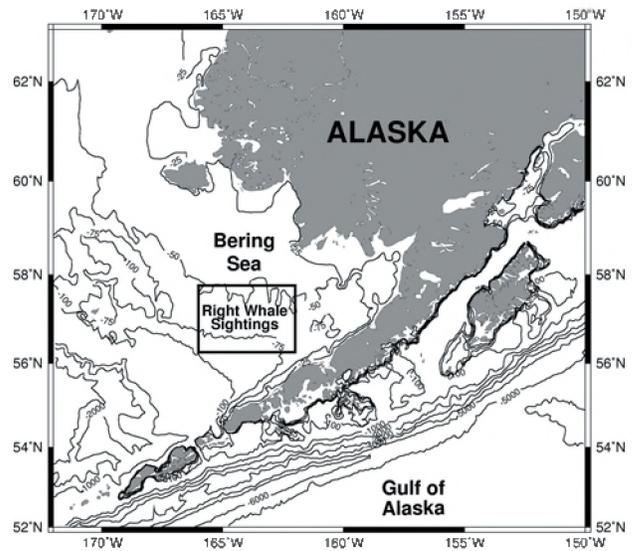
Fig. 1. Eastern Bering Sea. Bathymetric contours are every 25 m for the first 100 m, then every 1000 m. The box outlines where North Pacific right whales have been visually observed each summer since 1996, and acoustically observed since 1999. Bathymetry data from Smith and Sandwell (1997).

In shallow-water waveguides, calls that are more than several water depths in range away may become distorted due to multimode dispersion. Analysis of this distortion can provide an estimate of the distance to the caller. Normal-mode waveguide modeling helps to describe the observed waveform distortions. As the range between caller and receiver is increased, the original call will become increasingly distorted. The relatively shallow and flat continental shelf of the eastern Bering Sea provides an acoustic waveguide environment. In waveguides, the call or source waveform reflects off the seafloor and sea surface and these reflections will constructively and destructively interfere to create multiple mode arrivals and waveform dispersion, where different frequencies of the waveform travel at different velocities. The variation of velocity with frequency allows range estimates between source and receiver to be made for calls that sweep through a band of frequencies. The majority of North Pacific right whale calls upsweep in frequency and will become noticeably dispersed in shallow water after a few kilometers (McDonald and Moore, 2002). Another example showing shallow-water mode dispersion for right whale upswept calls was presented in a sonobuoy localization study in the Bay of Fundy where North Atlantic right whales (*Eubalaena glacialis*) were studied (Laurinolli *et al.*, 2003).

During 2000 to 2002 autonomous acoustic recording packages were deployed in the eastern Bering Sea to investigate the seasonal presence and population of North Pacific right whales. While some individual right whale calls were recorded on multiple instruments allowing for time-difference hyperbolic localization techniques to be

used, many calls were recorded only on one instrument. Normal-mode modeling allows for estimating ranges to these callers and provides information on their calling depth.

2. METHODS

2.1 Acoustic Data

During July 1999 sonobuoys were deployed in the eastern Bering Sea to provide acoustic data in conjunction with a right whale visual survey (McDonald and Moore, 2002). Sonobuoys provide real-time acoustic data using radio telemetry to a support ship where they can be recorded and analyzed with computer software. The DIFAR (DIrectional Frequency Analysis and Recording) sonobuoys used during this survey were configured with the hydrophone sensor at 28 m below the sea surface and provided bearing data in the band from 10 Hz to about 4 kHz. These sonobuoys were often deployed in array configurations at known GPS (global positioning system) coordinates and drift rates were calculated from bearings to the research ship so that caller locations could be calculated using multiple bearings and correlated with visual sightings. Concurrent visual and acoustic observations provided species identification of the caller.

McDonald and Moore (2002) analyzed over 500 North Pacific right whale calls and reported the predominant call type to be an upswept call which has similar characteristics to those reported by Clark (1982) for southern right whales (*Eubalaena australis*). From these sonobuoy data, typical 'up' calls sweep from about 90 Hz to 150 Hz in 0.7s and have sweep rates ranging from 35 to 150 Hz/s, although some of the variability reported in these calls may be caused by waveguide distortions. The acoustic waveform data are transformed into the frequency domain using Fourier transforms and viewed as spectrograms. Spectrograms allow narrow-band signals (such as right whale calls) to be detected above broad-band ocean noise. A single right whale call recorded on four different sonobuoys is shown as spectrograms in Fig. 2. Notice that the call becomes progressively distorted and extended in time with increasing range. These changes in the signal are a result of the waveguide propagation and associated dispersion.

In October 2000, four autonomous acoustic recording packages (ARPs) (Wiggins, 2003) were deployed in the eastern Bering Sea to monitor North Pacific right whales. The ARPs were configured to record continuously with a hydrophone sensor tethered approximately 10 m above the seafloor and with a bandwidth from 5 to 250 Hz. These instruments were deployed in the area where right whales have been observed since 1996, and were placed 60 to 80 km apart in about 70 m water depth (Fig. 3). The array was not configured to provide good localization geometry; however, because propagation in this environment was better than anticipated, there are many cases of multiple

instruments recording the same call. Using GPS instrument deployment locations, individual calls recorded with multiple ARPs were localized with time-difference hyperbolic localization software (Mellinger, 2002). In addition to providing call detection ranges for minimum population estimates, these localizations are used to evaluate the normal-mode range-estimate modeling.
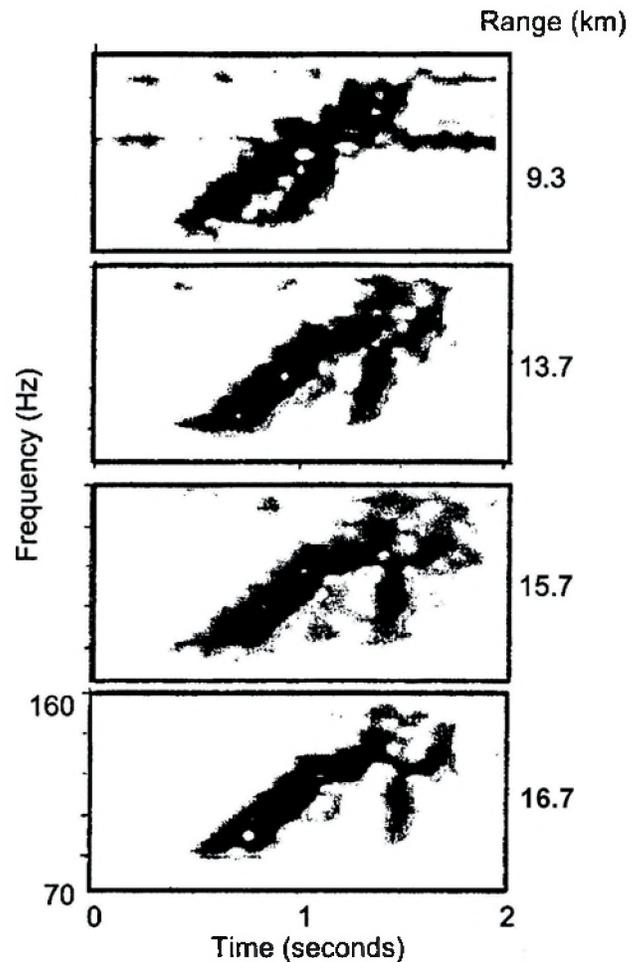
Fig. 2. An example spectrogram of a North Pacific right whale call recorded with four sonobuoys in the eastern Bering Sea. Notice that the call becomes spread-out in time, especially at lower frequencies, for the most distant sonobuoys. The spectral parameters used are 0.5 second FFT length with 87.5% overlap. See Fig. 3 for sonobuoy and whale locations during recordings. (Figure 7 from McDonald and Moore, 2002.)
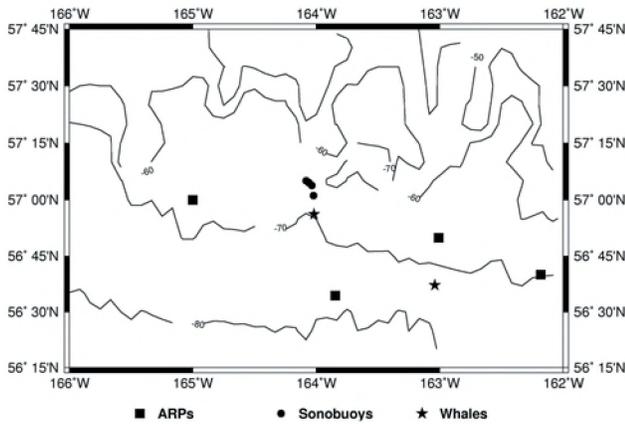
Fig. 3. Bathymetry and instrument locations of 'Right Whale Sighting' box from Figure 1. Autonomous acoustic recording packages (ARPs) are squares, sonobuoy locations during recording spectrograms in Figure 2 are circles, stars are locations of whales for example calls. Bathymetry data from Smith and Sandwell (1997).

2.2 Normal-Mode Modeling

To model sound transmission in shallow water environments, the normal-mode approach is often preferred over the ray-path method because the normal-mode approach provides better computational efficiency at long ranges (>10 times the water depth) and moderate to low frequencies (<500 Hz) (Officer, 1958); (Medwin and Clay, 1998). Shallow water environments can be described as waveguides in which ray paths of plane waves are trapped between two reflecting surfaces (Fig. 4). Normal-mode methods for a simple two-layer ocean model were first developed by Pekeris (1948) and have been used widely to investigate acoustic propagation in shallow water (e.g., Jensen *et al.*, 2000). The normal-mode solution considers all waveguide-trapped ray paths and their combined interference effects.
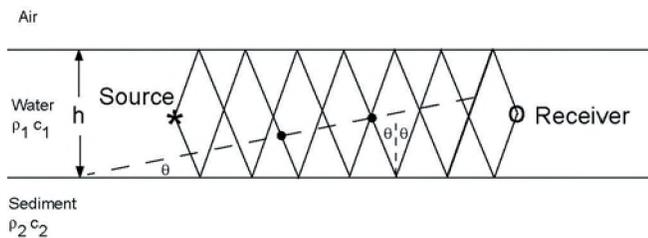
Fig. 4. Schematic of sound transmission in a shallow-water waveguide. Waves will reflect off the sea surface and seafloor boundaries and incur phase changes based on environmental physical properties.

Interference between up going and down going reflecting waves in a waveguide depends on their frequency. At frequencies where the phase difference between the interfering waves is an integer number of 2π, the waves constructively interfere. At all other frequencies the waves interfere out of phase and have negligible contribution at long ranges. For each angle of incidence, θ, of up going and down going waves, there is a set of discrete frequencies that constructively interfere. Each frequency corresponds to a mode and travels at a different velocity along the waveguide. These velocities are the group velocities. For angles of incidence, θ, more grazing than the critical angle ($\theta_c = \sin^{-1}(c_1/c_2)$ where c_1 is the sound speed of the water and c_2 is the sound speed of the sediment), a set of frequencies and group velocities can be calculated for each mode. A plot of these frequencies versus group velocity for each mode are dispersion curves and provide a means of estimating range to dispersed calls.

From Medwin and Clay (1998), the group velocity of the mth mode is

$$u_m = \frac{d\omega}{dk_m} \qquad (1)$$

where, the angular frequency, ω, and horizontal wave number, k_m, are expressed as

$$\omega = \frac{\gamma_m c_1}{\cos(\theta)} \qquad (2)$$

and

$$k_m = \left[\left(\frac{\omega}{c_1} \right)^2 - \gamma_m^2 \right]^{1/2}. \qquad (3)$$

The vertical wave number, γ_m, derived from the mode equation is

$$\gamma_m = \left(\pi(m-1) + \frac{\pi}{2} + \phi \right) \frac{1}{h} \qquad (4)$$

and the phase shift at the seafloor, ϕ, is

$$\phi = \tan^{-1} \left(\frac{\rho_1 c_1 g_2}{\rho_2 c_2 \cos(\theta)} \right) \qquad (5)$$

with g_2, the imaginary part of Snell's Law cosine at the seafloor for $\theta > \theta_c$, expressed as

$$g_2 = \left[\left(\frac{c_2}{c_1} \right)^2 \sin^2(\theta) - 1 \right]^{1/2}. \qquad (6)$$

The dispersion curves are calculated by numerically differentiating Equation (1) after substituting in Equations (2-6) and using a set of incident angles $\theta_c \le \theta \le \pi/2$. The waveguide parameters are water sound speed, c_1, sediment sound speed, c_2, water density, ρ_1, sediment density, ρ_2, and water depth, h.

Which frequencies constructively interfere to produce a dispersion curve is dependent upon the physical parameters of the waveguide. For up going waves, a phase change of π occurs at the sea surface. For down going waves, the phase change at the seafloor boundary (Equation 5) depends on water and seafloor sound speeds and densities, and on the angle of incidence. The depth or thickness of the waveguide also affects what frequencies constructively interfere because it defines the geometry in which the waves reflect and where phase changes take place.

The water depth of the study area is approximately 70 m (Fig. 3). The sound velocity (1470 m/s) and density (1026 kg/m^3) were obtained from Generalized Digital Environmental Model (GDEM) which is based on the US Navy's Master Oceanographic Observation Data Set (MOODS) (Teague *et al.*, 1990). The sound speed and density do not vary much with depth nor season probably because these waters are relatively well mixed, and are considered homogeneous for our purposes. The sediment velocity (1675 m/s) and density (1500 kg/m^3) are based on Deep Sea Drilling Project (DSDP) results from seismic reflection data and core samples in the Bering Sea, albeit off the shelf (Shipboard Scientific Party, 1971).

Using these physical parameters for a simple two layer lossless Pekeris waveguide (homogeneous water layer over a homogeneous half-space of sediment), dispersion curves for the first four modes were calculated and plotted (Fig. 5). For each mode at its cutoff frequency (frequency below which waves are not trapped in the waveguide and attenuate rapidly with distance), the group velocity is the sediment velocity. Above the cutoff frequency, the steep decrease in group velocity with increasing frequency occurs for seismic ground waves. At frequencies above the lowest velocity (Airy frequency or inflection point) of the dispersion curve, a more gentle increase in group velocity with increasing frequency occurs for water waves in the waveguide, and the group velocity approaches the water sound speed at high frequencies. Dispersion for water waves is greatest near the Airy frequency which leads to the greatest distortion in the received call. Also, the group velocities decrease as the mode number increases at a given frequency. This causes higher modes to arrive after lower modes, and the time difference between mode arrivals can be used to estimate the range to the source.
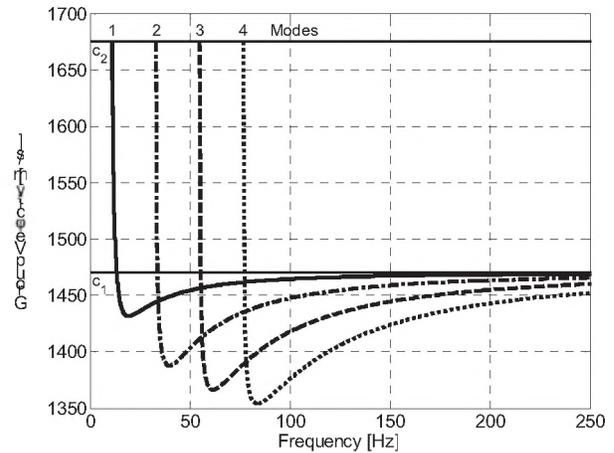
Fig. 5. Dispersion curves calculated from normal-mode modeling for first four modes. At high frequencies, as the frequency increases, the group velocity approaches the water sound speed (1470 m/s). At low frequencies, as the frequency decreases near the cutoff, the group velocity increases rapidly toward the sediment velocity (1675 m/s) and most of the energy is trapped in the sea floor as seismic ground waves.

Waveguide distorted calls can be modeled by applying the dispersion curve to an undistorted synthetic call at a given range. Estimates of modal arrival times from source to receiver are made simply by dividing the range by the velocity (dispersion) curve for each mode at the different frequencies and adding the result to the original synthetic call. If the call sweeps through a band of frequencies, then dispersion will distort the call with the low frequencies traveling slower than the higher frequencies (Figs. 2 and 5). The distortion becomes more pronounced at greater ranges because the energy at each frequency has had more time to travel at a different velocity. Also, higher modes are more dispersed leading to more distortion and increasing modal separation with increased range.

Sensitivity of the group velocity model can be evaluated by changing each environmental parameter independently and comparing the results to the unchanged model. The model is most sensitive to parameter change near the cutoff frequency of each mode, but model change decreases rapidly with increasing frequency. Based on environmental and bathymetric data for the Bering Sea, two extreme values for each parameter are tested: water sound speed, c_1, (1450 and 1490 m/s), sediment sound speed, c_2, (1550 and 1800 m/s), water density, ρ_1, (1025.5 and 1025.5 kg/m^3), sediment density, ρ_2, (1300 and 1700 kg/m^3), and water depth, h, (65 and 75 m) (Teague *et al.*, 1990); (Shipboard Scientific Party, 1971); (Smith and Sandwell, 1997). The change in group velocity models for the first four modes using these parameters are less than 1.5% (c_1), 0.5% (c_2), 0.01% (ρ_1), 0.1% (ρ_2), and 0.5% (h) for frequencies above the Airy frequency.

The interference effects in a waveguide also define how modes will be excited with depth. The depth dependence of the mode excitation is expressed as

$$Z_m(z) = \sin(\gamma_m z) \qquad (7)$$

where z is depth. As above, Equations (4-6) are substituted into Equation (7) and a set of incident angles $\theta_c \leq \theta \leq \pi/2$ are used to solve for various angles and frequencies. As an example, using a 100 Hz source, the normalized sound pressure as a function of depth is calculated from Equation (7) and plotted using the same environmental parameters for the dispersion curves (Fig. 6). A mode zero crossing (node) indicates the depth at which a mode is not excited. For example, a source at 20 m deep will not excite the fourth mode but will fully excite the second mode, however, if the receiver is at 40 m deep, neither the second nor the fourth mode will be received. No modes are excited at the pressure release boundary sea surface. Since the depths of the receivers are known for this study, the relative amplitude of excited modes provides information on the source (calling whale) depth.
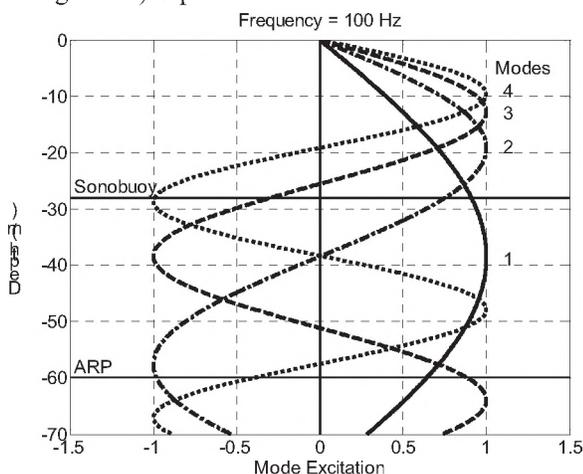
Fig 6. Normalized sound pressure versus depth from normal-mode modeling for first four modes at 100 Hz. No modes are excited at the sea surface which is a pressure release boundary.

3. RESULTS

3.1 Sonobuoy recordings and modeling

Upswept North Pacific right whale calls propagate long ranges as dispersed modes in the shallow and relatively flat waveguide of the eastern Bering Sea. The effects of modal dispersion of a right whale call recorded by four sonobuoys concurrently with shipboard visual observations of the calling whale are shown in Figure 2 (McDonald and Moore, 2002) with the sonobuoy and whale positions plotted in Figure 3. For the closest recording (9.3 km), separation of modes is evident by the gap in power of the spectrogram near 0.7 s at low frequencies. At greater ranges, further separation of modes is apparent and the highest propagating

mode of the farthest recording (16.7 km) has distorted so much that the original gentle upsweep appears near vertical.

To investigate how normal-mode modeling fits the sonobuoy right whale data, an initial synthetic three-part upswept call was used to best fit the first mode of the closest recording. The synthetic received call was calculated using the modal dispersion curves, the known range, and the initial synthetic call. Synthetic received calls are overlaid as thin black lines on spectrograms of the recorded calls for the close and far sonobuoys recordings (Fig. 7). The match between the modeled calls and the recorded calls for both the closest and farthest sonobuoy is good. However, notice in the recorded data, the third mode appears minimally excited compared to modes one, two and four.
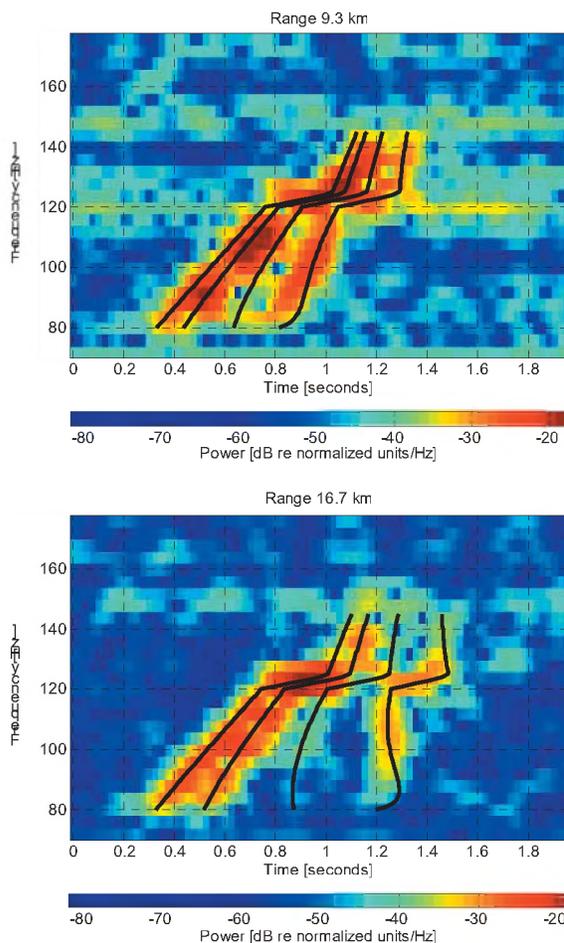
Fig. 7. Spectrograms of closest and farthest sonobuoy right whale recordings (from Fig. 2) overlaid with normal-mode modeling synthetic received call. Note the good match between the model and the data, and that the third mode is minimally excited in the recorded data.

The lack of third mode excitation in the sonobuoy example can be attributed to source and/or receiver depth. The depth of the calling whale is unknown, but the sonobuoy hydrophone is suspended from a sea surface float at approximately 28 m. This depth is near a node for the third

mode on the excitation plot (Fig. 6). For this waveguide model and for sources around 100 Hz, the third mode will be minimally excited for receivers and sources near 25 m deep. A synthetic spectrogram illustrating this point is shown in Figure 8. The relative received power has been combined with the synthetic received call by choosing a source and a receiver depth and evaluating the contributions from the modal excitations at these depths over the frequencies of the call. The power for the synthetic spectrogram is simply the product of the absolute values of the modal excitations at the source and receiver depths. For example, if the source and receiver were placed at 40 m depth, then for a source at 100 Hz the first and third modes would have powers near one (0 dB), whereas the second and fourth modes would have powers near zero (large negative dB) (Fig. 6). In this example, the source was chosen to be at 25 m and the receiver at 28 m. From the synthetic spectrogram it is apparent that the third mode is minimally excited compared to the other three modes by comparing their relative power (Fig. 8), similar to the recorded data.
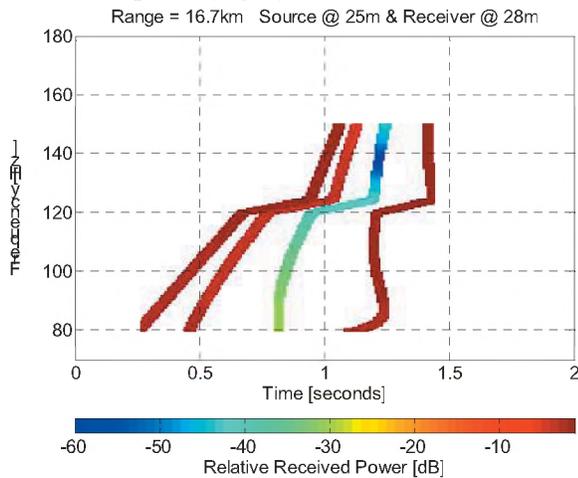
Fig. 8. Synthetic spectrogram for farthest sonobuoy example (16.7 km range). The synthetic received call from Figure 7 is combined with the relative received power calculated as the product of the absolute values of the modal excitation over the call frequency range and for source depth at 25 m and receiver depth at 28 m. The third mode is minimally excited compared to modes one, two and four, which is consistent with the sonobuoy recordings.

3.2 ARP recordings and modeling

The seafloor ARPs also recorded North Pacific right whale calls and can be modeled in a similar fashion to the sonobuoy recordings to verify the normal-mode modeling. An individual call recorded on three or more instruments (to allow localization) is required to compare the modeling results to the recorded calls. Once the model has been verified, it can be used to estimate range to calling whales recorded on single instruments.

An example right whale call is overlaid with synthetic received calls for the two most eastern ARPs (23 km and 56

km ranges) in Fig 9. The ranges were calculated using a hyperbolic localization technique, 1470 m/s water sound speed, and arrival times at 150 Hz. The initial synthetic call is an upsweep from 95 Hz to 170 Hz with a sweep rate of 120 Hz/s. This is a simplification of the actual call, but incorporates the range-dependent varying section of the upsweep. The mostly-constant tonal at the beginning of the call below 100 Hz does not add much range information and is omitted. The modal arrival times are calculated using the same sound speed, density and waveguide thickness parameters that were used for the sonobuoy modeling. The model fits the data well for both ranges including modes five and six on the closest ARP recording (23 km range).
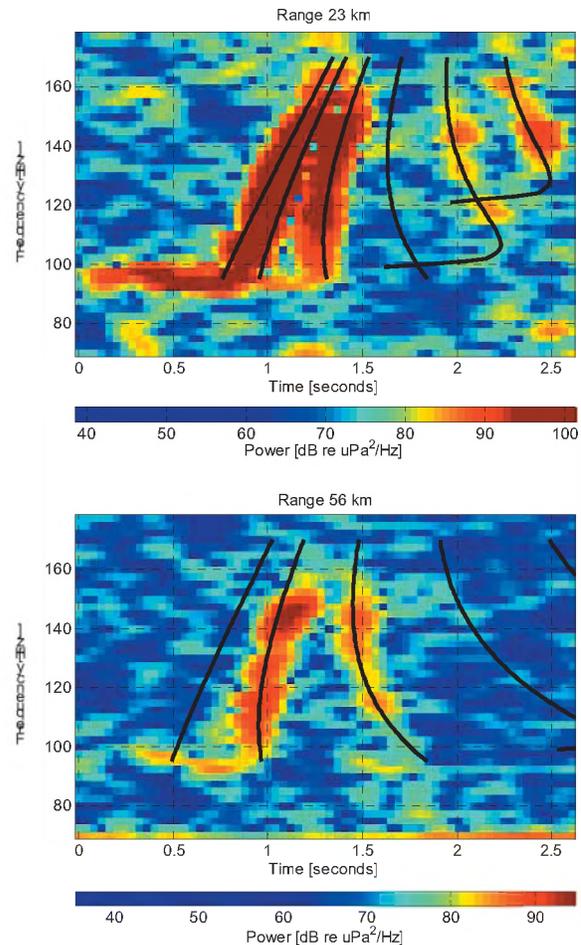
Fig 9. Spectrograms of right whale recordings from seafloor ARPs overlaid with synthetic received calls from normal-mode modeling. The model fits the data well for both ranges including modes five and six on the closest ARP (23 km range). Note mode four is not excited in both spectrograms and only modes two and three are above the background noise at the most distant ARP (56 km range).

On both the closest and farthest ARP recordings, the fourth mode is not excited above the background noise. The receiver hydrophone for the ARP is about 10 m off of the seafloor or at about 60 m depth which is near a zero-

crossing node for the fourth mode, but still some excitation should be present. Perhaps for this call, the whale is near the 20 m deep node for mode four, so the combined modal excitation contribution is negligible (Fig. 6). This explanation of source depth is also valid for why mode two is more strongly excited than mode three and why mode one is not observed at 56 km range (Fig 9). A synthetic spectrogram of the farthest ARP recording was produced similarly to the sonobuoy synthetic spectrogram, but with a source depth of 18 m and receiver depth of 60 m (Fig. 10). The synthetic spectrogram agrees with relative power of the recorded data where modes two and three are the strongest and mode one is weaker.
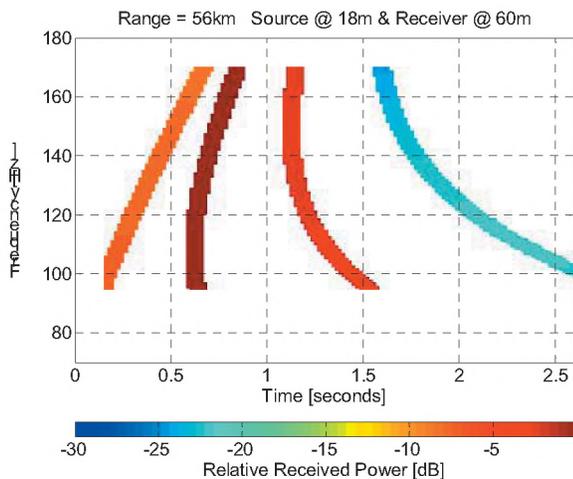
Fig 10. Synthetic spectrogram for the most distant ARP example (56 km range). The synthetic received call from Figure 9 is combined with the relative received power calculated as per Figure 8, but with the source at 18 m and the receiver at 60 m depth. The strongest arrival is mode two followed by mode three then mode one, which is not visible above the noise in the ARP recording.

4. DISCUSSION

Examples of North Pacific right whale calls distorting into dispersed modes with increasing range in the shallow-water waveguide eastern Bering Sea have been shown to contain source-receiver range and source calling depth information. With normal-mode modeling, estimating range to calling whales from single instruments is possible and can add to hyperbolic localizations from multiple instrument arrays. By understanding the detection range, whale abundance estimates for long term acoustic recordings can be improved (Buckland *et al.*, 1993).

The group velocity dispersion model presented has been qualitatively and successfully fit (forward modeled) to recorded data for different right whale calls, different types of instruments, and during different time periods. Preliminary work with downswept minke (*Balaenoptera acutorostrata*) and fin (*Balaenoptera physalus*) whale calls

recorded on the ARPs also has shown that this model fits well for these localized calls. However, minke and fin whale calls sweep through lower frequencies (80 to 50 Hz, and 35 to 15 Hz, respectively) than right whales, and only excite the first few modes because these frequencies are below the cutoff frequencies of higher modes.

How well these simple models fit the recorded calls depends upon the environmental parameters used in the models. The bathymetry of the eastern Bering Sea shelf changes only a few meters over the propagation path (10's of km) from caller to receiver (Fig. 3) (Marlow *et al.*, 1999). This amount of change has minimal impact on group velocities at frequencies above the Airy frequency, and provides consistent results between the recordings at various ranges. The sediment sound speed had the greatest uncertainty, but was adjusted from 1600 m/s to 1675 m/s to best fit the observed data for modes four and above. This adjustment had minimal effect on lower modes and on shorter range (<20 km) modeled calls. Uncertainties in water sound speed have the greatest impact on group velocity, however, the sound speed is almost constant with depth (+/- 5 m/s) during the late summer and early fall when the calls were recorded (Teague *et al.*, 1990). Low variability in the model parameters allows the same model to be used for the sonobuoy and ARP data.

While qualitative curve fitting to the data works well, a method more capable of automation could be developed for large data sets. One approach would be to choose one frequency of a right whale call and measure the time difference between modes at that frequency and estimate range from caller to receiver by using the modeled group velocities for those modes. For example,

$$R = \frac{u_i u_j}{|u_i - u_j|} |t_i - t_j| \qquad (8)$$

where R is range, u and t are respectively the group velocities and arrival times for modes i and j. It is best to choose a low frequency, close to the Airy frequency of the dispersion curve, because group velocities are slowest and the time difference will be greatest there. With large time differences, arrival time picking errors can be minimized. Also, because higher order modes are more dispersed, choosing these modes would minimize arrival time picking and range estimate errors. However, if only one frequency is used, then it is essential to know which modes are being excited and recorded. For example, if the time difference at 100 Hz between modes two and three for the ARP recorded call at 56 km range (Fig. 9) was presumed to be for modes one and two or for modes three and four, then the range would have been incorrectly estimated to be 95 km or 37 km, respectively. To prevent such gross errors, it would be best to analyze the full-sweep spectrogram because it requires that the dispersed modes fit for all frequencies in

the call band. Match field processing or inverse modeling techniques then could be applied to a set of mode time-frequency data to solve for range and statistical error estimates.

REFERENCES

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. R. Borchers, and L. Thomas 1993. Distance sampling: estimating abudance of biological populations. London, Chapman and Hall.

Clapham, P. J., S. B. Young, and R. L. Brownell, Jr. 1999. Baleen whales: conservation issues and the status of the most endangered populations. *Mammal Review* **29**: 35-60.

Clark, C. W. 1982. The acoustic repertoire of the Southern right whale, a qualitative analysis. *Anim. Behav.* **30**: 1060-1071.

Goddard, P. D. and D. J. Rugh 1998. A group of right whales seen in the Bering Sea in July 1996. *Mar. Mammal Sci.* **14**(2): 344-349.

Jensen, E. B., W. A. Kuperman, M. B. Porter, and H. Schmidt 2000. Computational Ocean Acoustics. New York, Springer-Verlag.

Laurinolli, M. H., A. E. Hay, F. Desharnais and C. T. Taggart 2003. Localization of North Atlantic right whale sounds in the Bay of Fundy using a sonobuoy array. *Mar. Mammal Sci.* **19**(4): 708-723.

LeDuc, R. G., W. L. Perryman, J. W. J. Gilpatrick, J. Hyde, C. Stinchcomb, J. V. Carretta and R. L. J. Brownell 2001. A note on recent surveys for right whales in the southeastern Bering Sea. *J. Cetacean Res. Manage.*(Spec. Issue; 2): 287-289.

Marlow, M. S., A. J. Stevenson, H. Chezar, and R. A. McConnaughey 1999. Tidally generated sea-floor lineations in Bristrol Bay, Alaska, USA. *Geo-Marine Letters.* **19**: 219-226.

McDonald, M. A. and S. E. Moore 2002. Calls recorded from North Pacific right whales (*Eubalaena japonica*) in the eastern Bering Sea. *J. Cetacean Res. Manage.* **4**(3): 261-266.

Medwin, H. and C. S. Clay 1998. Fundamentals of Acoustical Oceanography. San Diego, Academic Press.

Mellinger, D. K. 2002. Ishmael 1.0 User's Guide, NOAA Technical Memorandum OAR PMEL-120, available from NOAA/PMEL, 7600 Sand Point Way NE, Seattle, WA 98115-6349.

Moore, S. E., J. M. Waite, L. L. Mazzuca and R. C. Hobbs 2000. Mysticete whale adundance and observations on prey association on the central Bering Sea shelf. *J. Cetacean Res. Manage.* **2**(3): 227-234.

Officer, C. B. 1958. Introduction to the theory of sound transmission. New York, McGraw-Hill Book Company, Inc.

Pekeris, C. L. 1948. Theory of propagation of explosive sound in shallow water. New York, NY, Geological Society of America.

Shipboard Scientific Party 1971. Site 184 and Site 185, University of California, Scripps Institution of Oceanography.

Smith, W. H. F. and D. T. Sandwell 1997. Global seafloor topography from satellite altimetry and ship depth soundings. *Science* **277**: 1957-1962.

Teague, W. J., M. J. Carron, and P. J. Hogan 1990. A comparision between the Generalized Digital Environmental Model and Levitus climatologies. *J. Geophys. Res.* **95**(C5): 7167-7183.

Tynan, C. T., D. P. DeMaster and W. P. Peterson 2001. Endangered right whales on the southeastern Bering shelf. *Science* **294**: 1894.

Wiggins, S. M. 2003. Autonomous Acoustic Recording Packages (ARPs) for long-term monitoring of whale sounds. *Mar. Technol. Soc. Journal* **37**(2): 13-22.

ACKNOWLEDGEMENTS

DIFAR HYDROPHONE USAGE IN WHALE RESEARCH

Mark A. McDonald

WhaleAcoustics, 11430 Rist Canyon Road, Bellvue, CO 80512, USA, www.whaleacoustics.com

ABSTRACT

DIrectional Frequency Analysis and Recording (DIFAR) sonobuoys have been used by the Navy for many decades, providing magnetic bearings to low frequency (less than 4 kHz) sound sources from a single sensor. Computing advances have made this acoustic sensor technology increasingly easy to use and more powerful. The information presented here is intended to help new users determine when DIFAR sensors are or are not appropriate in whale acoustics research. Acoustic detection ranges for baleen whales average near 20 km but vary from 5 to 100 km depending on conditions. Radio reception range from DIFAR sonobuoys to a typical research vessel averages 18 km with an omni directional antenna on the ship and standard antenna on the sonobuoy. DIFAR bearing accuracy is analyzed for a set of whale calls where the track of the whale was well known. Bearings from the DIFAR sensor were found to have a standard deviation of 2.1 degrees. Systematic error and magnetic deviation can be removed using DIFAR bearings to the sound of the research vessel at a known location. A DIFAR sensor array requires fewer sensors than a conventional hydrophone array and sometimes provides more accurate source locations than the "time of arrival" hyperbolic methods used with conventional hydrophones. Continuous sounds such as ships are more easily localized with DIFAR sensors than with conventional hydrophones, because it is often difficult to find transient features upon which to estimate the time differences needed for hyperbolic fixing with a conventional hydrophone array. DIFAR hydrophone systems are well suited to right, blue, minke, fin and other baleen whale calls, as well as numerous other sound sources including ships.

RÉSUMÉ

Les bouées acoustiques directionnelles DIFAR sont utilisées par la marine depuis plusieurs décennies, fournissant des relèvements magnétiques provenant d'un détecteur unique pour des sources sonores à basse fréquence (moins de 4 kHz). Les avancées computationnels ont fait de cette technologie un outil puissant et simple à utiliser. L'information présentée dans le présent article a pour but d'aider les nouveaux utilisateurs à déterminer quand les détecteurs DIFAR sont ou ne sont pas appropriés dans l'étude acoustique des baleines. La portée de détection acoustique pour baleines mysticètes atteint une moyenne voisine de 20 km mais varie de 5 à 100 km dépendant des conditions. La portée de la réception radio des bouées acoustiques à un navire de recherche typique atteint une moyenne d'environ 18 km avec une antenne omni directionnelle sur le bateau et une antenne standard sur la bouée acoustique. La précision du relèvement DIFAR est analysée pour un certain nombre de vocalisations de baleine où le parcours de la baleine est bien connu. Les relèvements provenant du détecteur DIFAR ont démontré une déviation standard de 2.1 degrés. Les erreurs systématiques et la déviation magnétique peuvent être corrigées en utilisant les relèvements DIFAR vers le son d'un navire de recherche qui a une position connue. Un réseau de détecteurs DIFAR a besoin de moins de détecteurs qu'un réseau d'hydrophones conventionnel et procure parfois une localisation de la source plus précise que la méthode hyperbolique des "temps d'arrivée" utilisée avec les hydrophones conventionnels. Les sons continus, comme ceux des bateaux, sont plus facile à localiser avec le détecteur DIFAR qu'avec les hydrophones conventionnels parce qu'il est souvent difficile de trouver des signaux transitoires permettant d'estimer les différences temporelles nécessaires pour le positionnement hyperbolique avec un réseau d'hydrophones conventionnel. Les systèmes d'hydrophones DIFAR conviennent aux vocalisations de baleines franches, bleues, de petits rorquals, rorquals communs et autres mysticètes, aussi bien qu'un bon nombre d'autres sons incluant les navires.

1. INTRODUCTION

Acoustic surveying for whales is becoming commonplace, either in conjunction with shipboard visual surveys or land based visual surveys or independently (Širović *et al.,* in press; Laurinolli *et al.,* 2003; McDonald and Moore, 2002; Noad and Cato, 2001; Clark and Ellison, 2000; Norris *et al.,* 1999). The tools for these acoustic studies include shore cabled hydrophones, autonomous hydrophone recorders, towed hydrophones, drifting sonobuoys and moored sonobuoys. Acoustic surveys can be used for line transect, relative, minimum and potentially even for absolute abundance estimation. In some cases acoustics are used to locate whales of a given species for biopsy, photo-id or tagging or to document the presence of migrating whales in locations which may not have any visual survey data.

For whale species which produce most of their acoustic calls above 200 Hz, conventional towed hydrophones work well. If a ship is also conducting a visual line transect survey with an emphasis on covering the greatest distance and the species of primary interest produce calls above about 200 Hz, the large number of sonobuoy deployments required is more expensive and less efficient than using a towed hydrophone array. For the species which call below 200 Hz, sonobuoys and fixed hydrophones have significant advantages over towed hydrophones, being more distant from the typically noisy research vessel and avoiding flow noise as it is costly to slow or stop the research vessel to better hear on towed hydrophones. A conventional hydrophone provides no directional information to localize low frequency acoustic sound sources unless it is used in an array the length of which is determined primarily by the frequency of the whale calls of interest and then the whale must call multiple times to break the left-right ambiguity inherent in direction finding with a single array.

A DIFAR sensor makes use of particle motion in the sea water due to acoustic wave propagation, allowing for a compact sensor which indicates horizontal direction to each sound source present (D'Spain, 1994; D'Spain *et al.,* 1991). DIFAR hydrophones are sensitive to overloading from motion and thus have not been suitable for use on a ship hull or in towed arrays. In fixed configurations, they typically must be shielded from current flow by some form of shroud.

The sensor portion of a DIFAR sonobuoy consists of two orthogonal horizontal directional acoustic particle velocity sensors, a magnetic compass, and an omni directional pressure sensor. Within conventional DIFAR sonobuoys the magnetic North-South (NS) and magnetic East-West (EW) components of particle motion are computed by the sensor electronics at the hydrophone, the three signals including pressure are multiplexed and transmitted by radio. In the case of autonomous recorders or dipping hydrophones the three data sets can be recorded separately without multiplexing. In a type 53 sonobuoy the frequency response begins to rolloff at about 2 KHz, but not rapidly, such that sufficient response remains to about 4 kHz if the sound source is relatively loud.

A disadvantage of a DIFAR sensor when compared to ordinary hydrophones is that it requires three times the data bandwidth, with all three of the output channels, pressure, East-West particle motion and North-South particle motion being required to compute an unambiguous bearing (D'Spain, 1994). DIFAR sonobuoys of type AN/SSQ53B, AN/SSQ53D and AN/SSQ53E were used in the work presented here, the author having deployed nearly 500 of these in the course of various whale research projects. In a type 53 sonobuoy, the useful bandwidth of about 4 kHz takes up nearly 20 kHz of bandwidth after the analog multiplexing done by the electronics built in to the sensor head.

2. PROCESSING AND PERFORMANCE

2.1. Demultiplexing and Display

Commercial software from GreeneRidge Sciences Inc. was used to process raw DIFAR sonobuoy signals into three channels, 1) east-west particle motion, 2) north-south particle motion and 3) omni-directional pressure. Direction finding theory and methods for DIFAR sonobuoy processing are discussed in the published literature (D'Spain *et al.,* 1991; D'Spain *et al.,* 1992; D'Spain *et al.,* 1994). A MATLAB program was written based on the published theory to compute bearings to sound sources.

Processing speed for demultiplexing and bearing computation is faster than real time, although applications used to date always use a human operator selecting segments of data from a spectrogram and keeping each calling animal tracked on a plot or chart. Prior to about 1992 DIFAR processing was done in hardware rather than software, making processing more expensive and less flexible.

A typical DIFAR blue whale recording is shown in Figure 1, illustrating overlapping whale calls and ship noise.
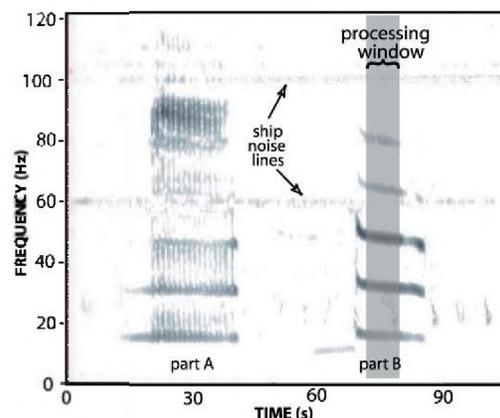
Figure 1. **This spectrogram shows a Northeastern Pacific blue whale call which is used to illustrate bearing processing with multiple sound sources.**

The display options for illustrating sound source bearings from sonobuoy data are nearly endless given the three independent variables, magnetic bearing, frequency and some measure of energy over time. The work presented here uses an averaged output for a given duration of data plotted as frequency versus azimuth (Figure 2). The sound source bearings are picked from the plot with a cursor.
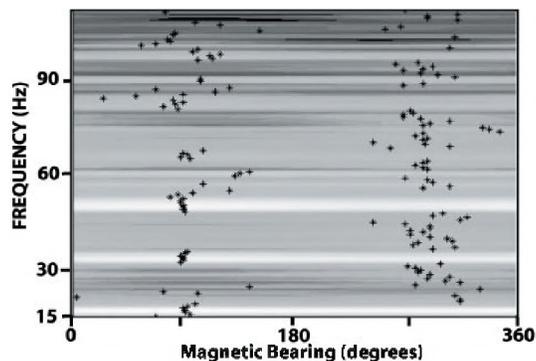
Figure 2. Bearing plot for six seconds of data containing a blue whale "B" call, as shown in Figure 1. Bearing is seen as high energy at the frequency bands of the sound source observed in the spectrogram. The asterisks mark the highest energy in each frequency bin. Lighter color indicates higher energy. This blue whale call is found at 95 degrees. The energy near 290 degrees is from the research ship.

2.2. Bearing Accuracy

In October of 1997 sonobuoy recordings of blue whales were collected during a whale photo-id cruise. A goal of this cruise was to acoustically record and genetically sample blue whales to examine sex bias in calling behavior, requiring localization of each acoustic call (McDonald *et al.*, 2001). A whale track was determined by recording the GPS position of the final surfacing of each surface sequence from a small boat following the whale. Whale positions at the time of each call are interpolated between surfacing's. Only one whale track was used for the analysis presented here, that being whale number one in McDonald *et al.* (2001). Overlapping calls from multiple animals always resulted in two distinct correct bearings, rather than a weighted average bearing between the two whales, which might have been supposed from theory.

DIFAR sonobuoy bearings are compared to bearings computed using GPS coordinates in Figure 3.
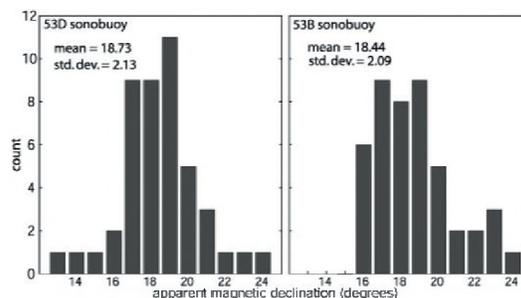
Figure 3. Difference between sonobuoy bearing and GPS bearing are plotted as histograms uncorrected for magnetic declination for two different sonobuoys, one type 53D and one type 53B. These data are for blue whale type "A" and "B" calls. The navigation chart for this area indicates the magnetic declination to be 17 degrees with significant local variability.

These calls were recorded at ranges from 3 km to 8 km. Short range calls were discarded because the whale position errors translate to increasingly large bearing errors at short ranges, these errors becoming potentially greater than the DIFAR bearing errors. One standard deviation of these data is 2.1 degrees, well within the sonobuoy specification requirement for a maximum error of 10 degrees. In this case the different model sonobuoys had very similar mean values (18.7 and 18.4) suggesting the compasses were either correct or had very nearly the same error. The two sonobuoys used were different models, manufacturers and vintages, so it is unlikely there was a common error. The standard deviation being acceptably small, methods of improved processing have not been pursued though more optimal processing or bearing picking algorithms may be possible. Bias error may be related to sensor construction (i.e. compass not mounted accurately) and/or to uncertainty in the actual deviation of the earth's magnetic field from true north.

2.3. Sonobuoy Radio Range

Production type 53 sonobuoys use a one watt VHF radio transmitter and an antenna only about 0.5 meters above sea level at its top. Radio frequencies are selectable between 136 and 172 MHz. Commercial VHF radios intended primarily for voice communication are typically not acceptable for sonobuoy work because the frequency response of the audio sections in these are limited to the band needed for intelligible voice communication only. GrenneRidge Sciences provided modified ICOM commercial radios used for this work. Receiver sensitivity is not though to be a primary factor in determining the working radio range of these sonobuoy systems.

Experience tells us the VHF radio range from these buoys is not determined strictly by line of sight between the two antennas as even the average radio ranges for a 3 dBi antenna are well beyond line of sight. Radio ranges are plotted in Figure 4 for two different cruises. Note that in

each case there are occasional ranges out to 24 nautical miles, about twice the average.

Experience suggests the greatest factor in radio reception range is atmospheric conditions, the detection ranges typically being similar on a given day and often changing when the weather changes. This phenomenon is well known to VHF radio hobbyists. Good conditions are most often thought to be caused by tropospheric enhancement, often associated with temperature inversions (Pocock, 1992). Equally important is receiving antenna gain, although practicality often dictates using a relatively low gain (3 dBi) omni directional antenna which allows maneuvering the vessel without the need to rotate a directional antenna. These low gain antennas also stand up well to wind and icing conditions.
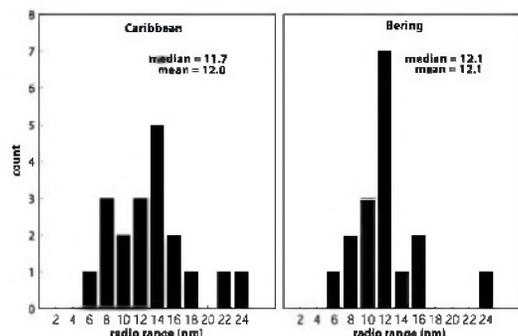
Figure 4. Radio reception distances from DIFAR sonobuoys are plotted for two cruise legs, one in the southern Caribbean and one in the Bering Sea. Antenna heights were 61 ft. (18.6 m) in the Bering Sea and 85 ft. (25.9 m) in the Caribbean. The same 3 dBi gain antenna was used on both cruises and average reception range was 12 nautical miles (22 km).

Comparison of an omni antenna with 3 dBi gain against a YAGI antenna with 12 dBi gain, typically results in more or less a doubling of effective range, assuming the YAGI is correctly pointed at the sonobuoy. The least important factor appears to be sea state or swell height as long as sea state is below 6. At or above sea state 6, it appears the sonobuoy suspension no longer functions well and buoys have a high failure rate in addition to much higher noise levels.

3. APPLICATIONS

3.1. Detection Ranges for Baleen Whales

Detection ranges vary for many reasons including 1) ambient noise due to ships, ice or sea state, 2) acoustic propagation being relatively good or bad, good typically because of an isothermal surface layer creating a sound channel with both receiver and whale in it, or a flat bottom shallow water sound channel or bad because of irregular seafloor bathymetry or a shadowing sound speed profile and 3) the source level of the whale calls. Listed in Table 1 are observed detection ranges with corresponding estimates of ambient noise level and propagation environment for each case.

There are descriptions of detections of baleen whale calls at many hundreds of kilometers range (Charif, et al., 2001; Stafford et al. 1998), but often these use hydrophone arrays with substantial gain and/or are in the deep sound channel and are thus not applicable comparisons for hydrophone or single sonobuoy recordings.

species	location	Range (km)	ambient noise	propagation	References
humpback	Caribbean	50 +	moderate	good, surface sound channel	Swartz et al., 2003; McDonald et al. 2000
right	Bering	50 +	low to mod.	excellent, shallow water wave guide	McDonald and Moore, 2002; Wiggins et al., this issue
right	off Cape Cod	5-10	high, shipping	poor, rugged bathymetry	IFAW, 2001; Doug Gillespie, Pers. comm.
blue	NE Pacific	20	moderate	moderate to poor, shadowing sound speed profile	McDonald et al., 2001; unpublished authors data
blue,	Antarctic	60-100	low	moderate, surface trapped sound speed profile	Širović et al., in press; unpublished authors data
fin,	NE Pacific	20	moderate	moderate to poor, shadowing sound speed profile	McDonald and Fox, 1999
sperm, male	N. Pacific	30-40	moderate	moderate, deep sound source	Barlow and Taylor, 1998
Sperm, female	N. Pacific	5-10	moderate	Moderate, deep sound source	Barlow and Taylor, 1998

Table 1. Acoustic detection ranges for various whale species based on observations, noting qualitatively both noise environment and propagation. Some estimates are based on hydrophones other than DIFAR sonobuoys. Beam steering gain from DIFAR sensors was not used in generating this table, but could improve detection ranges beyond the values given here. Source levels and call frequencies are not tabulated, but play an important role.

3.2.Localization

Hyperbolic fixing depends on finding the arrival time difference for a whale call between two or more hydrophones to solve for a location. To localize a whale call with hyperbolic fixing requires three sonobuoys or hydrophones in a good geometry while only two DIFAR sensors in a good geometry are required for localization. This important distinction is often critical to obtaining a good call location.

In the case of the blue whale calls presented in McDonald *et al.* (2001), both hyperbolic and DIFAR bearing localization was applied. The multi-path environment combined with the long duration of the blue whale calls resulted in average time difference errors of about one second while the time difference between array elements was only a few seconds. While these results are not necessarily typical of all arrays or all types of whale calls, it does demonstrate a significant advantage with DIFAR sensors in many baleen whale localization arrays. If even only one DIFAR sensor is employed in a two hydrophone array, a call location can be determined, given a good array geometry.

3.3.Sperm Whales

DIFAR sonobuoys have sufficient frequency response to about 4 kHz such that DIFAR localization works well for the lower frequency sounds of killer, and pilot whales and would appear to adequately record the lower frequency sounds of sperm whales.

Very short duration whale calls such as sperm whale clicks have never produced good bearings on DIFAR sonobuoys in the experience of the author. The reasons are unclear, but it may have to do with either the short duration of the signal or with the fact these whales are often producing their clicks at depths comparable to the horizontal ranges, resulting in a significant vertical angle to the incoming acoustic energy.

3.4.Directivity Index

Beam steering of DIFAR sensors is a simple matter in software and potentially provides over 4 dB of directivity index gain from the resulting cardioid beam pattern. While beam steering of sonobuoys is undoubtedly useful in some situations, it would appear to require a level of adaptive processing beyond that which has been used to date in whale research.

4. SUMMARY

DIFAR sonobuoys are not the ideal tool for every whale acoustics research question, but are irreplaceable in certain applications. An example of a near ideal application would be locating right whales in the Bering Sea for photo-id, biopsy or tagging studies (McDonald and Moore, 2002).

Right whale calls are mostly below 200 Hz where towed array performance suffers from flow noise and ship noise. The propagation environment in the Bering Sea allows long distance reception and mode dispersion allows range estimation from a single hydrophone (Wiggins *et al.*, this issue). Because visual searching stops for darkness and the ship stops also, a single sonobuoy stays within radio range, and it has often been possible to locate calling animals acoustically during the night such that the ship can plan to arrive in the vicinity of the whale or whales the following morning.

Bering Sea right whale calls occur infrequently and often in clusters from widely separated counter-calling animals. A single towed array would not be suitable for locating these animals because of the inherent left-right ambiguity. A double towed array would add expense and logistical difficulty.

An intermediate application might be the survey of humpback whales in the Southern Carribean (Swartz *et al.*, 2003), where either a towed array or DIFAR sonobuoys could provide a good acoustic survey with similar logistical effort and cost. A clear advantage goes to towed arrays for sperm whale surveys.

When the goals of a whale acoustics research project are clearly defined, the information presented here should help the potential DIFAR user compare the logistical effort of using DIFAR sonobuoys versus using some other acoustic method such as a towed array system. The performance results presented here for DIFAR sonobuoys and DIFAR sensors provide guidelines for what can be accomplished with a given effort for a number of different species of whales.

5. REFERENCES

Barlow, Jay and Barbara L. Taylor. Preliminary abundance of sperm whales in the northeastern temperate Pacific estimated from a combined visual and acoustic survey. International Whaling Commission, Scientific Committee document SC/50/CAWS20, 19p. 1998.

Charif, R.A., Clapham, P.J. and C.W. Clark. Acoustic detections of singing humpback whales in deep waters off the British Isles. *Mar. Mamm. Sci.* 17(4):751-768. 2001.

Clark, C.W. and W.T. Ellison. Calibration and comparison of the acoustic location methods used during the spring migration of the bowhead whale, *Balaena mysticetus*, off Pt. Barrow, Alaska, 1984-1993. *J. Acoust. Soc. Am*, 107(6), 2000, p.3509-3517. 2000.

Stafford, K.M., C.G. Fox, and D.S. Clark, Long-range detection and localization of blue whale calls in the northeast Pacific Ocean. *J. Acoust. Soc. Am.*, 104(6), 3616-3625, 1998.

D'Spain, G.L., Relationship of Underwater Acoustic Intensity Measurements to Beamforming, *Canadian Acoustics*, 22 (3), 157-158, 1994.

D'Spain, G.L., W.S. Hodgkiss, and G.L. Edmonds, Energetics of the deep ocean's infrasonic sound field, *J. Acoust. Soc. Am.*, 89(3), 1134-1158, 1991.

D'Spain, G.L., W.S. Hodgkiss, G.L. Edmunds, J.C. Nickles, F.H. Fisher, and R.A. Harris, Initial Analysis of the data from the Vertical DIFAR Array, in Mastering the Oceans through Technology (OCEANS 92), pp. 346-351, I.E.E.E., Newport, Rhode Island, 1992.

IFAW, Report of the Workshop on Right Whale Acoustics: Practical Applications in Conservation. (eds. Gillespie, D., and Leaper,R.) International Fund for Animal Welfare, Yarmouth Port, MA. 27pp., 2001.

Pocock, E., The weather that brings VHF DX, pp. 21-26 in "Beyond Line of Sight: A history of VHF Propagation from the pages of QST", American Radio Relay League, Newington, CT., 1992

Swartz, S.L., Cole, T., McDonald, M.A., Hildebrand, J.A., Oleson, E.M., Martinez, A., Clapham, P.J., Barlow, J. and Jones, M.L., Acoustic and Visual Survey of Humpback Whale *(Megaptera Novaeangliae)* Distribution in the eastern and Southeastern Caribbean Sea. *Caribbean Journal of Science*, **39**(2), pp. 195-208, 2003.

Laurinolli, M. H., A. E. Hay, et al.. Localization of North Atlantic right whale sounds in the Bay of Fundy using a sonobuoy array. *Mar. Mammal Sci.* **19**(4): 708-723, 2003.

McDonald, M. A. and S. E. Moore, Calls recorded from North Pacific right whales (*Eubalaena japonica*) in the eastern Bering Sea, *J. Cetacean Res. Manage.* **4**(3): 261-266, 2002.

McDonald, M. A., Calambokidis, J., Teranishi, A. M. and Hildebrand, J.A. The Acoustic Calls of Blue Whales off California with Gender Data, *J. Acous. Soc. Am.*, **109**(4), pp. 1728-1735, 2001.

McDonald, M.A., Oleson, E. M., and Hildebrand, J.A., Windwards 2000 Acoustic Cruise Report, NOAA Technical Memorandum NMFS-SEFSC-441, 32 pp., U.S. Dept. of Commerce, May 2000.

Norris, T. F., Mc Donald M. and Barlow J., Acoustic detections of singing humpback whales (*Megaptera novaeangliae*) in the eastern North Pacific during their northbound migration. *J. Acoust. Soc. Am.* **106**(1): 506-514, 1998.

Noad, M.J., and Cato, A combined acoustic and visual survey of humpbacks off southeast Queensland, *Memoirs of the Queensland Museum* **47**(2)145-161. 2001.

Širović, A, Hildebrand· J. A., Wiggins, S.M., McDonald· M.A., Moore· S.E., and Thiele, D. Seasonality of blue and fin whale calls west of the Antarctic Peninsula, Deep Sea Research II, in press.

Wiggins, S.M., McDonald, M.A., Munger, L.A., Hildebrand, J.A. and Moore, S.E. Waveguide propagation allows range-dependent estimates for North Pacific right whales in the Bering Sea, *Canadian Acoustics*, **32** (2), 2004.

ACKNOWLEDGEMENTS

ACCURACY IN THE LOCALIZATION OF SPERM WHALES RESIDENT IN THE STRAIT OF GIBRALTAR USING ONE HYDROPHONE

Christophe Laplanche[1], Olivier Adam, Jean-François Motsch

LiiA (Laboratoire d'Informatique Industrielle et d'Automatique), Université Paris XII
61 av. du Général de Gaulle, 94010 Créteil, France

ABSTRACT

The geographical distribution in the Strait of Gibraltar of sperm whales (*Physeter macrocephalus*) and of four species of dolphins suggests some common foraging territories between the species, and a subsequent share of the water column. A localization method using one hydrophone at an unknown depth has been used here to estimate the foraging depth of sperm whales. The sperm whales tracked in the Main Channel of the Strait of Gibraltar have been found to be hunting at the bottom of the water column.

The localization method calls for the use of a large vertical 4-hydrophone array. The accuracy of this method is assessed using random variables. A simple analytic expression the error on the depth of the source is then calculated. The results of the localization method are checked by considering an second sperm whale and a second hydrophone.

Three ray propagation models have been compared to study the importance of the heterogeneity of the speed of sound in the localization process.

RÉSUMÉ

Des études visuelles en surface dans le Détroit de Gibraltar montrent que les grands cachalots (*Physeter macrocephalus*) et plusieurs espèces de dauphins partagent leurs territoires de chasse. Une méthode de localisation par acoustique passive utilisant un hydrophone à une profondeur inconnue a été utilisée ici, afin d'estimer la profondeur de chasse des cachalots, et prouver un éventuel partage de la colonne d'eau entre les différentes espèces. Les cachalots poursuivis dans le Détroit de Gibraltar chassent dans la partie inférieure de la colonne d'eau.

La méthode de localisation simule l'utilisation d'un réseau vertical à 4 hydrophones de grande dimension. La précision dans la méthode de localisation est estimée à l'aide de variables aléatoires. Une expression analytique de l'erreur sur la profondeur de la source est donnée. Un second cachalot et un second hydrophone ont été considérés afin de vérifier les résultats de la méthode de localisation. Trois modèles de propagation ont été comparés afin d'étudier l'importance de l'hétérogénéité de la célérité du son dans l'estimation de la position de la source.

1 INTRODUCTION

The Strait of Gibraltar, situated between western Europe and northern Africa, is a common place for many species of toothed whales. Visual countings from the sea surface enable a plotting of the geographical distribution of the whales, and show that some species do share common surface territories, especially long finned pilot whales (*Globicephala melas*) and sperm whales (*Physe-*

ter macrocephalus). The vertical repartition of these two species during hunting, and likely a clear share of the water column, can be assessed by locating the animals using passive acoustics. The localization method using one hydrophone at an unknown depth [Thode et al., 2002] has been applied, as it is cheap and easy to set *in situ*.

The localization method may be very sensitive to errors in the measured data, and its accuracy is estimated by

---

[1]Email : laplanche@univ-paris12.fr
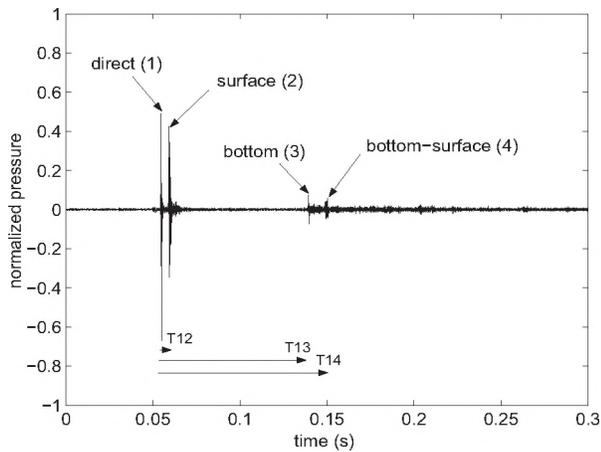
Figure 1 - **Received signal of a click**. It is composed of the click transmitted *via* the direct path (1), reflected once by the sea surface (2), reflected once by the sea bottom (3), and reflected twice at first by the sea bottom then by the sea surface (4).

using random variables. The theory of this uncertainty estimation technique is developed here in general terms, and can also be applied in other localization methods.

A constant speed of sound along the water column has been at first applied when locating. But variations in speed of sound may be important, and the consequences of these variations on the measurement of the data are investigated. The previous localization method is then enhanced, to take into account these variations in speed of sound.

2 CLICKS AND ECHOES

The oceanic propagation field is limited by two interfaces, the sea bottom and the sea surface, which diffuse and reflect acoustic waves. The signal emitted from a source travels *via* different paths, and arrives at a receptor as the sum of the signals transmitted *via* a direct path and its multireflected echoes. The shallow basin of the Strait of Gibraltar allows such multipath propagation to take place. Surface and bottom echoes of sperm whale clicks are commonly detected in the Strait of Gibraltar (Figure 1).

The sperm whale signal received by the hydrophone, as the sum of the sperm whale clicks transmitted *via* the direct path and its three delayed echoes, can be interpreted as the sum of four signals received on four hydrophones (the real one and three additional virtual ones). Each sperm whale click is then recorded on a large vertical
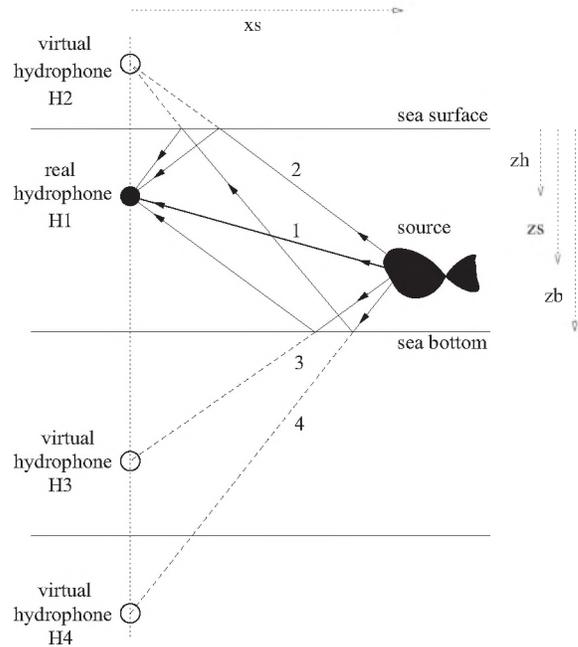
Figure 2 - **Virtual large vertical 4-hydrophone array**. The depth z_h of the hydrophone, the depth z_s of the source, and the range x_s of the source from the hydrophone are estimated from the measurement of the depth z_b of the sea bottom, the speed of sound, and the delays between the four received signals.

4-hydrophone array (Figure 2).

3 CONSTANT SPEED OF SOUND LO-CALIZATION

By measuring delays, a source can be localized in range and in depth by using three hydrophones. These hydrophones are to be located on a common vertical line at known depths. Each sperm whale click is recorded here on a virtual 4-hydrophone array. The depth z_s of the sperm whale can be found by using any of the four different triplets of hydrophones ($\{H_1, H_2, H_3\}$, $\{H_1, H_2, H_4\}, \{H_1, H_3, H_4\}$, or $\{H_2, H_3, H_4\}$), as a function of the depth z_h of the hydrophone H_1. Considering the 3-hydrophone array $\{H_1, H_2, H_3\}$, the depth of the source is :

$$z_s = \frac{4z_b(z_b - z_h)\tau_{12} - c_0^2\tau_{12}\tau_{13}\tau_{23}}{4z_b\tau_{12} + 4z_h\tau_{23}} \quad (1)$$

and by considering the array $\{H_1, H_3, H_4\}$, it is found as :

$$z_s = \frac{4z_b(z_b + z_h)\tau_{13} - 4z_b(z_b - z_h)\tau_{14} - c_0^2\tau_{13}\tau_{14}\tau_{34}}{-4z_b\tau_{34} + 4z_h\tau_{14}} \quad (2)$$
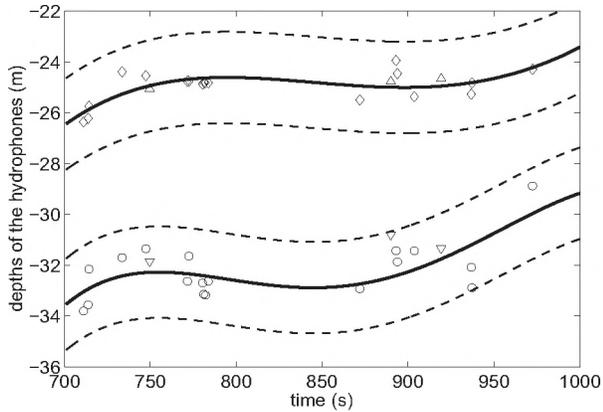
Figure 3 - **Depths of the hydrophones**. The depths of the hydrophones are found from the delays on clicks emitted from the first sperm whale ($\bigcirc$ and $\diamond$) and from the second sperm whale ($\triangle$ and $\triangledown$). The results are interpolated, and a 95 % uncertainty band is plotted, of width $4\sigma_{z_h}$ ($\sigma_{z_h} = 0.9$ m).
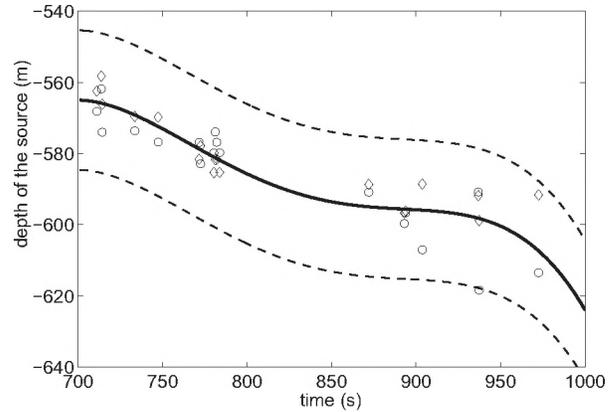
Figure 4 - **Depth of the sperm whale**. Depth of the sperm whale found from clicks measured on the first hydrophone ($\bigcirc$) and on the second hydrophone ($\diamond$). A 95 % uncertainty band is plotted, of width $4\sigma_{z_s}$ ($\sigma_{z_s} = 9.8$ m).

The depth z_s of the source is a function of the following parameters, the depth z_b of the sea bottom, the speed of sound c_0 (a supposed constant along the water column), and the delays τ_{1i} of the signal transmitting *via* the path i on the signal transmitting *via* the direct path (Figure 1). The time differences $\tau_{ji} = \tau_{1i} - \tau_{1j}$ represent the differences between the arrival times of the signals following path i and path j or, equivalently, arriving at hydrophone H_i and hydrophone H_j.

The depth of the source also depends on the depth z_h of the hydrophone. A unique solution for the depths z_s of the source and the depth z_h of the hydrophone can then be found from (1) and (2) given some asumptions on the depth of the source ($z_b - z_h < z_s < z_h$).

The virtual 4-hydrophone array considered here is not overdetermined, as the depth of the real hydrophone H_1 is unknown. Only two 3-hydrophone arrays ($\{H_1, H_2, H_3\}$ and $\{H_1, H_3, H_4\}$) have been used in the localization process. The use of additional arrays (such as $\{H_1, H_2, H_4\}$ or $\{H_2, H_3, H_4\}$) is useless in the localization process. Indeed, the whole information on the location of the source, here the time differences τ_{ji}, can be found by measuring the 4 delays τ_{1i} when using any pair of 3-hydrophone arrays. The source location is to be found equal when considering any pair of 3-hydrophone arrays (Figure 7). The overall error on the estimated location cannot then be reduced by repeating the localization process on different sets of virtual hydrophones. Nevertheless, the detection of more surface/bottom echoes of sperm whale clicks could help.

4 RESULTS

20 clicks from a single sperm whale have been selected in a 5-minute recording. A single cable 2-hydrophone array has been used during the recording campaign in the Strait of Gibraltar, towed from the CIRCé ship Elsa. The localization method has been tested twice on two different hydrophones as to confirm the results. The delays on sperm whale click echoes have then been measured on two hydrophones, these estimated depths are plotted (Figure 3). Both hydrophones are attached to the same cable, and a common movement of the array of 1 m/min in depth is detected. A second sperm whale was clicking in the same recording, and the measurement of the delays on three of its clicks confirms the results of the depths of the hydrophones found by using the clicks of the first sperm whale.

The measurement of the delays on the 20 selected clicks on the two hydrophones enables a calculation of the depth of the source on 40 points (Figure 4). The tracked sperm whale moves downwards at a speed close to 0.2 m/s.

The results, the depth of the source and the depths of the hydrophones, are interpolated. Each calculated point is included in a band in which width represents the uncertainty on the result as calculated using the following method.

σ_τ	σ_{z_b}	σ_{c_0}	σ_{z_h}	σ_{z_s}	σ_{x_s}
0.1 ms	10 m	2 m/s	0.9 m	9.8 m	125 m

Table 1 - **Uncertainties on the measurements and the results**. The depth of the sea bottom is $z_b = -870 \pm 2\sigma_{z_b} = -870 \pm 20$ m with a 95 % probability.

5 UNCERTAINTY ESTIMATION

5.1 Theory

Let $\widehat{x}_1, ..., \widehat{x}_n$ be the measurements of $x_1, ..., x_n$ with the accuracies $\sigma_1, ..., \sigma_n$. Let $\widehat{y} = f(\widehat{x}_1, ..., \widehat{x}_n)$ be the estimated value of $y = f(x_1, ..., x_n)$ with the uncertainty σ_y. The uncertainty σ_y can be estimated from a linear uncertainty estimation method by writing :

$$\sigma_y^2 = \sum_{i=1}^{n} \left[\frac{\partial f}{\partial x_i}(\widehat{x}_1, ..., \widehat{x}_n) \right]^2 \sigma_i^2 \qquad (3)$$

σ_i is the uncertainty on the measurement $\widehat{x}_i$ of x_i, $|\partial f/\partial x_i(\widehat{x}_1, ..., \widehat{x}_n)|\sigma_i$ is the uncertainty on $\widehat{y}$ due to the uncertainty on the measurement $\widehat{x}_i$, and σ_y as given by (3) is then the quadratic mean of the uncertainties on $\widehat{y}$ due to the uncertainties on each of the measurements $\widehat{x}_i$. But errors on $\widehat{y}$ due to errors on some parameters x_i may be compensated by errors due to other parameters x_j. The linear error estimation given by (3) does not take into account such compensations. Random variables naturally do so, and may be useful to estimate the uncertainty in the result of a function of several parameters.

The uncertainties on the measurements of the parameters x_i can be modelized using random variables. Let $X_i \sim N(\widehat{x}_i, \sigma_i^2)$ be n Gaussian distributed random variables distributed around the means (the measurements) $\widehat{x}_i$ with the standard deviations (the accuracies) σ_i. σ_i, representing the uncertainty on $\widehat{x}_i$, is chosen such that the real value x_i would be a sample of the random variable X_i. Then, given properties on Gaussian distributed random variables, x_i has a 68 % probability to belong to $[\widehat{x}_i - \sigma_i, \widehat{x}_i + \sigma_i]$, a 95 % probability to belong to $[\widehat{x}_i - 2\sigma_i, \widehat{x}_i + 2\sigma_i]$, and a 99 % probability to belong to $[\widehat{x}_i - 3\sigma_i, \widehat{x}_i + 3\sigma_i]$.

The uncertainty on the estimated value $\widehat{y}$ is also modelized using random variables, here $Y = f(X_1, ..., X_n)$. The mean, the standard deviation, and the distribution of Y can be estimated using a large number m of samples $y^{(k)}$ of Y. Each of these samples is calculated from samples $x_i^{(k)}$ of the random variables X_i, by writing
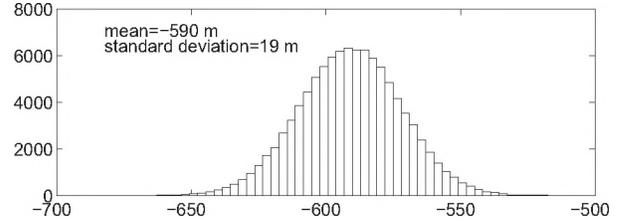
Figure 5 - **Histogram of the depth of the source**. The random variable representing the depth of the source is almost Gaussian, with a mean $z_s = -590$ m and a standard deviation $\sigma_{z_s} = 19$ m. The tracked sperm whale is then at a depth $z_s = -590 \pm 38$ m with a 95 % probability.

$y^{(k)} = f(x_1^{(k)}, ..., x_n^{(k)})$. The mean y_m and the standard deviation σ_y of Y can then be estimated by :

$$y_m \simeq \frac{1}{N} \sum_{k=1}^{N} y^{(k)} \qquad \sigma_y^2 \simeq \frac{1}{N} \sum_{k=1}^{N} [y^{(k)} - y_m]^2 \qquad (4)$$

and its distribution is estimated by plotting a histogram of the samples $y^{(k)}$ (Figure 5).

Let σ_τ, σ_{z_b}, and σ_{c_0} be the uncertainties on the parameters τ_{1i}, z_b and c_0 (Table 1). The resulting uncertainty on $z_s = f(\tau_{1i}, z_b, c_0)$ can then be found from (4) by generating m samples $z_s^{(k)} = f(\tau_{1i}^{(k)}, z_b^{(k)}, c_0^{(k)})$ ($m = 100000$ has been drawn to plot the histogram of Figure 5). An approached value of the uncertainty σ_{z_s} found from (4) can be calculated analytically, by using some properties of Gaussian distributed random variables, as described next.

5.2 Analytic expression

Let $X_1 \sim N(\widehat{x}_1, \sigma_1^2)$ and $X_2 \sim N(\widehat{x}_2, \sigma_2^2)$ be two independant Gaussian distributed random variables. Then $X_1 + X_2 \sim N(\widehat{x}_1 + \widehat{x}_2, \sigma_1^2 + \sigma_2^2)$ is also a Gaussian distributed random variablof mean $\widehat{x}_1 + \widehat{x}_2$ and of standard deviation $\sqrt{\sigma_1^2 + \sigma_2^2}$. σ_1 being the uncertainty on the measurement $\widehat{x}_1$, and σ_2 being the uncertainty on the measurement $\widehat{x}_2$, the uncertainty on $\widehat{x}_1 + \widehat{x}_2$ is then the quadratic mean of two uncertainties, one due to the uncertainty on $\widehat{x}_1$ (σ_1), and a second one due to the uncertainty on $\widehat{x}_2$ (σ_2). Similar evaluations of uncertainties can be calculated for different basic operations of random variables (product, ratio, square, square root). The uncertainty σ_y on $\widehat{y} = f(\widehat{x}_1, ..., \widehat{x}_n)$ can then be estimated by decomposing f in basic operations and by calculating σ_y step by step.

Analytic expressions of the uncertainties on the depth z_s of the tracked sperm whale and the depth z_h of the hy-

propagation model	τ_{12}	τ_{13}	τ_{14}
rectilinear propagation constant speed of sound	7.8 ms	100.6 ms	114.9 ms
rectilinear propagation non constant speed of sound	7.8 ms	97.0 ms	111.3 ms
non rectilinear propagation non constant speed of sound	7.7 ms	96.9 ms	111.3 ms

Table 2 - **Propagation models.** The propagation of sperm whale clicks is simulated in a configuration of the source and the receiver close to what is found from the constant speed of sound localization method. Delays on bottom echoes can here be overestimated by 3 ms when not taking into account variations in speed of sound.

drophone can then be calculated from the expressions of z_s and z_h found from (1) and (2). The uncertainties σ_{z_s} and σ_{z_h} are found as quadratic means of the uncertainties driven by the uncertainties on each of the parameters τ_{1i}, z_b, and c_0. The uncertainty in the depth of the source is found close to :

$$\sigma_{z_s}^2 \simeq z_s^2 \left[\frac{(c_0^4/16)[2\tau_{13}^2 + \tau_{23}^2]\sigma_\tau^2 + z_b^2\sigma_{z_h}^2 + (4z_b^2 + z_h^2)\sigma_{z_h}^2}{[z_b(z_b - z_h) - (c_0^2/4)\tau_{13}\tau_{23}]^2} \right.$$

$$+ \frac{(c_0^2/4)\tau_{13}^2\tau_{23}^2\sigma_{c_0}^2}{[z_b(z_b - z_h) - (c_0^2/4)\tau_{13}\tau_{23}]^2}$$

$$\left. + \frac{z_h^2[2\tau_{12}^2 + \tau_{23}^2]\sigma_\tau^2 + \tau_{12}^4\sigma_{z_b}^2 + \tau_{12}^2\tau_{23}^2\sigma_{z_h}^2}{\tau_{12}^2[\tau_{12}z_b + \tau_{23}z_h]^2} \right]$$

$$(5)$$

The result is nevertheless an approximation of what could be found from (3) or (4), as making strong hypotheses on the gaussianity and the independance of the parameters step after step in the calculation process.

The uncertainties on the depth of the hydrophone, the depth of the source and the range of the source are found close to each other when estimated by sampling (Table 1) or analytically ($\sigma_{z_h} = 0.8$ m, $\sigma_{z_s} = 15.5$ m and $\sigma_{x_s} = 145$ m), and close to the results (Figure 3 and Figure 4).

6 VARYING SPEED OF SOUND LOCALIZATION

6.1 Variations in the speed of sound

The previous localization and uncertainty estimation techniques assumed the speed of sound to be constant in the sea water. The value of the speed of sound in sea water depends on local conditions of temperature,
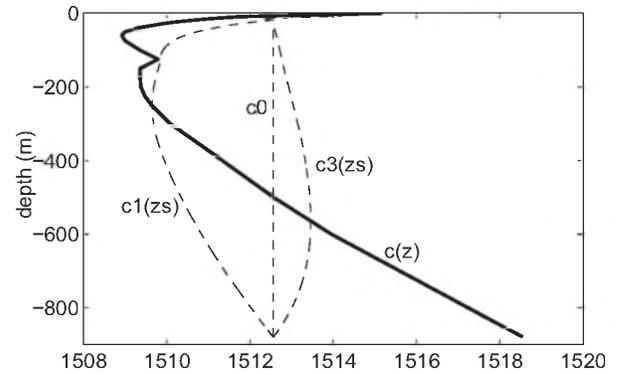
Figure 6 - **Speed of sound.** The speed of sound $c(z)$ in sea water varies mostly with depth. The mean speed of sound along the whole water column is $c_0 = 1512.6$ m/s. The mean speed of sound along the ascending rays (paths (1) and (2)) is close to $\tilde{c}_1(z_s)$. The mean speed of sound along descending rays (paths (3) and (4)) is close to $\tilde{c}_3(z_s)$. Sperm whale clicks reflected by the sea bottom will then propagate faster than sperm whale clicks reflected by the sea surface (as $\tilde{c}_3(z_s) \geq \tilde{c}_1(z_s)$).

salinity, pressure and current. These parameters vary, mostly with depth. The resulting variations of the speed of sound in the Strait of Gibraltar are plotted in Figure 6. The importance of the speed of sound variations on the accuracy of the previous localization method is now studied.

The propagation of sound in sea water tends to be more complex with such variations in the speed of sound. Rays are curved due to refraction effects, and the speed of sound varies along the rays. The propagation times of signals from a source to a receiver, depending approximately on the length of the rays and of the mean speed of sound along the rays (as the speed of sound values are much greater than its variations), may then be wrongly estimated when supposing a constant speed of sound.

Three propagation models are compared, as to estimate the effects of the speed of sound variations (curvature of the rays, and different mean speeds of sound along the rays for different rays) on the misevaluation of the delays on received sperm whale click signals (Table 2). The curvature of the rays introduces negligible errors on the measurement of delays (the deviations of the rays from the rectilinear paths are inferior to 8 m as found by raytracing). The variations in the mean speed of sound with rays ($\tilde{c}_1(z_s) \neq \tilde{c}_3(z_s)$, Figure 6) however do induce important errors on the measurement of delays. The previous localization technique assuming a constant speed of sound may then be quite inaccurate, and is then enhanced by taking into account the variations in speed of sound.

6.2 Localization

The propagation time of a signal from a source to a receiver along a given ray is then given by the length of the ray, and by the mean speed of sound along the ray. The rays of the signals transmitted *via* the direct path and the rays of the signals reflected once by the sea surface are close to each other (the hydrophone is close to the surface). And so is the mean speed of sound along these two rays. And so it is for the rays of the two signals reflected by the sea bottom. Then, by labelling $\widetilde{c_1}$, $\widetilde{c_2}$, $\widetilde{c_3}$, and $\widetilde{c_4}$ the mean speeds of sound along the rays (1), (2), (3), and (4) (Figure 2), one can find $\widetilde{c_2}(z_s, z_h) \simeq \widetilde{c_1}(z_s, z_h) \simeq \widetilde{c_1}(z_s)$ and $\widetilde{c_4}(z_s, z_h) \simeq \widetilde{c_3}(z_s, z_h) \simeq \widetilde{c_3}(z_s)$.

The localization method is then close to the one used in the previous constant speed of sound method. Using different triplets of hydrophones, 4 polynomial equations on the depth z_s of the source and the depth z_h of the receiver are found, such as, considering $\{H_1, H_2, H_3\}$:

$$\left[4z_b(z_b - z_s) - \frac{\widetilde{c_3}(z_s)^2 - \widetilde{c_1}(z_s)^2}{4}\tau_{12}^2 - \widetilde{c_3}(z_s)^2\tau_{13}\tau_{23} \right] +$$

$$\left[-4(z_b - z_s) - 2\frac{\widetilde{c_3}(z_s)^2 - \widetilde{c_1}(z_s)^2}{\widetilde{c_1}(z_s)^2}z_s - 4\frac{\widetilde{c_3}(z_s)^2\tau_{13}}{\widetilde{c_3}(z_s)^2\tau_{12}}z_s \right]z_h +$$

$$\left[-4\frac{\widetilde{c_3}(z_s)^2 - \widetilde{c_1}(z_s)^2}{[\tau_{12}\widetilde{c_1}(z_s)^2]^2}z_s^2 \right]z_h^2 = 0$$

$$(6)$$

The solution sets of these equations are plotted (Figure 7), and the intersection points are found numerically. The results are consequently close to those found by the constant speed of sound localization method (the mean deviation of the results on z_s found by this method and the constant speed of sound one, as plotted at Figure 4, is close to 1.9 m).

7 CONCLUSIONS

Sperm whales can be localized in range and in depth using only one hydrophone at an unknown depth. The feasibility of the method has been asserted by using a second hydrophone, by detecting clicks emitted from a second sperm whale, and by estimating the accuracy on the estimated locations using random variables. The uncertainty estimation technique developed here may be used in other localization systems using more than one hydrophone.

The depth of the sea bottom and the value of the mean speed of sound are required in the present localization method. The only requirements are the surface coordi-
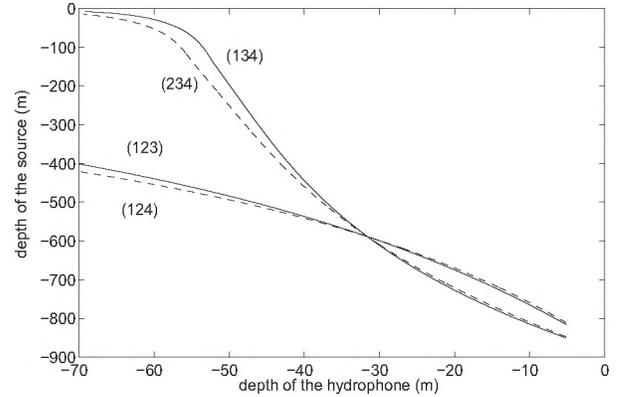
Figure 7 - **Varying speed of sound localization**. The depth z_s of the source and the depth z_h of the receiver are found from the intersection of the 4 solution sets corresponding to the 4 different triplets of hydrophones of the virtual 4-hydrophone array.

nates, for instance given by a GPS receiver, of the place of the recordings. The resulting values of the depth of the sea bottom and the speed of sound would then be given by bathymetric charts and oceanic models. The MERCATOR project enables estimations of the variations of speed of sound in the whole Mediterranean Sea. The use of past or future recordings, made with a GPS receiver and a single hydrophone, could then be used to estimate the depths of diving sperm whales.

Statistics on the diving depths of sperm whales could be made in a given area by using recordings stored in databases. The diving behaviour of sperm whales could then be better understood, in high traffic areas like the Strait of Gibraltar or the Ligurian Sea, or why not in the whole Mediterranean Sea.

ACKNOWLEDGEMENTS

REFERENCES

[Thode et al., 2002] Thode, A., Mellinger, D., and Stienessen, S. (2002). Depth-dependant acoustic features of diving sperm whales (physeter macrocephalus) in the gulf of mexico. *J. Acoust. Soc. Am.*, 112 :308–321.

YOU CAN'T GET THERE FROM HERE:
SHALLOW WATER SOUND PROPAGATION AND WHALE LOCALIZATION

David M.F. Chapman
Defence R&D Canada – Atlantic, P.O. Box 1012, Dartmouth, N.S., B2Y 3Z7.
dave.chapman@drdc-rddc.gc.ca

SUMMARY

Sound propagation in summer conditions in the Bay of Fundy is modelled here for the case of a shallow source (a whale at 10 m depth) communicating with a bottomed receiver (an ocean bottom hydrophone at 163.1 m depth). It is shown that the signal strength along the direct path at long ranges (5–8 km) is extremely weak, for three reasons: (1) destructive interference of the shallow source and its image in the sea surface, (2) destructive interference between paths arriving at the bottom and their bottom-reflected counterparts, and (3) upward refraction by the positive sound speed gradient at the seabed. The first significant signals arriving at long ranges are paths that reflect from the surface and the bottom several times, the number of times increasing with range. Consequently, localization algorithms based on the assumption of direct straight-line paths are prone to bias and error. It is suggested that a simple straight-line, average-speed model could be made to work if the algorithm were to admit the hypothesis that the paths could be reflected paths, which could be accommodated simply by using the method of images.

SOMMAIRE

On décrit dans ce document la modélisation de la propagation du son dans des conditions estivales dans la baie de Fundy dans le cas d'une source à faible profondeur (une baleine à une profondeur de 10 m) communiquant avec un récepteur sur le fond (un hydrophone sur le fond océanique par une profondeur de 163,1 m). Il est démontré que la puissance du signal à longue portée (5 à 8 km) suivant la trajectoire de propagation directe est extrêmement faible et ce pour trois raisons : 1) interférence destructive entre la source peu profonde et son image à la surface de la mer, 2) interférence destructive entre les trajectoires arrivant au fond et leur réflexion sur le fond et 3) réfraction vers le haut attribuable au gradient positif de vitesse du son au fond marin. Les premiers signaux significatifs arrivant à de longues portées sont ceux dont la trajectoire est réfléchie plusieurs fois à la surface et au fond, le nombre de réflexions augmentant en fonction de la portée. En conséquence, les algorithmes de localisation basés sur l'hypothèse voulant que les trajectoires de propagation directe en ligne droite sont sujets à des biais et des erreurs. Il est suggéré qu'un modèle simple basé sur la propagation en ligne droite à vitesse moyenne pourrait fournir de bons résultats si l'algorithme était modifié de manière à tenir compte de l'hypothèse voulant que les trajectoires puissent être celles de rayons réfléchis, ce qui pourrait se faire simplement par la méthode des images.

1. INTRODUCTION

It is evident that successful acoustical localization of whales depends heavily on the fidelity of the sound propagation model used, at least with respect to travel times, but possibly also with respect to waveform shape, in cases where correlation techniques are used. Typical "hyperbolic" underwater position fixing often assumes direct straight-line paths with a constant sound speed. This model can be adequate at short range (several water depths), but may break down in some environments at longer ranges owing to a combination of several physical acoustic effects. (A definition of hyperbolic position fixing: For two receivers

of known location, if one knows the difference between arrival times of a pulse from a source of unknown location, the locus of possible source positions forms a hyberbolic surface, if the signal speed is constant. In order to reduce the positional ambiguity, arrival time differences from multiple pairs of receivers are needed: the near-intersection of the hyperbolic surfaces fixes the source position, within some error bound.)

The environment in question is a portion of the Bay of Fundy, a shallow water region of average depth 164 m over a seabed composed of a surficial layer of LaHave clay (1–10 m thick) over a basement of Scotian Shelf drift, or till. In
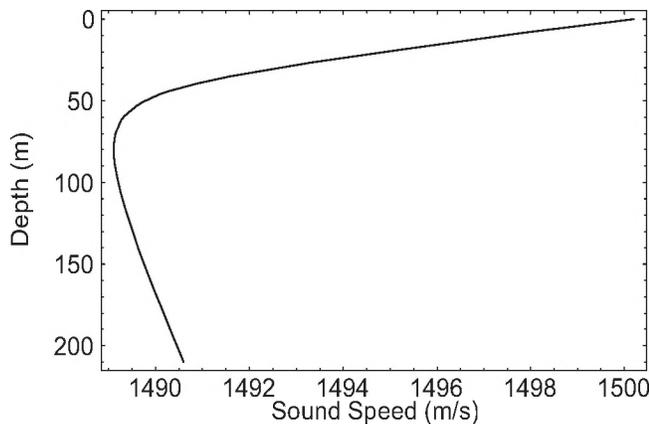
Figure 1. *A simplified sound speed profile for the Bay of Fundy in summer. Note the upward-refracting gradient at the bottom, the location of the receivers.*

summer there is a strong downward-refracting sound speed profile near the sea surface and a weak upward-refracting profile at the seabed, which can be approximated by a smoothed bilinear profile with a minimum at about 75 m depth, as shown in Figure 1.

The Northern Right Whale typically vocalizes near the surface, not while diving, so we assume a shallow source depth of 10 m for modelling purposes. For the experiments described elsewhere in these Proceedings, the receivers are ocean bottom hydrophones (OBHs) mounted 0.9 m above the seafloor, so we use a receiver depth of 163.1 m for modelling purposes. The elements of the array of receivers used for whale localization are widely separated, several kilometers apart, and it is expected that a whale could be localized both inside and outside this array pattern, perhaps up to a few tens of kilometres away. To give some idea of the angles involved, the direct line-of-sight path from whale to OBH is only about 4 degrees below the horizontal plane at 2 km range, and about 1 degree at 8 km range.

Considering the environment and the geometry, there are three fundamental limits on the assumption that direct-straight line paths are adequate for localization algorithms:

1. The proximity of the source to the surface results in an effective source beam pattern that creates enhancements and nulls at specific angles, owing to constructive and destructive interference between a directly radiated path and its reflection in the surface, which has inverted phase. There is always a null in the horizontal direction, which reduces the effective source strength at long range. This well-known phenomenon is called "Lloyd's mirror" [Jensen *et al.* 1994].

2. The placement of the receiver at the seabed introduces a similar effect: for every path to the receiver through the

water, there is an associated path that reflects from the seabed just before combining with its mate. In effect, arrivals along these two paths arrive simultaneously, but the bottom-reflected path has suffered an amplitude loss and a phase shift. This effect would be the same for a receiver within a small fraction of a wavelength of the seabed. At near-horizontal angles, for realistic seabeds, the reflection is almost perfect in amplitude with a reversal in phase. The combination of these arrivals results in poor sensitivity of an OBH at low grazing angles.

3. The positive sound speed gradient near the seabed tends to refract sound upwards away from the bottom, decreasing overall signal amplitude there; in extreme cases, there may exist a shadow zone for some source depths (deeper than 41 m in this case), preventing acoustic rays from reaching the OBH directly.

In this short paper we will briefly explain the origin of these effects. We then examine their combination using an underwater acoustic model that correctly combines the relevant physical factors. Finally, we suggest a possible work-around for those who are constrained to use localization algorithms that assume direct straight-line propagation of rays.

2. EFFECT OF NEAR-SURFACE SOURCE

The change in effective level (in decibels) of an omnidirectional source near a perfectly reflecting (but phase-inverting) sea surface is given by the Lloyd's mirror expression

$$\Delta L_S = 20\log_{10}\left[1 - \exp\left[i4\pi\left(fz/c\right)\sin\theta\right]\right] \text{ dB}, \tag{1}$$

in which f is the frequency, z is the source depth, c is the sound speed, and θ is the angle of propagation relative to
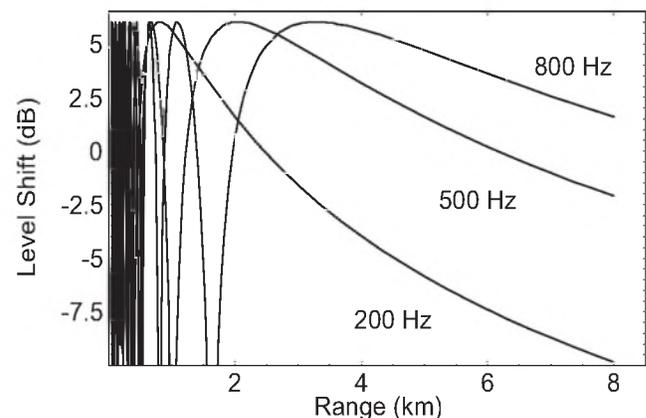
Figure 2. *Effect of locating the source 10 m from the surface, for a receiver at 162.1 m depth, at several frequencies, in an isospeed environment.*

the horizontal plane. Even for an unbounded isospeed environment, this has consequences as source-receiver range increases. Figure 2 shows the effect for a shallow source communicating with a deep receiver at three frequencies. (The effects of the reflecting bottom, multipath propagation, and spherical spreading are not yet included.) The effect is complex at short ranges, with both dropouts and enhancements in signal level, highly frequency-dependent. More importantly, at long ranges, there is a significant decrease of level, particularly at low frequencies, owing to the presence of the horizontal null.

3. EFFECT OF BOTTOMED RECEIVERS

The change in effective response (in decibels) of an omnidirectional bottomed receiver is given by the expression

$$\Delta L_R = 20\log_{10}\left[1 + R(\theta)\right] \text{ dB}, \qquad (2)$$

in which $R(\theta)$ is the complex plane-wave reflection coefficient of the seabed at grazing angle θ. (This expression results simply from adding the contribution from a ray path and its bottom-reflected mate at the seabed; there is no phase delay between them other than that introduced by the reflection.) Again, for an unbounded isospeed environment, this has consequences as source-receiver range increases. Figure 3 shows the effect for a bottomed receiver receiving signals from an elevated source for the two bottom types, as a function of range. (The effects of surface reflection, multipath propagation, and spherical spreading are not yet included.) Note that the response is enhanced at short range, owing to reflected in-phase energy; however, at long range the out-of-phase reflected energy partially cancels the direct arrival. The magnitude of the effect is sensitive to the acoustic properties of the seabed.

The calculation of $R(\theta)$ for a semi-infinite homogeneous elastic seabed is standard and can be found in [Jensen et al. 1994], although they do not show the angular dependence of the phase shift. Table 1 gives the values of the seabed parameters we used: density, compressional wave speed, compressional attenuation, shear wave speed, and shear attenuation. (These parameters are not unique: one can observe the same acoustic effect with seabeds having different acoustic parameters.)

Table 1: Acoustic parameters of the seabed types

	ρ	c_p	α_p	c_s	α_s
	[gm/cm^3]	[m/s]	[dB/λ]	[m/s]	[dB/λ]
Clay	1.54	1520	0.2	50	2
Till	2.1	1830	0.6	400	1

4. EFFECT OF REFRACTION BY THE SOUND SPEED PROFILE

Finally, there is the effect of the sound speed profile itself. A variable sound speed profile refracts rays and modifies the variation of signal amplitude along the rays. The presence of a positive gradient of sound speed at the bottom may create a shadow zone, depending on the source depth. One method of illustrating this is to trace a ray that leaves the receiver in the horizontal direction, and see how far it must travel to reach a given source depth. Using the simplified sound speed profile in Figure 1, we plot such a ray in Figure 4. At a given source depth, sources at ranges shorter than the maximum range may communicate with the bottomed receiver along a direct path ray that arrives at the receiver with positive angle; sources at longer ranges have no direct path, but may be able to reach the receiver through a path reflected from one or more boundaries, or refracted

Figure 3. *Effect of locating receiver on the bottom, receiving signals from a source elevated 153.1 m, for two seabed types in an isospeed environment.*
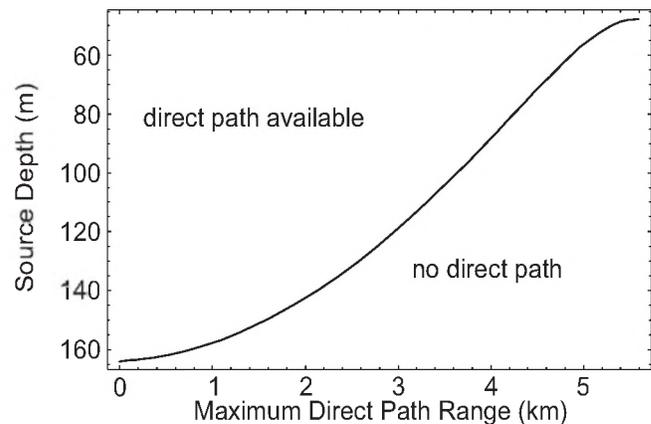
Figure 4. *Maximum range of a direct path between a source and a receiver bottomed at 163.1 m, for the spond speed profile in Figure 1. Sources shallower than 41 m always have a direct path available.*

back by the down-ward refracting gradient at the surface. (In a sense, this is a kind of direct path, but not for the purpose at hand.) Any of these paths will have longer travel times than a straight-line path would suggest. The full effects of refraction will be covered in the next section.

5. THE COMBINED EFFECT

To treat all physical effects properly, one needs to use a pulse propagation model that can handle the proximity of the source to the surface, the refractive effects of the sound speed profile, the reflective properties of the seabed, and the response of the bottom receiver. We used the OASES model [Schmidt, 1999 and 1988], which was originally developed for seismo-acoustic modelling in stratified ocean media. (OASES and other useful models can be found on the internet through SAIC's Ocean Acoustics Library, http://oalib.saic.com/). We calculated the band-limited impulse response of this channel between a shallow source and a near-bottom receiver at several ranges: 2 km, 5 km, and 8 km, shown in Figure 5. We considered the frequency band 100–800 Hz, roughly matching the Right Whale "gunshot" sounds. The seabed was made of 2 m of clay

over a halfspace of till. The time axis of the plot is adjusted so that zero time at each range corresponds to a the arrival time of a pulse covering the horizontal range at the average speed in the water column, i.e. the adjusted time is $t - r/1.491$, where t is time in seconds and r is range in kilometres. If there were a direct arrival with significant amplitude, it would occur shortly after zero time. Two results are shown, one for the bilinear sound speed profile (upper) and one for an isospeed profile with the average speed of 1491 m/s (lower).

Table 2: Pulse Arrival Time after r/1.491 s

Range r	Arrival #	KosmicRay	OASES isospeed	OASES bilinear
2 km	1	4 ms	4 ms	4 ms
5 km	1	2 ms	2 ms	-
	2	16 ms	16 ms	15 ms
8 km	1	1 ms	-	-
	2	10 ms	10 ms	-
	3	28 ms	29 ms	26 ms

Even for the isospeed environment, early arrivals are attenuated at long range relative to later arrivals; this is even more evident for the case of the bilinear profile, which includes refraction effects. Table 2 compares times of arrivals between the OASES results and a simple, isospeed, straight-line ray model (with multipaths) called KosmicRay, developed by the author. KosmicRay reproduces the travel times and waveforms of the isospeed OASES results, and is used to interpret the OASES calculations in ray terminology.

Figure 6 provides a schematic illustration of the relevant paths, ignoring refraction. The first OASES arrivals

Figure 5. *OASES model outputs of the 100-800 Hz impulse response of the shallow water channel between a shallow source and a near-bottom receiver at three ranges, including all effects of refraction, seabed interaction, and proximity of source and receiver to boundaries. Label numbers refer to arrivals in Table 2.*
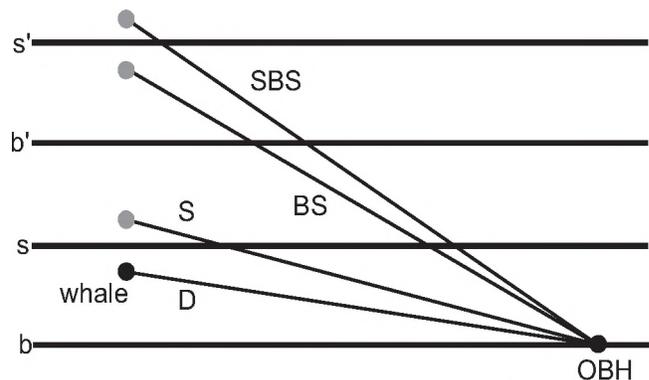
Figure 6. *Simplified ray path analysis of arrival structure, using the image method: on the left, s and b denote sea surface and seabed, respectively; s' and b' denote images of these planes. Within the figure, D denotes a direct path with no reflections, S denotes a surface-reflected path, BS denotes a bottom-surface-reflected path, and SBS denotes a surface-bottom-surface-reflected path*

(labeled "1" in Figure 5), interpreted in terms of rays, are a combination of direct path (D) and a single surface reflection (S). Arrival "2" is a combination of BS and SBS paths, arrival "3" is a combination of BSBS and SBSBS, etc. OASES takes care of the effect of placing the sensor directly on the seabed, discussed in Section 3, so the path with the additional bottom-reflected path is not shown.

For the isospeed OASES case, the first arrival is strong at 2 km, discernable at 5 km, and not noticeable at 8 km. The second arrival is prominent in the isospeed OASES case at all ranges. For the bilinear gradient OASES case, the first arrival at 2 km is modified by the sound speed profile, but still prominent; the first arrival at 5 km and the first and second arrival at 8 km are not noticeable. Note that KosmicRay, even though it is an isospeed model, provides a good estimate of the arrival time of the first significant arrival in the gradient OASES case, but it cannot tell you in advance which arrival that will be! This is because these paths travel at relatively steep angles, so the bending of the ray paths are not significant, and they traverse the entire water column one or more times, so an average sound speed is sufficient. (For estimating travel times, it is actually more appropriate to average the "slowness" (inverse of sound speed) rather than the sound speed, but the difference is slight unless the profile has large gradients.)

6. CONCLUSIONS / FUTURE WORK

We have shown that several environmental and geometric factors combine to suppress direct path signal arrivals in the case of whale localization in the summer conditions in the Bay of Fundy, particularly at longer ranges (5–8 km). The consequences of this for localization have not been investigated, but those using simple straight-line rays in their localization algorithms should be alert to bias and/or error in their position estimates introduced by this assumption. Although absolute delays in arrival times of pulses have been presented here, what is relevant is relative arrival times of first significant arrivals at sensors at different ranges from the source.

The success of a simple straight-line ray model in predicting the arrival times of reflected arrivals suggests that existing algorithms could be adapted by introducing image sources to account for multipath geometry. For example, refering to Figure 6, if the water depth is H and

the true source depth is z, then placing the source at a depth of $-z$ would account for path S; source depths of $2H \pm z$ account for paths SBS and BS, and so on. The image method would naturally result in greater travel times for these paths, and the arrival time differences would reflect the changed geometry. (Note that the travel time differences between receivers actually *decrease* for multipaths, even though the path travel times increase. This is a geometric effect.) To include multipaths in position-fixing algorithms, one is faced with which multipath to choose, which is not immediately obvious unless one uses a propagation model that includes the effects described above. One way of dealing with this issue would be to compute multiple solutions associated with multiple paths, and to select the solution that provides the lowest fix error.

Another consideration to be investigated is the consequences of interference-induced modification of the effective source spectrum by reflection at the sea surface and at the seabed. This may affect detection algorithms that rely on correlation in time or in frequency, as these reflection/interference effects alter both the spectrum and the waveform of the signal.

ACKNOWLEDGEMENTS

REFERENCES

Jensen, Finn B., William A. Kuperman, Michael B. Porter, and Henrik Schmidt, *Computational Ocean Acoustics*, American Institute of Physics (New York, 1994): page 16 for Lloyd's mirror, page 46 for seabed reflection.

Schmidt, Henrik, OASES Version 2.2 User Guide and Reference Manual, Henrik Schmidt (MIT, 1999) this is a file supplied with the software package

Schmidt, Henrik, SAFARI: Seismo-Acoustic Fast field Algorithm for Range-Independent Environments, SACLANTCEN REPORT SR-113, 1988.

NEWS / INFORMATIONS

CANADIAN ANNOUNCEMENTS

Eugene H. Bolstad, P.Eng., known by many as the grandfather of acoustics in western Canada, has retired from a long and respected career in acoustical engineering in Edmonton at the tender age of 80 (truth be told, this is technically his 3rd retirement). Mr. Bolstad started his career many years ago as a building mechanical system draftsman and designer. Along the way a colleague convinced him that he should challenge the Professional Engineering exams. After a few years and about a dozen senior level Engineering exams, he earned the right to call himself a Professional Engineer. Although his background was with building mechanical systems, he was naturally drawn to acoustical projects. He has been known to comment, "I spent enough years causing noise problems that I figured it was time to start fixing them". In the 70's, he built one of only two fully functioning reverberation chamber test facilities in the country (the other being the facilities at the NRC in Ottawa). This was a rather bold move at the time, especially considering that he did it without any external funding. In later years he sold the lab to the University of Alberta where it is still used for testing and research. In more recent years, he banded together with some fellow acoustical Engineers to start a consulting company lending much of his expertise experience in a mentorship role. Anyone who has known him or had the fortune to work along side him will remember his incredibly easy to get along with demeanor and infectious laugh. We are better for having known him and wish Mr. Bolstad well. Until the next retirement…

EXCERPTS FROM "*WE HEAR THAT*", IN *ECHOS*, ASA

Dick Botteldooren, Ghent University, is the new editor-in chief of *Acustica/Acta Acustica*, succeeding Michael Vorländer, who served in that position from 1998-2003.

Michael Möser is the new editor for *General Linear Acoustics*, and **Jian Kang** is the new editor for *Environmental Acoustics*. Special issues in Musical Acoustics and Spatial and Binaural Hearing are planned during 2004.

A joint meeting of **ASA** and **EAA** (European Acoustics Association) is being planned for Paris, June 22-28, 2008.

Gerald Kidd, Armin Kohlrausch, David Dowling, T. Douglas Mast and **James A. Simmons** have been appointed associate editors of JASA.

EXCERPTS FROM "*SCANNING THE JOURNALS*", IN *ECHOS*, ASA

A resource letter on **thermo-acoustic engines and refrigerators** by Steven Garrett appears in the January issue of *American Journal of Physics*. Resource letters, commissioned by the American Association of Physics Teachers, are intended to guide college physics teachers and students to some of the most important papers in various fields of physics. Thermo-acoustic engines and refrigerators incorporate acoustical components and networks to produce mechanical power or to pump heat, or both, with the use of traditional mechanical contrivances such as pistons, linkages, and valves. One of the 106 papers cited is "Build an Acoustic Laser" which appeared in the Fall 2000 issue of *ECHOES*.

When sopranos sing at frequencies that are higher than the lowest resonance (formant) of their vocal tract, their vocal power is reduced. To increase the loudness and uniformity of tone, sopranos learn to **tune their formants** to the frequency of the sung note. New data on formant tuning by sopranos appears in a communication in the 8 January issue of *Nature*. Vocal tract resonances were measured directly for five sung vowels. The large shift in formant frequencies at high sung pitch helps to explain the difficulty in identifying words sung in the high range by sopranos.

The 23 February issue of *Journal of Sound and Vibration* is a special issue with papers from the 2002 **I.M.A. Conference on Computational Aero-acoustics** held in London, April 9-11, 2002.

The January issue of *Acoustical Science and Technology* is a special issue in commemoration of the **China-Japan Joint Conference on Acoustics** (JCA2002) held in Nanjing Nov. 14-17, 2002. The theme of the conference was "Acoustics in Digitalized Era," and the special issue is edited by co-chairpersons of JCA2002, Yôiti Suzuki and Jing Tian.

"Infrasonic Symphony" is the title of an article about infrasound in the January 10 issue of *Science News*. It begins with a riddle: what do rhinoceroses, supersonic aircraft and hurricanes have in common? The answer, of course, is that they all generate infrasound below 20 Hz. Scientists first detected infrasonic waves in 1883 when the eruption of the Krakatoa volcano sent inaudible sound waves around the world, affecting barometric readings. Just as seismic waves travel through Earth, infrasonic waves travel through the air, and the lower the frequency, the farther they can travel. Infrasound is also generated by aurora, which is caused by charged particles in the air. Electricity heats atmospheric gases and the warmed molecules spread out and increase air pressure.

Insomniac Hawaiian islanders are hopping mad over a tiny frog whose **shrill mating call** disturbs their sleep, according to a note in the January issue of *Smithsonian* magazine. Although no larger than a quarter, the male coqui frog creates a shrill screeching sound by forcing air through its balloon-like vocal sac. Sound levels apparently reach as high as 90 dB.

Computer simulations that create maps of European cities in color are at the core of a plan to restore peace and quiet to a population driven to distraction by traffic noise, according to an article in the 5 February issue of *Nature*. Legislation passed last year by the European Union (EU) requires member states to make, by 2007, regular **noise maps** of all major cities, roads, railways, airports and industrial sites. These maps are a powerful way to visualize noise pollution. Noise mappers in Paris have made use of "virtual microphones," each of which is a point in a computer model that reports what the sound level would be at a certain place under given circumstances. In total, the 3D representation of Paris contains 26 million virtual microphones. The Paris map uses software called MITHRA developed by the French Scientific Centre for Building Physics. It models sound as rays and calculates how they interact with different surfaces.

EXCERPTS FROM *"ACOUSTICS IN THE NEWS"*, IN *ECHOS*, ASA

Two Canadian mathematicians have designed a Y-shaped guitar called a tritar, according to a note in the 23 January issue of *Science*. The mathematicians at the University of Moncton, while working on a problem involving infinite sums called p series, invented a series of hyperimaginary numbers lying on a Y-shaped number line. They wondered how waves would behave on a Y-shaped string. To find out, they built a model tritar and took it to a guitar builder, who made a playable version. Plucking each three-ended string creates unpredictable overtones quite unlike the usual ones of a guitar. _ Newly developed software turns compatible camera phones into visual aids for the blind by changing images snapped by the camera into sounds that the user's brain can reconstruct into mental pictures, according to a story in the February issue of *IEEE Spectrum*. Once per second, the computer scans a 64x64-pixel frame from left to right, one column at a time. Each pixel in a column produces a wave whose frequency indicates its position; the highest frequencies are at the top. Amplitude is based on the brightness of the pixel on a 16-tone gray scale. Frequency is then translated into pitch and amplitude into volume; what a listener hears is a musical chord of up to 64 notes. There is evidence that the part of the brain responsible for sight, the visual cortex will, after some training, respond to changes in pitch.

By means of massive digital signal processing, an electric guitar called Variax can model 50 historic guitars, according to a story in the February issue of *IEEE Spectrum*. The instrument, being played by such stars as Pete Townsend, Steve Howe and Joe Walsh, can reconstruct the distinct twang of the Fender Stratocaster or the singing sustain of the Gibson Les Paul.

Two deaf women in the United States have become the first people to undergo the risky procedure of having implants in their brainstems, according to a news note in the 10 January issue of *New Scientist*. The devices are designed to restore hearing by directly stimulating nerves. Implants that sit outside the brainstem do not work as well. Cochlear implants bypass the hair cells and stimulate the auditory nerve directly, but they cannot help people with a damaged cochlea or auditory nerve. At the moment, the only way to restore hearing to people with type II neurofibromatosis (NF2) is to stimulate the brainstem directly. But the procedure is very risky because at the brainstem level every neuron that is damaged can have serious consequences.

A microphone that imitates the remarkably acute hearing of a tiny fly may one day help wearers of hearing aids understand conversation in a busy restaurant, according to a story in the December 11 *New York Times*. The microphone structure is based on the ears of the fly *Ormia ochracea*. The female of this species uses her fine hearing to pick out the sound of distant crickets, which serve as hosts on which she can deposit larvae. The two membranes on the fly's hearing organ are close together and are mechanically coupled by a piece of tissue. Imitating this design in silicon and using optical sensors gives a directional microphone.

Carleen Maley Hutchins, now 92 years old, made her first viola at age 35, according to a story in the September 4 issue of *Granite State News* (New Hampshire). A major milestone in her career came when she was introduced to Frederick Saunders, retired Harvard physics professor, who continued his violin research in Mt. Holyoke, Massachusetts (see *ECHOES*, Spring 1997 and Summer 2003). Another important event occurred when Henry Brant, professor of musical composition at Bennington College in Vermont, encouraged her to make a set of seven graduated-size string instruments. Thus was born the New Violin Family (AKA the Hutchins Violin Octet) of eight scaled instruments, which appeared in a "splendiferous" concert in Wolfeboro, New Hampshire on September 20. Carleen received the Silver Medal in Musical Acoustics from ASA in 1981 and became an Honorary Fellow of the Society in 1998.

An experimental sonar, that uses frequencies above the hearing range of whales, has been used to detect Pacific Gray whales without causing them to break away from their migratory path or show signs of injury, according to a February 2 story on *Newsday.com*. It has also invoked new debate. Supporters say a reliable high-frequency sonar could help protect whales from a variety of ocean hazards, such as long range military sonar, collisions with ships, underwater demolitions, and seismic mapping by oil and gas companies. Opponents worry that the sonar could distress the whales, drive them from their habitat, or separate migrating calves from their mothers.

From the manufacturer of Scamp® Sound Masking Systems comes the revolutionary LogiSon™ Acoustic Network: the first and only sound masking, paging and music system to provide full, digital control of individual speaker settings from a central location.

Acoustic Control
At Your Fingertips™

The Network offers the highest level of component integration in the industry. Each hub contains a random masking sound generator, amplifier and even independent volume and equalizer controls for masking and paging, eliminating the need for most centralized audio equipment.

Timer and paging zones, volume and equalizer settings, and paging channel selection are configured through a centrally-located control panel or remotely with user-friendly LogiSon™ Acoustic Network Manager software.

This digital contol and accuracy, combined with small zone sizes and one-third octave equalization capabilities, allows you to custom tune the Network to suit the unique requirements of the space, increasing the masking's effectiveness.

And, as your needs change, the Network can be completely re-adjusted without rewiring or requiring access to the ceiling.

Though the LogiSon Network is typically installed above the suspended ceiling, it has been designed for visible applications and will compliment the investment you make in the professional appearance of your facility.

For more information, call 1 (866) LOGISON or visit www.logison.com to request an Information Package.

Better testing...
better products.

The Blachford Acoustics Laboratory

Bringing you superior acoustical products from the most advanced testing facilities available.
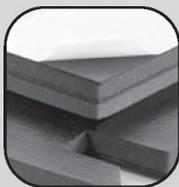
Our newest resource offers an unprecedented means of better understanding acoustical make-up and the impact of noise sources. The result? Better differentiation and value-added products for our customers.

Blachford Acoustics Laboratory features

- Hemi-anechoic room and dynamometer for testing heavy trucks and large vehicles or machines.
- Reverberation room for the testing of acoustical materials and components in one place.
- Jury room for sound quality development.

Blachford acoustical products

- Design and production of simple and complex laminates in various shapes, thicknesses and weights.
- Provide customers with everything from custom-engineered rolls and diecuts to molded and cast-in-place materials.

Blachford QS 9000 REGISTERED

www.blachford.com | Ontario 905.823.3200 | Illinois 630.231.8300

Blachford

ACOUSTICS LABORATORY

Canadian Acoustical Association
Minutes of the Board of Directors Meeting
30 May 2004

Aiolos Engineering Inc., 2150 Islington Ave.
Toronto, Ontario

Present: S. Dosso, D. Giusti, D. Quirt, C. Buma, M. Hodge, R. Ramakrishnan, A. Behar, , C. Giguère, V. Parsa, J. Bradley

Regrets: D. Stredulinsky, M. Cheng, R. Panneton

The meeting was called to order at 10:00 a.m. Minutes of Board of Directors meeting on 16 October 2003 were approved as published in Canadian Acoustics (Dec. 2003 issue).

(Approval moved by R. Ramakrishnan, seconded by D. Giusti, carried).

President's Report

Stan Dosso commented that from his initial review of current activities, CAA seems to be running well. He identified some issues to be added to the agenda under Other Business.

Treasurer's Report

Dalila Giusti provided an itemized report of the Association's finances, including a summary for the last four years. The report shows a solid financial position. As of 30 May, total assets are $258,902, most of which is invested to fund our awards. However, due to low interest rates, continuing management of new expenses is needed, to ensure sufficient funds are available for prizes. In 2003/04, only $4400 was distributed for awards, because the Shaw and Bell prizes were not awarded.

In the last 3 years, revenues have exceeded operating costs and the same is projected for this fiscal year. Conferences have consistently achieved a financial surplus (see further comments under conferences) although the surplus in the last two years was modest.

Our VISA merchant's account continues to be used for annual membership dues. This is working well except for occasional problems with incorrect numbers. Despite the charge for incorrect numbers in batch submissions, the Board recommended that this mode be continued to minimize administrative routine.

The Treasurer reiterated the hope that her successor will be recruited soon. Other board members expressed their satisfaction with the status quo.

(Acceptance moved by A. Behar, seconded V. Parsa, carried)

Secretary's Report

David Quirt reported on membership and operation of the Association. In brief, there are no glaring problems. Last year's report noted that the paid membership had returned to the range (300-320) typical of the preceding 5 years, but the number of Members from Canada is up slightly this year. The number of Sustaining Subscribers also rose slightly. To promote renewals, stamped return envelopes were included with Canadian invoices, and reminders were sent in April to members who had not responded to the December invoice. Each year, over 10% of members simply did not renew; half of these were students, who presumably shift to other interests, but many of the others drift back a year or two later. This year's modest increase in membership reflects the current push to overcome that lethargy.

Other administrative issues:

- Payments by VISA have increased to about 35% of the renewals; this is probably contributing to the better renewal rate, but seems to generate more processing problems for the Treasurer.

- Circulation of INCE quarterly news magazine *Noise News International* to the Canadian Members and Sustaining Subscribers began, as an option with the annual membership renewal. Thirty-eight members requested the magazine, and the $5 fee covers most of the mailing cost. Unfortunately, bulk transfer from the printer is sometimes late.

- Secretarial operating costs for the first nine months of FY03/04 were $939, which includes mailing costs and maintaining the address database including the annual membership renewal process. The secretarial account balance should cover most expected costs to the end of the fiscal year (31 August). However, it was agreed that the Treasurer should transfer an additional $500 to provide an operating margin.

(Acceptance moved by D. Giusti, seconded R. Ramakrishnan, carried)

Editor's Report

Ramani Ramakrishnan presented a number of issues related to publication of *Canadian Acoustics*. Major issues were:

- Reprints are now available for authors who request them, and a sample was circulated. The Board recommended also providing the corresponding pdf file to the author, and setting pricing to cover typical mailing costs.

- It was noted that advertising in *Canadian Acoustics* has declined slightly, but the number of Sustaining Subscribers has increased. Steps to increase advertising were discussed. Another administrative concern is arranging translation, both for abstracts and announcements; expenses to deal with this as needed were confirmed as a normal part of the Editor's budget.
- Several aspects of content were also discussed. The June 2004 issue will be the proceedings of a symposium on underwater sound, with 20 refereed papers. The conference organizers will buy 100 copies for their participants, and much of the cost is covered by page charges to authors. This sort of special topic issue was strongly endorsed by the Directors. The Editor also noted that there is a significant backlog of submissions in the review process, so larger issues should be expected in December and thereafter.

(A. Behar moved acceptance of Editor's report, V. Parsa seconded, carried.)

Awards

Christian Giguère submitted a report for the Awards Committee. Good applicants have applied for all prizes except the Hétu Book Prize, and proposed changes for the travel support for conferences on underwater sound and signal processing are expected to provide more candidates. Winners will be announced in October. Some key details were:

- Major updates of award pages on the website have been made, to rationalize the English and French versions, and provide downloadable forms. The next priorities are a preamble to the Shaw Prize, plus pages for the conference presentation awards and travel subsidies for students. For the latter, the Board debated changes to the process. It was agreed that receipts will be required to substantiate expenses, and a task group (John Bradley, Christian Giguère, and Dalila Giusti) was established to propose a fair way to allocate the funds, to be decided at the next Board meeting.

- An e-mail template was used to help individual coordinators to send directed advertisements for their prizes. New templates are being created for *Canadian Acoustics*, beginning March 2005, to deal with all phases of the awards process from advertising to announcing winners.

- The Board agreed to change rules for applications. For most prizes (Bell, Eckel, Fessenden, Shaw, Hétu) the deadline moves from April 15 to April 30. For travel support for underwater sound and signal processing conferences, two deadlines (September 30 and March 31) will permit two rounds of awards better matched to the academic planning cycle. Specific changes to procedural details for the latter were also decided. All of these changes should be incorporated on the website by mid-

summer.

(D. Giusti moved acceptance of awards changes, C. Giguère seconded, carried.)

CAA Website

Dave Stredulinsky, our webmaster, submitted a report on recent progress in the CAA website (caa-aca.ca), which is hosted by Telus. Traffic on the site has steadily increased, and is now typically over 200 visits/day; in April there were visitors from 70 countries.

Board members agreed that the website has become the most accessible and complete repository for information about CAA. Content includes:

- information on CAA awards (Significant updates and translation improvements have been made this year. Board confirmed that for awards, the website should be treated as the primary source for CAA information.)
- the CAA operations manual (Board agreed it should not be translated. It was suggested that as specific sections of the website become the authoritative source, corresponding parts of the manual should be reduced to links to pertinent pages.)
- Sustaining Subscribers' page (Has slowly increasing listings.)
- membership and subscription forms (Primary method for joining CAA)
- job-posting page (This is frequently updated, and has both job postings and work wanted items. Board agreed no translation is required.)
- the site for the annual conference (This has become central to the organizing process, especially for submission of papers.)

The increasing reliance on the website places a duty on Board members to keep their parts of the information current, by sending updates to the webmaster. Overall, there was enthusiastic support for the steady improvements.

Past and Future Conferences

2003 Conference in Edmonton: Corjan Buma and Megan Hodge provided a verbal report on the final status for the 2003 meeting. There were 55 papers and nearly 100 registrants. Total receipts were $20,342.29, and this resulted in a small surplus. The Board commended several operational innovations, especially the remotely controlled tone to ensure consistent schedules in concurrent sessions, and the pre-session submission of PowerPoint files to avoid delays for computer swapping. Special thanks were extended to Steven Bilawchuk for his huge contribution to both the website and A/V support for the sessions.

2004 Conference in Ottawa: John Bradley reported that the Ottawa organizing team has proceeded with the key parts of organization for the conference. Some features are: banquet at the National Arts Centre, exhibits will last only 1.5 days,

two plenary speakers by distinguished Canadian researchers whose topics mesh with following organized sessions. See the meeting notice published in this issue (June 2004) of *Canadian Acoustics*.

2005 Conference in London: Vijay Parsa reported on preliminary arrangements for the conference, which Meg Cheesman will convene in London. The hotel and dates are not confirmed yet, but will be reported in October.

CAA members are also organizing the ASA conference in Vancouver (16-20 May 2005), which is officially co-sponsored by CAA. It was tentatively agreed that the spring Board meeting in 2005 should be held there during the conference. Stan Dosso will poll Board members, to confirm that this will permit adequate attendance without substantial travel subsidies.

2006 Conference: Possible sites were discussed. Quebec or Nova Scotia were the consensus choices. Stan Dosso will investigate local interest.

Other Business

Three items were discussed:

1. Nomination of Emeritus Members. Alberto Behar and David Quirt agreed to assemble a list of candidates, and to circulate this to the Board before the October meeting, where this will be discussed.

2. Appointment of a CAA member to attend the meeting of IIAV in St. Petersburg (Russia) in August was discussed. The President agreed to consult confirmed Canadian participants in the conference, and to designate a representative if feasible.

3. Desirability and extent of CAA relationship with advocacy associations (such as Noise Pollution Clearinghouse) was discussed. It was agreed that Mark Cheng should prepare and circulate a brief report to the Board to provide a policy framework, and assemble key data for decisions at the October meeting on those with links on the CAA website.

Adjournment

D. Giusti moved to adjourn the meeting, seconded by R. Ramakrishnan, carried. Meeting adjourned at 2:50 p.m.

Special Action Items Arising from the Meeting

S. Dosso: In collaboration with Past President (J.S. Bradley), identify candidates for expected vacancies in Executive and other Directors. For meetings, confirm that Vancouver will be acceptable as site for Board meeting in May 2005, and seek convener for 2006 conference.

Designate Canadian representative for IIAV meeting in August.

Mark Cheng: prepare report for October meeting on policy framework for relationship with advocacy associations

J.S. Bradley: Ensure Ottawa Conference organization proceeds, and report to next meeting of the Board.

D. Giusti: Collaborate with Ottawa conference committee to establish process for payments by VISA. Transfer funds to Secretary (done).

D. Quirt: Provide suitable e-mail and mailing lists to C. Giguère for mailings related to awards. Provide financial and membership data for auditor at end of August. With Alberto Behar, prepare a list of candidates for Emeritus Member.

Each Member: Review CAA website contents within agreed areas of responsibility, and send updates to Webmaster periodically.

There´s a *New Kid in Town*

Created for You

Brüel & Kjær presents its innovative 4th generation hand-held instrument for sound and vibration. Experienced users from all over the world assisted us in setting the requirements for the new Type 2250.

- Color Touch Screen is the easiest user interface ever
- Non-slip surfaces with contours designed to fit comfortably in any hand
- Incredible 120 dB measurement range so you can't mess up your measurement by selecting an improper measurement range
- User log-in so the meter is configured the way you want it to be, and templates to make it easy to find user defined setups
- Optional Applications for Frequency Analysis and Logging are seamlessly integrated with the standard integrating SLM application
- High Contrast display and "traffic light" indicator make it easy to determine the measurement status at a distance even in daylight
- SD and CF memory and USB connectivity make Type 2250 the state-of-the-art sound analyzer

Created, built and made for you personally, you'll find Type 2250 will make a wonderful difference to your work and measurements tasks.

For more info please go directly to
www.type2250.com

Type *2250*

Brüel & Kjær

CAA Annual Conference in Ottawa
October 6-8, 2004
www.caa-aca.ca/ottawa-2004.html

Third Announcement

The 2004 annual conference of the Canadian Acoustical Association will be held in Ottawa October 6-8, 2004. With a location in Ottawa, and the theme 'Acoustics: A National Issue', it should be one of our more significant conferences. You can participate in three days of three parallel sessions of papers on all areas of acoustics. In addition to various associated meetings, there will be tours of local acoustical laboratories. Mark your calendars and plan now to participate!

Special Sessions

We are planning many special sessions including invited and contributed papers. The planned special sessions and organisers include those listed below. Please contact the organiser to participate in one of these or the conference Technical Chair to add other special sessions.

"Outdoor Sound Propagation",
   contact: Cameron Sherry, CWSherry@aol.com
"Instrumentation and Measurements",
   contact: George Wong, George.Wong@nrc.ca
"Acoustical Materials: Simulation and Characterisation",
   contact: Raymond Panneton, Raymond.Panneton@USherbrooke.ca
"Acoustics of Educational Facilities",
   contact: John Bradley, John.Bradley@nrc.ca
"Audio Systems and Signal Processing",
   contact: Scott Norcross, Scott.Norcross@crc.ca
"Acoustic and Non-Acoustic Influences on Speech Understanding",
   contact: Kathy Pichora-Fuller, kpfuller@utm.utoronto.ca
"Hearing Aids",
   contact: Vijay Parsa, parsa@nca.uwo.ca
"Hearing and the Workplace",
   contact: Christian Giguère, cgiguere@uottawa.ca
"Musical Acoustics",
   contact: David Gerhard, david.gerhard@uregina.ca
"Underwater Sound",
   contact: Nicole Collison, Nicole.collison@drdc-rddc.gc.ca
"Signal Processing Applications",
   contact: Dave Havelock, David.Havelock@nrc.ca
"Guidelines for Environmental Noise",
   contact: Stephen Bly, Stephen_Bly@hc-sc.gc.ca
"Noise Emission Declaration for Machinery Noise",
   contact: Stephen Keith, Stephen_Keith@hc-sc.gc.ca

Deadlines

Abstracts	20 June 2004
Two-page papers	15 August 2004

Plenary Speakers

Two distinguished Canadian acousticians will give plenary lectures coordinated with two of the special sessions above. They are:

Dr. Gilles Daigle, "Outdoor Sound Propagation", and

Dr. Bruce Schneider, " From Acoustics to Cognition: Some Surprising Connections".

Submissions and Important Dates

Submissions on all aspects of acoustics are welcomed. The deadline for submission of abstracts is **20 June 2004**. They should be submitted by Email to **abstracts@caa-aca.ca** and should be no more than 250 words and include the usual contact information. Other details of requirements for abstracts will be posted on the CAA website. Notices of acceptance will be sent out shortly after this deadline.

The deadline for the subsequent submission of a two-page summary paper for the conference issue of Canadian Acoustics is **15 August 2004**. If you miss this deadline your two-page summary paper will not be included in the conference issue of Canadian Acoustics. This conference issue has become the archival record of new acoustical research activities in Canada each year. Make sure you are included!

Associated Events

-"What's New in Building Acoustics at IRC-NRC?"
  organiser: Dave Quirt (Friday AM, open to all CAA conference attendees)
-CSA Z107 Committee Acoustics and Noise Control meeting (Wednesday evening),
  contact: Cameron Sherry, CWSherry@aol.com
-Lab tours various NRC labs, Health Canada (Friday PM).

Exhibits

There will be a one and a half day exhibit of measurement equipment and other acoustical products. The exhibit will run all day Thursday and Friday morning (October 7-8). As usual the exhibit area will also be the central coffee break area. Please contact our exhibit coordinator for exhibitor information and sponsorship of various aspects of this meeting.

Student Participation

CAA has a very strong emphasis on encouraging students. Student members of CAA who make presentations can apply for travel support and can win one of a number of student presentation awards. See our website for details.

Venue and Accommodation

The conference will be held at the newly renovated and enlarged Lord Elgin hotel, centrally located in Ottawa a few blocks from Parliament Hill. Participants registering with the hotel by September 5, 2004 will receive a room rate of $128/night (single or double). (1-800-267-4298). Please stay at this hotel to be with your friends and to support CAA.

Hospitality

CAA conferences are always an opportunity to meet old friends and to make new ones over a coffee during the conference, or over a drink after the sessions are over. There are many nearby bars and restaurants and of course there will be a banquet as part of the conference. Why not make it a holiday too, and stay on to see the Fall colours in Ottawa and Gatineau?

Contacts

Convenor	John Bradley	(john.bradley@nrc.ca)
Technical Chair	Brad Gover	(brad.gover@nrc.ca)
Publicity	Christian Giguère	(cgiguere@uottawa.ca)
Exhibits	Hugh Williamson	(hughwilliamson@sympatico.ca)
Webmaster	Alf Warnock	(alf.warnock@nrc.ca)
Audio-Visual	Frances King	(frances.king@nrc.ca)

Congrès annuel de l'ACA à Ottawa
6 au 8 octobre 2004
www.caa-aca.ca/ottawa-2004.html

Troisième avis

Le congrès annuel de l'Association canadienne d'acoustique se tiendra à Ottawa du 6 au 8 octobre 2004. Avec comme site Ottawa et comme thème « L'Acoustique : Une question nationale », il s'agira sûrement d'un congrès des plus mémorables. Trois jours de communications scientifiques comprenant trois sessions parallèles sont prévus sur tous les domaines de l'acoustique. En plus des réunions habituelles, des visites de laboratoires seront au programme. Veuillez planifier dès maintenant de participer à cet événement !

Sessions spéciales

Des sessions spéciales seront structurées autour de conférenciers invités et des communications soumises par les délégués. Si vous désirez participer à l'une des sessions spéciales ci-dessous, veuillez communiquer avec le responsable de session. Pour organiser d'autres sessions spéciales, veuillez communiquer avec le Président du congrès ou le Directeur scientifique.

« Propagation sonore à l'extérieur »
responsable: Cameron Sherry, CWSherry@aol.com

« Instrumentation et Méthodes de mesures »
responsable: George Wong, George.Wong@nrc.ca

« Matériaux acoustiques : simulation et caractérisation »
responsable: Raymond Panneton, Raymond.Panneton@USherbrooke.ca

« L'Acoustique des établissements éducationnels »
responsable: John Bradley, John.Bradley@nrc.ca

« Systèmes audio et traitement du signal »
responsable: Scott Norcross, Scott.Norcross@crc.ca

« Aspects acoustiques et non-acoustiques de la perception de la parole »
responsable: Kathy Pichora-Fuller, kpfuller@utm.utoronto.ca

« Aides auditives »
responsable: Vijay Parsa, parsa@nca.uwo.ca

« Audition et Milieu de travail »
responsable: Christian Giguère, cgiguere@uottawa.ca

« Acoustique musicale »
responsable: David Gerhard, david.gerhard@uregina.ca

« Acoustique sous-marine »
responsable: Nicole Collison, Nicole.collison@drdc-rddc.gc.ca

« Applications du traitement du signal »
responsable Dave Havelock, David.Havelock@nrc.ca

« Lignes directrices pour le bruit environnemental »
responsable: Stephen Bly, Stephen_Bly@hc-sc.gc.ca

« Déclaration de l'émission sonore des équipements bruyants »
responsable: Stephen Keith, Stephen_Keith@hc-sc.gc.ca

Échéances

Résumés	20 juin 2004
Articles de deux pages	15 août 2004

Orateurs pléniers

Deux acousticiens canadiens renommés feront des présentations plénières dans le cadre de deux des sessions spéciales. Ils ont,

Dr. Gilles Daigle, "Outdoor Sound Propagation", and

Dr. Bruce Schneider, " From Acoustics to Cognition: Some Surprising Connections".

Appel de communications et Dates importantes

Les soumissions portant sur tous les domaines de l'acoustique sont les bienvenues. La date d'échéance pour la soumission de résumés est le **20 juin 2004**. Les résumés doivent être envoyés par courriel à **abstracts@caa-aca.ca**, inclure les renseignements habituels sur les auteurs et ne pas dépasser 250 mots. Les autres détails et exigences pour les résumés seront affichés sur le site Internet de l'ACA. Les avis d'acceptation seront envoyés peu après la date d'échéance. La date d'échéance pour la soumission de l'article de deux pages pour la revue Acoustique Canadienne, édition spéciale du congrès, est le **15 août 2004**. Si vous ne rencontrez pas cette échéance, votre article de deux pages ne sera pas publié dans l'édition spéciale du congrès. Cette édition spéciale est un portrait des nouvelles recherches en acoustique de l'année. Soyez certain d'en faire partie!

Événements particuliers

- Quoi de Neuf en Acoustique des Bâtiments à l'IRC-CNRC?
  organisateur: Dave Quirt (Vendredi en avant-midi, ouvert aux délégués du congrès de l'ACA)
- CSA Z107 Réunion du Comité en Acoustique et Contrôle du Bruit
  organisateur: Cameron Sherry, CWSherry@aol.com (mercredi soir)
- Visite de divers labos de le CNRC et de Santé-Canada (vendredi en après-midi).

Exposition technique

Il y aura une exposition d'instruments et d'autres produits en acoustique. L'exposition durera toute la journée jeudi et le vendredi en avant-midi (7-8 octobre). La salle d'exposition agira comme lieu central lors des pauses. Veuillez communiquer dès maintenant avec le coordonnateur de l'exposition pour de plus amples renseignements ou pour la commandite d'événements particuliers lors du congrès.

Participation étudiante

L'ACA accorde beaucoup d'importance à la participation des étudiants. Les membres étudiants qui présenteront une communication pourront soumettre une demande de subvention pour frais de déplacement au congrès et pourront se voir mériter l'un des prix pour communications étudiantes. Veuillez consulter notre site Internet.

Lieu du congrès et Hébergement

Le congrès se tiendra à l'hôtel Lord Elgin, tout récemment agrandi et rénové, situé au coeur d'Ottawa à proximité de la colline parlementaire. Les délégués qui réserveront leur chambre à cet hôtel (1-800 267-4298) avant le 5 septembre 2004 bénéficieront d'un tarif préférentiel de $128/nuit (occupation simple ou double). Choisissez cet hôtel pour participer pleinement au congrès et encourager l'ACA.

Votre séjour à Ottawa

Le congrès de l'ACA est toujours une excellente occasion de renouer avec vos collègues acousticiens et de faire de nouvelles connaissances. Il y a plusieurs bistros et restaurants à proximité de l'hôtel et il y aura aussi le banquet du congrès. Pourquoi aussi ne pas en profiter et rester un peu plus longtemps pour découvrir les couleurs d'automne de la région Ottawa-Gatineau?

Personnes contacts

Président	John Bradley	(john.bradley@nrc.ca)
Directeur scientifique	Brad Gover	(brad.gover@nrc.ca)
Publicité	Christian Giguère	(cgiguere@uottawa.ca)
Exposition	Hugh Williamson	(hughwilliamson@sympatico.ca)
Webmaster	Alf Warnock	(alf.warnock@nrc.ca)
Audio-Visuel	Frances King	(frances.king@nrc.ca)

ACOUSTICS WEEK IN CANADA 2004
SEMAINE CANADIENNE D'ACOUSTIQUE 2004

October 6-8, 2004 / *du 6 au 8 octobre 2004*
The Lord Elgin Hotel, Ottawa

REGISTRATION FORM / *FORMULAIRE D'INSCRIPTION*

(1) Full Three Day Registration. Includes:

Conference + exhibits
Lunch each day for 3 days
Coffee breaks, morning and afternoon
Entertainment
Banquet Thursday night (except students)
All taxes & gratuities

(1) Inscription complète. *Comprend:*

La participation à la conférence + l'exhibition
Le dîner pendant 3 jours
Les pauses café, le matin et l'après-midi
L'hospitalité/le divertissement
Le banquet du jeudi soir (sauf étudiants)
Toutes les taxes et les pourboires

Registration/ *Inscription*	CAA Members *Membres de l'ACA*	Non-members *Autres*	Students *Etudiant(e)s*	
			CAA/ ACA	Others *Autres*
Before Sept. 1/ *Avant le 1 sept.*	$300.00	$360.00*	$30.00	$50.00*
After Sept. 1/ *Après le 1 sept.*	$330.00	$360.00*	$30.00	$50.00*

*non-member registration includes a 1 year CAA membership
*inclut l'adhésion à l'ACA pendant un an

(2) Daily Rates. Includes:

Conference
Lunch for one day
All taxes & gratuities

(2) Tarif à la journée. *Comprend:*

Une journée à la conférence
Le dîner pour une journée
Toutes les taxes et les pourboires

CAA Members *Membres de l'ACA*	Students *Etudiant(e)s*	Non-Members *Autres*
$125.00	$25.00	$125.00

Note: All Conference passes are non-transferable.
Les billets d'accès à la conférence sont personnels et ne peuvent être transférés.

(3) Extras/ *Suppléments*

Student Banquet Tickets / *Etudiant Billet pour le Banquet* $20.00 each/*par personne*

Additional Banquet Ticket / $50.00 each/*par personne*
Billet Supplémentaire pour le Banquet

The Canadian Acoustical Association
l'Association Canadienne d'Acoustique

REGISTRATION FORM / *FORMULAIRE D'INSCRIPTION*

Name/*Nom*:_____

Company/*Institution*:_____

Address/*Adresse*:_____

_____Postal Code/*Code Postal*:_____

Tel:_____E-mail/*courriel*:_____

Full 3 day Conference	$_____	***Inscription complète***
		Billet(s)
Additional banquet ticket(s)	$_____	*supplémentaire(s) pour le banquet*
Total	$_____	*Total*

Daily Rate	Wednesday *mercredi*	Thursday *jeudi*	Friday *vendredi*	***Inscription à la journée***
Check applicable day(s)				*Entourez le(s) jour(s) choisi(s)*
Daily rate	$_____			*Montant*
Banquet ticket(s)	$_____			*Billet(s) pour le banquet*
Total	$_____			*Total*

Payable by cheque in Canadian Funds made out to CAA Conference 2004, or payable by **VISA**
***Payable par chèque, en dollars canadiens, à l'ordre de* CAA Conference 2004*,** ou par carte* **VISA**

VISA Number/*Numéro*:_____ Expiry Date/*Date d'expiration*:_____

Name on VISA card / *Nom sur la carte VISA* : _____

Signature (if paying by VISA / *en cas de paiement par VISA)* :_____

Mail/Fax to:	Dr. Brad Gover
Expédier ou faxer à:	CAA-ACA 2004.
	P.O. Box 74068
	Ottawa, Ontario
	K1M 2H9
	Fax: (613) 954-1495
	Tel: (613) 993-7985

Please note: You must book your own hotel room by Sept. 5, 2004 to guarantee availability.
Veuillez noter : Pour des questions de disponibilité, il est recommandé de réserver votre chambre d'hôtel avant le 5 septembre 2004.

The Canadian Acoustical Association l'Association Canadienne d'Acoustique

The Canadian Acoustical Association
L'Association Canadienne d'Acoustique

PRIZE ANNOUNCEMENT • ANNONCE DE PRIX

A number of prizes and subsidies are offered annually by The Canadian Acoustical Association. Applicants can obtain full eligibility conditions, deadlines, application forms, past recipients, and the names of the individual prize coordinators on the CAA Website (http://www.caa-aca.ca). • Plusieurs prix et subventions sont décernés à chaque année par l'Association Canadienne d'Acoustique. Les candidats peuvent se procurer de plus amples renseignements sur les conditions d'éligibilités, les échéances, les formulaires de demande, les récipiendaires des années passées ainsi que le nom des coordonnateurs des prix en consultant le site Internet de l'ACA (http://www.caa-aca.ca).

CAA conference Student Travel subsidies: **consult www.ottawa2004.ca**

Subventions pour étudiants pour frais de déplacement au congrès annuel de l'ACA : **consulter le www.ottawa2004.ca**

EDGAR AND MILLICENT SHAW POSTDOCTORAL PRIZE IN ACOUSTICS • PRIX POST-DOCTORAL EDGAR AND MILLICENT SHAW EN ACOUSTIQUE

$3,000 for full-time postdoctoral research training in an established setting other than the one in which the Ph.D. was earned. The research topic must be related to some area of acoustics, psychoacoustics, speech communication or noise. • $3,000 pour une formation recherche à temps complet au niveau postdoctoral dans un établissement reconnu autre que celui où le candidat a reçu son doctorat. Le thème de recherche doit être relié à un domaine de l'acoustique, de la psycho-acoustique, de la communication verbale ou du bruit.

ALEXANDER GRAHAM BELL GRADUATE STUDENT PRIZE IN SPEECH COMMUNICATION AND BEHAVIOURAL ACOUSTICS • PRIX ÉTUDIANT ALEXANDRE GRAHAM BELL EN COMMUNICATION VERBALE ET ACOUSTIQUE COMPORTEMENTALE

$800 for a graduate student enrolled at a Canadian academic institution and conducting research in the field of speech communication or behavioural acoustics. • $800 à un(e) étudiant(e) inscrit(e) au 2e ou 3e cycle dans une institution académique canadienne et menant un projet de recherche en communication verbale ou acoustique comportementale.

FESSENDEN GRADUATE STUDENT PRIZE IN UNDERWATER ACOUSTICS • PRIX ÉTUDIANT FESSENDEN EN ACOUSTIQUE SOUS-MARINE

$500 for a graduate student enrolled at a Canadian academic institution and conducting research in underwater acoustics or in a branch of science closely connected to underwater acoustics. • $500 à un(e) étudiant(e) inscrit(e) au 2e ou 3e cycle dans une institution académique canadienne et menant un projet de recherche en acoustique sous-marine ou dans une discipline reliée à l'acoustique sous-marine.

ECKEL GRADUATE STUDENT PRIZE IN NOISE CONTROL • PRIX ÉTUDIANT ECKEL EN CONTRÔLE DU BRUIT

$500 for a graduate student enrolled at a Canadian academic institution and conducting research related to the advancement of the practice of noise control. • $500 à un(e) étudiant(e) inscrit(e) au 2e ou 3e cycle dans une institution académique canadienne et menant un projet de recherche relié à l'avancement de la pratique du contrôle du bruit.

RAYMOND HÉTU UNDERGRADUATE PRIZE IN ACOUSTICS • PRIX ÉTUDIANT RAYMOND HÉTU EN ACOUSTIQUE

One book in acoustics of a maximum value of $100 and a one-year subscription to *Canadian Acoustics* for an undergraduate student enrolled at a Canadian academic institution and having completed, during the year of application, a project in any field of acoustics or vibration. • Un livre sur l'acoustique et un abonnement d'un an à la revue *Acoustique Canadienne* à un(e) étudiant(e) inscrit(e) dans un programme de 1er cycle dans une institution académique canadienne et qui a réalisé, durant l'année de la demande, un projet dans le domaine de l'acoustique ou des vibrations.

CANADA-WIDE SCIENCE FAIR AWARD • PRIX EXPO-SCIENCES PANCANADIENNE

$400 and a one-year subscription to *Canadian Acoustics* for the best project related to acoustics at the Fair by a high-school student • $400 et un abonnement d'un an à la revue *Acoustique Canadienne* pour le meilleur projet relié à l'acoustique à l'Expo-sciences par un(e) étudiant(e) du secondaire.

DIRECTORS' AWARDS • PRIX DES DIRECTEURS

One $500 award for the best refereed research, review or tutorial paper published in *Canadian Acoustics* by a student member and one $500 award for the best paper by an individual member • $500 pour le meilleur article de recherche, de recensement des travaux ou d'exposé didactique arbitré publié dans l'*Acoustique Canadienne* par un membre étudiant et $500 pour le meilleur article par un membre individuel.

STUDENT PRESENTATION AWARDS • PRIX POUR COMMUNICATIONS ÉTUDIANTES

Three $500 awards for the best student oral presentations at the Annual Symposium of The Canadian Acoustical Association. • Trois prix de $500 pour les meilleures communications orales étudiant(e)s au Symposium Annuel de l'Association Canadienne d'Acoustique.

STUDENT TRAVEL SUBSIDIES • SUBVENTIONS POUR FRAIS DE DÉPLACEMENT POUR ÉTUDIANTS

Travel subsidies are available to assist student members who are presenting a paper during the Annual Symposium of The Canadian Acoustical Association if they live at least 150 km from the conference venue. • Des subventions pour frais de déplacement sont disponibles pour aider les membres étudiants à venir présenter leurs travaux lors du Symposium Annuel de l'Association Canadienne d'Acoustique, s'ils demeurent à au moins 150 km du lieu du congrès.

UNDERWATER ACOUSTICS AND SIGNAL PROCESSING STUDENT TRAVEL SUBSIDIES • SUBVENTIONS POUR FRAIS DE DÉPLACEMENT POUR ÉTUDIANTS EN ACOUSTIQUE SOUS-MARINE ET TRAITEMENT DU SIGNAL

One $500 or two $250 awards to assist students traveling to national or international conferences to give oral or poster presentations on underwater acoustics and/or signal processing. • Une bourse de $500 ou deux de $250 pour aider les étudiant(e)s à se rendre à un congrès national ou international pour y présenter une communication orale ou une affiche dans le domaine de l'acoustique sous-marine ou du traitement du signal.

WHAT'S NEW ??   QUOI DE NEUF ?

Promotions	Retirements	Promotions	Retraites
Deaths	Degrees awarded	Décès	Obtention de diplômes
New jobs	Distinctions	Offre d'emploi	Distinctions
Moves	Other news	Déménagements	Autres nouvelles

Do you have any news that you would like to share with Canadian Acoustics readers? If so, send it to:

Avez-vous des nouvelles que vous aimeriez partager avec les lecteurs de l'Acoustique Canadienne? Si oui, écrivez-les et envoyer à:

Steven Bilawchuk, aci Acoustical Consultants Inc., Edmonton, Alberta, Email: **stevenb@aciacoustical.com**

INSTRUCTIONS TO AUTHORS
FOR THE PREPARATION OF MANUSCRIPTS

Submissions: The original manuscript and two copies should be sent to the Editor-in-Chief.

General Presentation: Papers should be submitted in camera-ready format. Paper size 8.5" x 11". If you have access to a word processor, copy as closely as possible the format of the articles in Canadian Acoustics 18(4) 1990. All text in Times-Roman 10 pt font, with single (12 pt) spacing. Main body of text in two columns separated by 0.25". One line space between paragraphs.

Margins: Top - title page: 1.25"; other pages, 0.75"; bottom, 1" minimum; sides, 0.75".

Title: Bold, 14 pt with 14 pt spacing, upper case, centered.

Authors/addresses: Names and full mailing addresses, 10 pt with single (12 pt) spacing, upper and lower case, centered. Names in bold text.

Abstracts: English and French versions. Headings, 12 pt bold, upper case, centered. Indent text 0.5" on both sides.

Headings: Headings to be in 12 pt bold, Times-Roman font. Number at the left margin and indent text 0.5". Main headings, numbered as 1, 2, 3, ... to be in upper case. Sub-headings numbered as 1.1, 1.2, 1.3, ... in upper and lower case. Sub-sub-headings not numbered, in upper and lower case, underlined.

Equations: Minimize. Place in text if short. Numbered.

Figures/Tables: Keep small. Insert in text at top or bottom of page. Name as "Figure 1, 2, ..." Caption in 9 pt with single (12 pt) spacing. Leave 0.5" between text.

Line Widths: Line widths in techincal drawings, figures and tables should be a minimum of 0.5 pt.

Photographs: Submit original glossy, black and white photograph.

Scans: Should be between 225 dpi and 300 dpi. Scan: Line art as bitmap tiffs; Black and white as grayscale tiffs and colour as CMYK tiffs;

References: Cite in text and list at end in any consistent format, 9 pt with single (12 pt) spacing.

Page numbers: In light pencil at the bottom of each page. Reprints: Can be ordered at time of acceptance of paper.

DIRECTIVES A L'INTENTION
DES AUTEURS PREPARATION DES MANUSCRITS

Soumissions: Le manuscrit original ainsi que deux copies doivent être soumis au rédacteur-en-chef.

Présentation générale: Le manuscrit doit comprendre le collage. Dimensions des pages, 8.5" x 11". Si vous avez accès à un système de traitement de texte, dans la mesure du possible, suivre le format des articles dans l'Acoustique Canadienne 18(4) 1990. Tout le texte doit être en caractères Times-Roman, 10 pt et à simple (12 pt) interligne. Le texte principal doit être en deux colonnes séparées d'un espace de 0.25". Les paragraphes sont séparés d'un espace d'une ligne.

Marges: Dans le haut - page titre, 1.25"; autres pages, 0.75"; dans le bas, 1" minimum; latérales, 0.75".

Titre du manuscrit: 14 pt à 14 pt interligne, lettres majuscules, caractères gras. Centré.

Auteurs/adresses: Noms et adresses postales. Lettres majuscules et minuscules, 10 pt à simple (12 pt) interligne. Centré. Les noms doivent être en caractères gras.

Sommaire: En versions anglaise et française. Titre en 12 pt, lettres majuscules, caractères gras, centré. Paragraphe 0.5" en alinéa de la marge, des 2 cotés.

Titres des sections: Tous en caractères gras, 12 pt, Times-Roman. Premiers titres: numéroter 1, 2, 3, ..., en lettres majuscules; sous-titres: numéroter 1.1, 1.2, 1.3, ..., en lettres majuscules et minuscules; sous-sous-titres: ne pas numéroter, en lettres majuscules et minuscules et soulignés.

Equations: Les minimiser. Les insérer dans le texte si elles sont courtes. Les numéroter.

Figures/Tableaux: De petites tailles. Les insérer dans le texte dans le haut ou dans le bas de la page. Les nommer "Figure 1, 2, 3,..." Légende en 9 pt à simple (12 pt) interligne. Laisser un espace de 0.5" entre le texte.

Largeur Des Traits: La largeur des traits sur les schémas technique doivent être au minimum de 0.5 pt pour permettre une bonne reproduction.

Photographies: Soumettre la photographie originale sur papier glacé, noir et blanc.

Figures Scanées: Doivent être au minimum de 225 dpi et au maximum de 300 dpi. Les schémas doivent être scannés en bitmaps tif format. Les photos noir et blanc doivent être scannées en échelle de gris tifs et toutes les phoots couleurs doivent être scannées en CMYK tifs.

Références: Les citer dans le texte et en faire la liste à la fin du document, en format uniforme, 9 pt à simple (12 pt) interligne.

Pagination: Au crayon pâle, au bas de chaque page. Tirés-à-part: Ils peuvent être commandés au moment de l'acceptation du manuscrit.

The Canadian Acoustical Association
l'Association Canadienne d'Acoustique

Application for Membership

CAA membership is open to all individuals who have an interest in acoustics. Annual dues total $55.00 for individual members and $15.00 for Student members. This includes a subscription to *Canadian Acoustics*, the Association's journal, which is published 4 times/year. New membership applications received before September 1 will be applied to the current year and include that year's back issues of *Canadian Acoustics*, if available. New membership applications received after September 1 will be applied to the next year.

Subscriptions to *Canadian Acoustics* or Sustaining Subscriptions

Subscriptions to *Canadian Acoustics* are available to companies and institutions at the institutional subscription price of $55.00. Many companies and institutions prefer to be a Sustaining Subscriber, paying $250.00 per year, in order to assist CAA financially. A list of Sustaining Subscribers is published in each issue of *Canadian Acoustics*. Subscriptions for the current calendar year are due by January 31. New subscriptions received before September 1 will be applied to the current year and include that year's back issues of *Canadian Acoustics*, if available.

Please note that electronic forms can be downloaded from the CAA Website at **caa-aca.ca**

Address for subscription / membership correspondence:

Name / Organization _____

Address _____

City/Province _____ Postal Code _____ Country _____

Phone _____ Fax _____ E-mail _____

Address for mailing Canadian Acoustics, if different from above:

Name / Organization _____

Address _____

City/Province _____ Postal Code _____ Country _____

Areas of Interest: (Please mark 3 maximum)

1. Architectural Acoustics
2. Engineering Acoustics / Noise Control
3. Physical Acoustics / Ultrasound
4. Musical Acoustics / Electro-acoustics

5. Psychological / Physiological Acoustic
6. Shock and Vibration
7. Hearing Sciences
8. Speech Sciences

9. Underwater Acoustics
10. Signal Processing / Numerical Methods
11. Other

For student membership, please also provide:

_____ _____ _____ _____
(University)   *(Faculty Member)*   *(Signature of Faculty Member)*   *(Date)*

I have enclosed the indicated payment for:
[ ] CAA Membership $ 55.00
[ ] CAA Student Membership $ 15.00
[ ] Institutional Subscription $ 55.00
[ ] Sustaining Subscriber $ 250.00
   includes subscription (4 issues /year)
   to *Canadian Acoustics.*

Payment by: [ ] Cheque
            [ ] Money Order
            [ ] VISA credit card *(Only VISA accepted)*

For payment by VISA credit card:
   Card number _____
   Name of cardholder _____
   Expiry date _____

   _____
   (Signature)   (Date)

Mail this application and attached payment to:
D. Quirt, Secretary, Canadian Acoustical Association, PO Box 74068, Ottawa, Ontario, K1M 2H9, Canada

l'Association Canadienne d'Acoustique
The Canadian Acoustical Association

Formulaire d'adhésion

L'adhésion à l'ACA est ouverte à tous ceux qui s'intéressent à l'acoustique. La cotisation annuelle est de 55.00$ pour les membres individuels, et de 15.00$ pour les étudiants. Tous les membres reçoivent *l'Acoustique Canadienne*, la revue de l'association. Les nouveaux abonnements reçus avant le 1 septembre s'appliquent à l'année courante et incluent les anciens numéros (non-épuisés) de *l'Acoustique Canadienne* de cette année. Les nouveaux abonnements reçus après le 1 septembre s'appliquent à l'année suivante.

Abonnement pour la revue *Acoustique Canadienne* et abonnement de soutien

Les abonnements pour la revue *Acoustique Canadienne* sont disponibles pour les compagnies et autres établissements au coût annuel de 55.00$. Des compagnies et établissements préfèrent souvent la cotisation de membre bienfaiteur, de 250.00$ par année, pour assister financièrement l'ACA. La liste des membres bienfaiteurs est publiée dans chaque issue de la revue *Acoustique Canadienne*. Les nouveaux abonnements reçus avant le 1 juillet s'appliquent à l'année courante et incluent les anciens numéros (non-épuisés) de *l'Acoustique Canadienne* de cette année. Les nouveaux abonnements reçus après le 1 septembre s'appliquent à l'année suivante.

Pour obtenir des formulaires electroniques, visitez le site Web: **caa-aca.ca**

Pour correspondence administative et financière:

Nom / Organisation _____

Adresse _____

Ville/Province _____ Code postal _____ Pays _____

Téléphone _____ Téléc. _____ Courriel _____

Adresse postale pour la revue *Acoustique Canadienne*

Nom / Organisation _____

Adresse _____

Ville/Province _____ Code postal _____ Pays _____

Cocher vos champs d'intérêt: (maximum 3)

1. Acoustique architecturale
2. Génie acoustique / Contrôle du bruit
3. Acoustique physique / Ultrasons
4. Acoustique musicale / Electro-acoustique

5. Physio / Psycho-acoustique
6. Chocs et vibrations
7. Audition
8. Parole

9. Acoustique sous-marine
10. Traitement des signaux /Méthodes numériques
11. Autre

Prière de remplir pour les étudiants et étudiantes:

_____ _____ _____ _____
(Université) (Nom d'un membre du corps professoral) (Signature du membre du corps professoral)   (Date)

Cocher la case appropriée:
[   ] Membre individuel $ 55.00
[   ] Membre étudiant(e) $ 15.00
[   ] Abonnement institutionnel $ 55.00
[   ] Abonnement de soutien $ 250.00
     (comprend l'abonnement à
     L'acoustique Canadienne)

Méthode de paiement:
[   ] Chèque au nom de l'Association Canadienne d'Acoustique
[   ] Mandat postal
[   ] VISA — *(Seulement VISA)*
     Pour carte VISA: Carte n° _____
                         Nom _____
                         Date d'expiration _____

                         _____
                         (Signature)   (Date)

Prière d'attacher votre paiement au formulaire d'adhésion. Envoyer à:
D. Quirt, Secrétaire exécutif, Association Canadienne d'Acoustique, Casier Postal 74068, Ottawa, K1M 2H9, Canada

The Canadian Acoustical Association
l'Association Canadienne d'Acoustique

PRESIDENT **PRÉSIDENT**	**Stan Dosso** University of Victoria Victoria, British Columbia V8W 3P6	(250) 472-4341 sdosso@uvic.ca
PAST PRESIDENT **PRÉSIDENT SORTANT**	**John Bradley** IRC, NRCC Ottawa, Ontario K1A 0R6	(613) 993-9747 john.bradley@nrc-cnrc.gc.ca
SECRETARY **SECRÉTAIRE**	**David Quirt** P. O. Box 74068 Ottawa, Ontario K1M 2H9	(613) 993-9746 dave.quirt@nrc-cnrc.gc.ca
TREASURER **TRÉSORIER**	**Dalila Giusti** Jade Acoustics 545 North Rivermede Road, Suite 203 Concord, Ontario L4K 4H1	(905) 660-2444 dalila@jadeacoustics.com
ADVERTISING **PUBLICITÉ**	**Brian Chapnik** HGC Engineering Ltd. 2000 Argentia Road, Plaza 1 Suite 203 Mississauga, Ontario L5N 1P7	(905) 826-4044 bchapnik@hgcengineeering.com
EDITOR-IN-CHIEF **RÉDACTEUR EN CHEF**	**Ramani Ramakrishnan** Dept. of Architectural Science Ryerson University 350 Victoria Street Toronto, Ontario M5B 2K3	(416) 979-5000 #6508 rramakri@ryerson.ca ramani@aiolos.com
DIRECTORS **DIRECTEURS**	**Alberto Behar** **Corjan Buma** **Mark Cheng** **Christian Giguère** **Megan Hodge** **Raymond Panneton** **Vijay Parsa** **Dave Stredulinsky**	(416) 265-1816 (780) 435-9172 (604) 276-6366 613-562-5800 Ext 3071 (780) 492-5898 (819) 821-8000 (519) 661-2111 Ex. 88947 (902) 426-3100
WORLD WIDE WEB HOME PAGE:	**http://www.caa-aca.ca** **Dave Stredulinsky**	(902) 426-3100

SUSTAINING SUBSCRIBERS / ABONNÉS DE SOUTIEN

The Canadian Acoustical Association gratefully acknowledges the financial assistance of the Sustaining Subscribers listed below. Annual donations (of $250.00 or more) enable the journal to be distributed to all at a reasonable cost. Sustaining Subscribers receive the journal free of charge. Please address donation (made payable to the Canadian Acoustical Association) to the Secretary of the Association.

L'Association Canadienne d'Acoustique tient à témoigner sa reconnaissance à l'égard de ses Abonnés de Soutien en publiant ci-dessous leur nom et leur adresse. En amortissant les coûts de publication et de distribution, les dons annuels (de $250.00 et plus) rendent le journal accessible à tous nos membres. Les Abonnés de Soutien reçoivent le journal gratuitement. Pour devenir un Abonné de Soutien, faites parvenir vos dons (chèque ou mandat-poste fait au nom de l'Association Canadienne d'Acoustique) au secrétaire de l'Association.

ACI Acoustical Consultants Inc.
Mr. Steven Bilawchuk - (780) 414-6373
stevenb@aciacoustical.com - Edmonton, AB

ACO Pacific
Mr. Noland Lewis - (650) 595-8588
acopac@acopacific.com - Belmont, CA

Acoustec Inc.
Dr. J.G. Migneron - (418) 682-2331
courrier@acoustec.qc.ca - Montréal, QC

Acoustik GE Inc.
C/o Gilles Elhadad - (514) 487
7159ge@acoustikge.com - Cote St Luc, QC

Aercoustics Engineering Limited
Mr. John O'Keefe - (416) 249-3361
aercoustics@aercoustics.com - Toronto, ON

Bruel & Kjaer International
Mr. Andrew Khoury - (514) 695-8225
- Pointe-Claire, QC

Dalimar Instruments Inc.
Mr. Daniel Larose - (450) 424-0033
daniel@dalimar.ca - Vaudreuil-Dorion, QC

Dodge-Regupol Inc.
Mr. Paul Downey - 416-440-1094
pcd@regupol.com - Toronto, ON

Earth Tech Canada Inc.
Mr. Christopher Hugh - (905) 886-7022,2625
chris.hugh@earthtech.ca - Markham, ON

Eckel Industries of Canada Ltd.
Mr. Blake Noon - (613) 543-2967
eckel@eckel.ca - Morrisburgh, ON

H. L. Blachford Ltd.
Mr. Dalton Prince - (905) 823-3200
amsales@blachford.ca -Mississauga ON

Hatch Associates Ltd.
Mr. Tim Kelsall - (905) 403-3932
tkelsall@hatch.ca - Mississauga, ON

HGC Engineering
Mr. Bill Gastmeier - (905) 826-4044
info@hgcengineering.com -Mississauga, ON

Hydro-Quebec
Mr. Blaise Gosselin - (514) 840-3000ext5134
gosselin.blaise@hydro.qc.ca - Montréal, QC

Industrial Metal Fabricators Ltd.
Mr. Frank Van Oirschot - (519) 354-4270
frank@indmetalfab.com -Chatham, ON

Integral DX Engineering Ltd.
Mr. Greg Clunis - (613) 761-1565
au741@ncf.ca - Ottawa, ON

J. E. Coulter Associates Ltd.
Mr. John Coulter - (416) 502-8598
jcoulter@on.aibn.com - Toronto, ON

J. L. Richards & Assoc. Ltd.
Mr. Fernando Ribas - (613) 728-3571
mail@jlrichards.ca -Ottawa, ON

Jade Acoustics Inc.
Ms. Dalila Giusti - (905) 660-2444
dalila@jadeacoustics.com - Concord, ON

John Swallow Associates
Mr. John C. Swallow - 905-271-7888
jswallow@jsal.ca - Mississauga, ON

MJM Conseillers en Acoustique
M. Michel Morin - (514) 737-9811
mmorin@mjm.qc.ca - Montréal, QC

Novel Dynamics Test Inc.
Mr. Andy Metelka - (519) 853-4495
metalka@aztec-net.com - Acton, ON

Owens Corning Canada Inc.
Mr. Keith Wilson - (705) 428-6774
keith.wilson@owenscorning.com-Stayner,ON

OZA Inspections Ltd.
Mr. Gord Shaw - (905) 945-5471
oza@ozagroup.com - Grimsby, ON

Peutz & Associés
Mr. Marc Asselineau - +33 1 45 23 05 00
marc.asselineau@club-internet.fr - Paris, France

Michel Picard
(514) 343-7617; FAX: (514) 343-2115
michel.picard@umontreal.ca - Brossard, QC

Pyrok Inc.
Mr. Howard Podolsky - (914) 777-7770
info@pyrokinc.com - Mamaroneck, NY US

Scantek Inc.
Mr. Peppin - (410) 290-7726
info@scantekinc.com - Columbia, MD

Silex Inc.
Mr. Mehmoud Ahmed - (905) 612-4000
mehmouda@silex.com - Mississauga, ON

SNC/Lavalin Environment Inc.
Mr. J-L. Allard - (514) 651-6710
jeanluc.allard@snclavalin.com-Longueuil,QC

Soft dB Inc.
M. André L'Espérance - (418) 686-0993
contact@softdb.com - Sillery, Québec

Spaarg Engineering Limited
Dr. Robert Gaspar - (519) 972-0677
gasparr@kelcom.igs.net - Windsor, ON

State of the Art Acoustik Inc.
Dr. C. Fortier - 613-745-2003
sota@sota.ca - Ottawa, ON

Tacet Engineering Ltd.
Dr. M.P. Sacks - (416) 782-0298
mal.sacks@tacet.ca - Toronto, ON

Valcoustics Canada Ltd.
Dr. Al Lightstone - (905) 764-5223
solutions@valcoustics.com - Richmond Hil

Vibro-Acoustics Ltd.
Mr. Tim Charlton - (800) 565-8401
tcharlton@vibro-acoustics.com - Scarborough, ON

West Caldwell Calibration Labs
Mr. Stanley Christopher - (905) 624-3919
info@wccl.com - Mississauga, On

Wilrep Ltd.
Mr. Don Wilkinson - (905) 625-8944
www.wilrep.com - Mississauga, ON