

canadian acoustics

acoustique canadienne

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

MARCH 2008
Volume 36 -- Number 1

MARS 2008
Volume 36 -- Numéro 1

GUEST EDITORIAL / ÉDITORIAL INVITÉ

2

PROCEEDINGS OF THE 3RD INTERNATIONAL WORKSHOP ON DETECTION AND CLASSIFICATION OF MARINE MAMMALS USING PASSIVE ACOUSTICS

Table of Contents / Table des matières

4

Overview of the 3rd International workshop on the detection and classification of marine mammals using passive acoustics

David Moretti, Nancy DiMarzio, Ronald Morrissey, David K. Mellinger, Sara Heimlich and Heather Pettis

7

Other Features / Autres Rubriques

Obituary / Nécrologie - Prof. John E. K. Foreman

184

CAA Prizes Announcement / Annonce de Prix

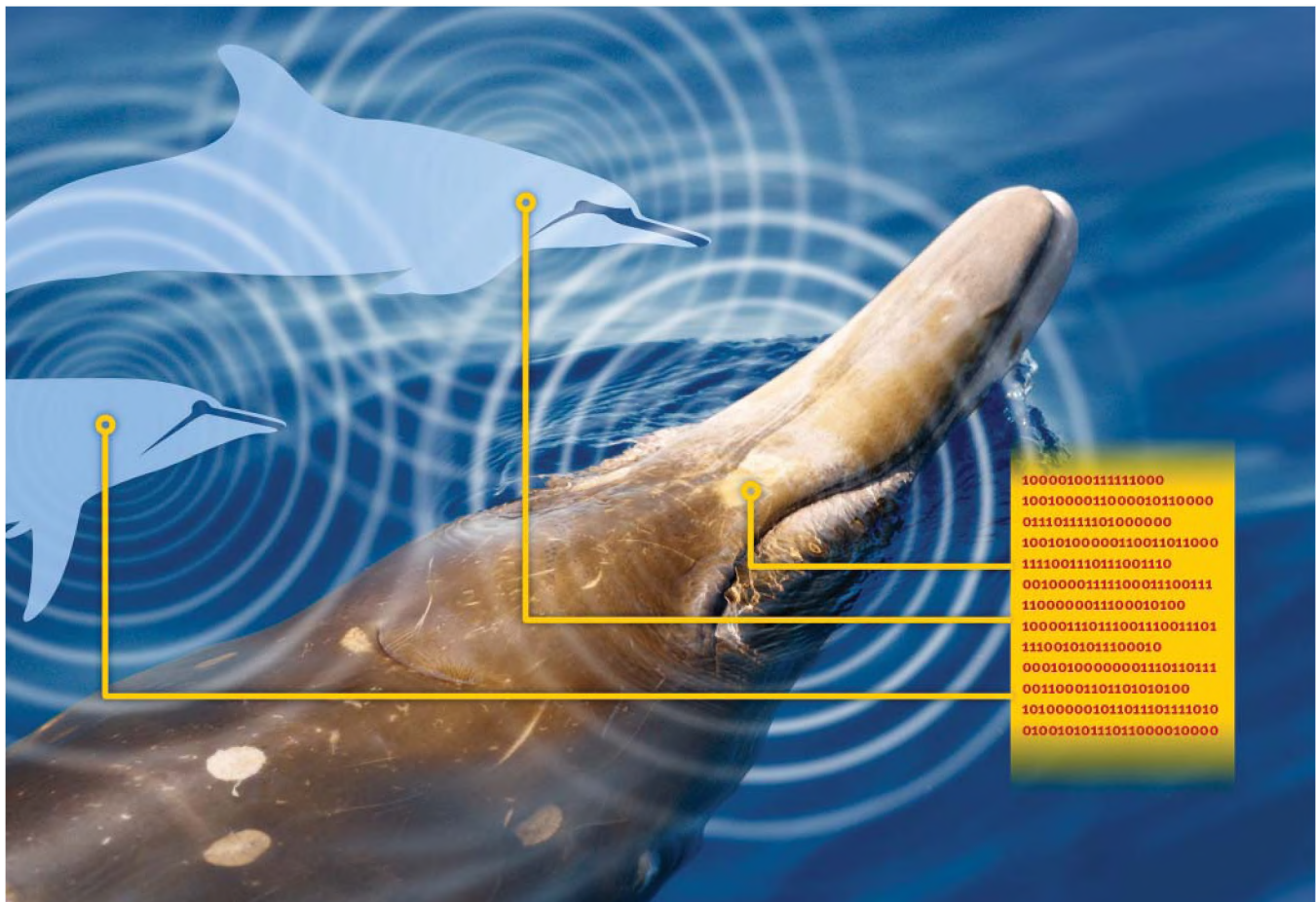
185

News / Informations

186

Canadian News - Acoustics Week in Canada 2008 / Semaine Canadienne d'acoustique 2008

190



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); \$300 (sustaining - see back cover). Back issues (when available) may be obtained from the CAA Secretary - price \$10 including postage. Advertisement prices: \$600 (centre spread); \$300 (full page); \$175 (half page); \$125 (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 (individu, société); \$300 (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: \$600 (page double); \$300 (page pleine); \$175 (demi page); \$125 (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

Jason Tsang
7 Parkwood Crescent
Ottawa, ONTARIO
K1B3J5
E-mail: jtsangeng@yahoo.ca

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 3rd International Workshop on the
Detection and Clasification of Marine Mammals using
Passive Acoustics**

Boston, MASS, USA

July 24-26, 2007

**Les actes de le 3^{ème} atelier international sur la
detection et la clasification des mammifères marins à
l'aide du repérage acoustique passif**

Boston, MASS, USA

July 24-26, 2007

Note: The Image for the cover was provided by Bahamas Marine Mammal Research Organisation and the graphic design for the Cover Page was by Simon Tuckett of GraphiCom Design.

Mitigating the effect of anthropogenic noise on marine mammals is rapidly becoming the focus of management agencies, world navies, and industry. Traditionally, visual methods have been used for monitoring, mitigation, and density estimation. Passive acoustic methods are now also commonly applied to detect and locate animal vocalizations. It has been widely postulated that acoustic methods can also classify vocalizations to species, and that this process can be substantially or entirely automated. The potential advantages of passive acoustic detection, classification, and localization (DCL) of animal vocalizations are clear: These techniques can be used day and night and are less weather-dependent than visual methods. They can be automated, thus ensuring reproducibility and reducing or eliminating observer bias. Sensors can be left on-site and used to monitor for vocalizations over extended periods of time.

In principle, passive acoustic techniques are straightforward. However, marine mammal vocalizations vary widely in frequency and structure over species, groups, individuals, and even for vocalizations from a single individual. For example, blue whales produce calls under 50 Hz with a variety of call types worldwide, while small odontocete clicks can exceed 100 kHz. Marine species may produce songs, clicks, whistles, moans, creaks, and other types of sounds. Thus designing an “optimal” detector is extremely challenging. Classifying vocalizations to the species level is even more so, especially for small odontocetes which may produce very similar vocalizations. Localization is highly dependent not only on the characteristics of the vocalizations, but also on the system used to collect and process data and the local environment in which it is used. Warm to temperate oceans with downward-refracting sound velocity profiles may present a very different problem than an upward-refracting Arctic environment.

To help develop and test new DCL and density estimation algorithms, common, verified data sets are critical. These data sets can be used by researchers to test the efficacy of algorithms on real-world signals and compare the results of such tests. With time, verified methods can be developed, refined, and fielded in real-time systems.

Passive acoustic methods continue to improve, but significant hurdles remain. Two previous workshops were held – the first in Halifax, Nova Scotia in 2003, and the second in Monaco in 2005 – to highlight research in detection and localization and to bring together researchers in the field. Another opportunity for this was provided by the 3rd International Workshop on the Detection and Classification of Marine Mammals Using Passive Acoustics, which is the focus of this Special Issue. For this workshop, held in Boston in July 2007, an associated data set was provided containing clicks from several visually-identified species of odontocetes. The set focused on Blainville’s (*Mesoplodon densirostris*) beaked whales, a species that have stranded during events associated with naval sonars. Researchers were encouraged to test their respective DCL algorithms and compare the results in a blind

Atténuer l’effet des bruits anthropiques sur les mammifères marins est rapidement devenu le focus des organismes d’aménagement de l’environnement, des marines de plusieurs pays, et de l’industrie. Traditionnellement, les méthodes visuelles ont été utilisées pour la surveillance, l’atténuation et l’estimation de la densité. Les méthodes acoustiques passives sont aussi utilisées couramment maintenant pour détecter et localiser les vocalisations d’animaux. Il a été largement postulé que les méthodes acoustiques peuvent également classifier les d’espèces selon leurs vocalisations, et que ce processus peut être partiellement ou entièrement automatisé. Les avantages potentiels de la détection, classification, et localisation (DCL) acoustiques passives de vocalisations d’animaux sont clairs: ces techniques peuvent être utilisées jour et nuit, et elles dépendent moins des conditions météorologiques que les méthodes visuelles. Elles peuvent aussi être automatisées, ce qui garantit la reproductibilité, et réduit ou élimine les biais d’observation. Les capteurs peuvent être laissés sur place et utilisés pour contrôler les vocalisations sur de longues périodes.

En principe, les méthodes acoustiques passives sont simples. Toutefois, les vocalisations des mammifères marins sont très variables en fréquence et en structure selon l’espèce, le groupe, l’individu, et même pour les vocalisations d’un seul individu. Par exemple, les baleines bleues produisent des appels en bas de 50 Hz avec une diversité à travers le monde des types d’appels, tandis que les clics de petits odontocètes peuvent aller jusqu’à plus de 100 kHz. Les espèces marines peuvent aussi produire des chansons, clics, sifflets, gémissements, grincements, et autres types de sons. Ainsi, la conception d’un détecteur “optimum” est extrêmement difficile. Classifier l’espèce selon ses vocalisations l’est encore plus, surtout pour les petits odontocètes qui peuvent produire des vocalisations très similaires. La localisation est très dépendante non seulement des caractéristiques des vocalisations, mais aussi du système utilisé pour recueillir et traiter les données, et de l’environnement local dans lequel il est déployé. Les océans tempérés ou chauds, avec des profils de la vitesse du son qui réfractent le son vers le fond marin, peuvent présenter un problème très différent des profils qui réfractent le son vers la surface, tel que dans l’environnement arctique.

Pour aider à développer et tester de nouveaux algorithmes de DCL et d’estimation de densité, des ensembles de données communs et vérifiés sont essentiels. Ces données peuvent être utilisées par les chercheurs pour tester l’efficacité de leurs algorithmes sur des signaux réels, et comparer leurs résultats. Avec le temps, des méthodes vérifiées peuvent être développées, affinées, et implantées dans des systèmes qui fonctionnent en temps réel.

Les méthodes acoustiques passives continuent de s’améliorer, mais des obstacles importants demeurent. Deux ateliers précédents ont eu lieu - la première à Halifax, en Nouvelle-Écosse en 2003, et la seconde à Monaco en 2005 - pour mettre en évidence la recherche sur la détection et la lo-

test. Workshop papers documenting both the algorithms used to analyze the data set, as well as a broad spectrum of current research topics in the field, are contained in this special edition of the journal. A paper which explains the data set and presents the results of the blind test is also provided. The data set remains available as a resource for current and future marine mammal passive acoustic research.

This third workshop was a success, with 120 registered participants, 54 oral presentations, 8 poster presentations, and a great deal of cross-fertilization of ideas among attendees. The next workshop will be held in Pavia, Italy in 2009. Having an annotated dataset used in common by many participants worked well and is recommended for future workshops. Organizers of future workshops should note that this data set required a surprising amount of advance preparation, and to be most effective, should be released many months before the workshop -- so start early!

David Moretti, David K. Mellinger,
and Francine Desharnais

calisation, et rassembler des chercheurs de ce domaine. Une autre opportunité en a été fournie avec le 3ème atelier international sur la détection et la classification des mammifères marins utilisant l'acoustique passive, qui est au centre de ce numéro spécial. Pour cet atelier, qui s'est tenu à Boston en juillet 2007, un ensemble de données a été fourni contenant des clics de plusieurs espèces (identifiées visuellement) d'odontocètes. Les données sont centrées sur les baleines à bec Blainville (*Mesoplodon densirostris*), des espèces qui ont échouées lors d'événements associés aux sonars navals. Les chercheurs ont été invités à tester leurs algorithmes DCL et comparer leurs résultats dans un test aveugle. Cette édition spéciale du journal contient les articles de cet atelier documentant les algorithmes utilisés pour analyser l'ensemble des données, ainsi qu'un large éventail de thèmes de recherche actuelle dans ce domaine. Un article qui explique l'ensemble de données, et présente les résultats du test aveugle, est également inclus. L'ensemble de données reste accessible en tant que ressource pour les recherches actuelles et futures sur l'acoustique passive des mammifères marins.

Ce troisième atelier a été un succès, avec 120 participants inscrits, 54 présentations orales, 8 posters et beaucoup d'échanges d'idées entre les participants. Le prochain atelier aura lieu à Pavie, en Italie en 2009. L'utilisation d'un ensemble données annotées commun par de nombreux participants est une approche qui a bien fonctionnée et qui est recommandée pour les ateliers futurs. Les organisateurs d'ateliers futurs devraient par contre noter que l'ensemble de données doit être préparé étonnamment à l'avance, et doit être mis à la disposition des participants plusieurs mois avant l'atelier pour être effectif - alors commencez tôt!

David Moretti, David K. Mellinger, et Francine Desharnais

WHAT'S NEW ??

Promotions
Deaths
New jobs
Moves

Retirements
Degrees awarded
Distinctions
Other news

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
Steven Bilawchuk, aci Acoustical Consultants Inc., Edmonton, Alberta, Email: stevenb@aciacoustical.com

QUOI DE NEUF ?

Promotions
Décès
Offre d'emploi
Déménagements

Retraites
Obtention de diplômes
Distinctions
Autres nouvelles

avec les lecteurs de l'Acoustique Canadienne? Si oui, écrivez-les et envoyer à:

THE 3RD INTERNATIONAL WORKSHOP ON DETECTION AND CLASSIFICATION OF MARINE MAMMALS USING
PASSIVE ACOUSTICS

LES ACTES DE LE 3ÈME ATELIER INTERNATIONAL SUR LA DETECTION ET LA CLASSIFICATION DES MAMMIFÈRES
MARINS À L'AIDE DU REPÉRAGE ACOUSTIQUE PASSIF

TABLE OF CONTENTS / TABLE DES MATIÈRES

Editorial	2
Table of Contents	4
Overview of the 3rd International workshop on the detection and classification of marine mammals using passive acoustics David Moretti, Nancy DiMarzio, Ronald Morrissey, David K. Mellinger, Sara Heimlich and Heather Pettis	7
Feature-Aided Tracking for Marine Mammal Detection and Classification Odile Gerard, Craig Carthel, Stefano Coraluppi, and Peter Willett	13
Statistical Classification of Odontocete Clicks Douglas Gillespie and Marjolaine Caillat	20
Processing the Workshop Datasets Using the trud Algorithm Edward Harland	27
A Novel Multi-class Support Vector Machine Classifier for Automated Classification of Beaked Whales and Other Small Odontocetes Susan Jarvis, Nancy DiMarzio, Ronald Morrissey and David Moretti	34
Comparison of machine learning techniques for the classification of echolocation clicks from three species of odontocetes Marie A. Roch, Melissa S. Soldevilla, Rhonda Hoenigman, Sean M. Wiggins, and John A. Hildebrand	41
Detection of Clicks based on Group Delay Varvara Kandia and Yannis Stylianou	48
A Neural Network for Classifying Clicks of Blainville's Beaked Whales (<i>Mesoplodon densirostris</i>) David K. Mellinger	55
Passive acoustic detection and localization of <i>Mesoplodon densirostris</i> (Blainville's beaked whale) vocalizations using distributed bottom-mounted hydrophones in conjunction with a Digital Tag (DTag) recording Jessica Ward, Ronald Morrissey, David Moretti, Nancy DiMarzio, Susan Jarvis, Mark Johnson, Peter Tyack, and Charles White	60
Three-Dimensional Single-Hydrophone Tracking Of A Sperm Whale Demonstrated Using Workshop Data From The Bahamas Christopher O. Tiemann	67
A Comparison Of Pitch Extraction Methodologies For Dolphin Vocalization Xanadu C. Halkias and Daniel P. W. Ellis	74

Attractive Time-Variant Orthogonal Schur-Like Representation For Click-Type Signal Recognition Maciej Lopatka, Olivier Adam, Jean-François Motsch, and Jan Zarzycki	81
Acoustic behavior of dolphins in the pacific ocean: implications for using passive acoustic methods for population studies Shannon Rankin, Julie N. Oswald, and Jay Barlow	88
Definition of the antarctic and pygmy blue whale call templates. Application to fast automatic detection Samaran Flore, Olivier Adam, Jean-François Motsch, Christophe Guinet	93
Detection And Localization Of Blue And Fin Whales From Large-Aperture Autonomous Hydrophone Arrays: A Case Study From The St. Lawrence Estuary Yvan Simard and Nathalie Roy	104
Detection And Recognition Of North Atlantic Right Whale Contact Calls In The Presence Of Ambient Noise Ildar R. Urazghildiiev, Christopher W. Clark and Timothy P. Krein	111
Detection of leopard seal (<i>Hydrurga leptonyx</i>) vocalizations using the envelope-Spectrogram technique (tEST) in combination with a Hidden Markov Model Holger Klinck, Lars Kindermann and Olaf Boebel	118
Joint Localization and Separation of Sperm Whale Clicks Paul M. Baggenstoss	125
Pair-Wise Spectrogram Processing Used To Track A Sperm Whale Eva-Marie Nosal and L. Neil Frazer	132
Whale Cocktail Party: Real-Time Multiple Tracking And Signal Analyses Hervé Glotin, Frédéric Caudal, and Pascale Giraudet	139
Introduction To Particle Filters For Tracking Applications In The Passive Acoustic Monitoring Of Cetaceans P. R. White and M. L. Hadley	146
Robust 2D Localization Of Low-Frequency Calls In Shallow Waters Using Modal Propagation Modelling C. Gervaise, S. Vallez, Y. Stephan, Y. Simard	153
Performance Of Three Acoustical Methods For Localizing Whales In The Saguenay—St. Lawrence Marine Park Nathalie Roy, Yvan Simard ¹ , and Jean Rouat	160
Passive Acoustic Measurement of Dive Vocal Behavior and Group Size of Blainville's Beaked Whale (<i>Mesoplodon densirostris</i>) in the Tongue Of The Ocean (TOTO) Nancy DiMarzio, David Moretti, Jessica Ward, Ronald Morrissey, Susan Jarvis, Anna Maria Izzi Mark Johnson, Peter Tyack, and Amanda Hansen	166
Modeling the Effect of Boat Traffic on the Fluctuation of Humpback Whale Singing Activity in The Abrolhos National Marine Park, Brazil Renata S. Sousa-Lima and Christopher W. Clark	174

831 sound level meter/real time analyzer

- Consulting engineers
- Environmental noise monitoring
- Highway & plant perimeter noise
- Aircraft noise
- General Surveys
- Community noise



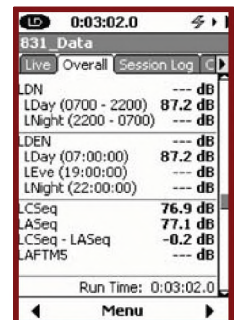
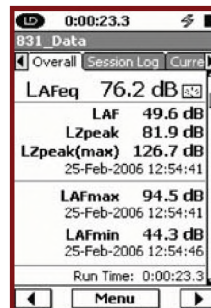
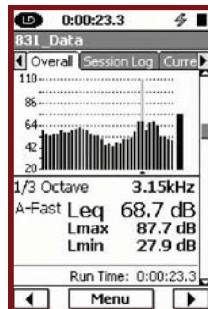
FEATURES

- Class 1/Type 1 sound level meter
- Small size with large display. Ergonomic
- User friendly operator interface
- 120MB standard memory expandable up to 2GB
- Single measurement range from 20 to 140 dB SPL
- Up to 16 hours of battery life
- Provided with utility software for instrument set-up and data download
- Field upgradeable
- AUX port for connection to USB mass storage & cellular modems



MEASUREMENT CAPABILITIES

- Real time 1/1 & 1/3 octave frequency analysis
- Simultaneous display of several noise measurements—ANY DATA (Leq, Lmax, Spectra, etc)
- Automatic logging of user selectable noise measurements (Leq, Lmax, Spectra, etc...)
- Exceedance logging with user selectable trigger levels
- Audio and voice recording with replay



OVERVIEW OF THE 3RD INTERNATIONAL WORKSHOP ON THE DETECTION AND CLASSIFICATION OF MARINE MAMMALS USING PASSIVE ACOUSTICS

David Moretti¹, Nancy DiMarzio¹, Ronald Morrissey¹, David K. Mellinger²
Sara Heimlich² and Heather Pettis³

1 – Naval Undersea Warfare Center Division, 1176 Howell St., Newport, RI 02841 USA

2 – Oregon State University, 2030 SE Marine Science Dr., Newport, OR 97365 USA

3 – New England Aquarium, Central Wharf, Boston, MA 02110 USA

ABSTRACT

The 3rd International Workshop on the Detection and Classification of Marine Mammals Using Passive Acoustics was held 24-26 July 2007 in Boston, MA. A dataset containing verified odontocete vocalizations from five different species, including Blainville's beaked whale (*Mesoplodon densirostris*), was provided for the testing and development of detection and classification algorithms. Data collected under different acoustic conditions were included along with a blind test dataset. Six research groups tested their respective algorithms against the unknown data and presented their results. Both the data set and the test results are presented.

SOMMAIRE

Le 3^{ème} atelier international sur la détection et la classification des mammifères marins employant l'acoustique passive a été tenu le 24-26 juillet 2007 à Boston, MA. Un ensemble de données vérifiées de vocalisations d'odontocètes de cinq espèces différentes, y compris la baleine à bec de Blainville (*Mesoplodon densirostris*), était disponible pour l'essai et le développement des algorithmes de détection et classification. Des données enregistrées dans des conditions acoustiques différentes étaient incluses, ainsi qu'un ensemble de données pour test aveugle. Six groupes de recherche ont testé leurs algorithmes respectifs avec les données inconnues, et ont présenté leurs résultats. L'ensemble de données et les résultats du test sont présentés ici.

1. INTRODUCTION

Marine mammal passive acoustic methods to monitor individual animals and populations are undergoing rapid development. Algorithms for the detection, classification, and localization (DCL) of marine mammal vocalizations are critical to these methods. To foster this development, the 3rd International Workshop on the Detection and Classification of Marine Mammals Using Passive Acoustics was held 24-26 July 2007 in Boston, MA, USA. Two previous workshops concentrated on detection and localization without emphasis on the problem of classification. This workshop emphasized classification as a core topic area. The workshop provided researchers from around the world the opportunity to present their work and to test the efficacy of their DCL algorithms on a common data set consisting of sounds recorded from odontocete species identified by experienced visual observers. The workshop drew participants from different fields and included specialists in biology, acoustics, signal processing, mathematics, electronics, and computer science. Topics for presentation were extended beyond passive acoustics to areas of research related to the effects of anthropogenic sound on marine

mammals. The scientific topics encompassed by the workshop were as follows:

1. Underwater acoustics
2. Detection and Classification
3. Localization
4. Biology of Marine Mammals
5. Density Estimation
6. Applications

A half-day was reserved for the comparison of scientific methods applied to a common workshop data set of recorded odontocete clicks. The dataset allowed researchers to develop and compare DCL algorithms. It was provided by the U.S. Navy's Naval Undersea Warfare Center (NUWC), Division Newport and hosted by Oregon State University on the MobySound website. It can be accessed at:

<http://hmsc.oregonstate.edu/projects/MobySound/MsSoundSets.html>

The first five journal papers are from participants who developed detection and classification algorithms and attempted to identify the species in the workshop dataset test files. A summary of their results is included below. The next four papers are from participants who used the workshop dataset in other ways. The remaining papers are organized according to the scientific topics listed above.

2. WORKSHOP DATA

Data for the workshop dataset were collected by the NUWC Marine Mammal Monitoring on Navy Ranges (M3R) program during species verification tests at two U.S. Navy ranges: the Atlantic Undersea Test and Evaluation Center (AUTEc) located off Andros Island, Bahamas, and the Southern California Offshore Range (SCORE) off southern California (Figure 1). Animal vocalizations were recorded on wide-band bottom-mounted hydrophones located at these ranges. Data were analyzed and prepared for the workshop at NUWC and Oregon State University.

2.1 Navy Ranges

The AUTEc range consists of 82 operational hydrophones covering a total area of over 1500 km². The hydrophones are at varying depths of approximately 1300-1900 meters. Sixty-two of the hydrophones are arranged in offset rows on approximately 3.8 km (2 nm) baselines, with a bandwidth of about 50 Hz to 45 kHz. Fourteen of the hydrophones are arranged into two 7-hydrophone hexagonal arrays with a center hydrophone. These hydrophones are on a baseline of about 1.8 km (1 nm), and have a bandwidth of 8 to 50 kHz. The AUTEc hydrophones are located east of Andros Island in the Bahamas in a deep ocean canyon known as the Tongue of the Ocean (TOTO).

The SCORE range contains 88 hydrophones covering an area of over 1800 km². The SCORE hydrophone baselines range between about 2.5 km (1.3 nm) close to shore to 6.5 km (3.5 nm) farther west, although a few hydrophone pairs have shorter baselines, with one as small as 1.65 km (0.89 nm). The hydrophones vary in depth from about 800 to 1760 meters. Twenty of the hydrophones are individually cabled with a nominal bandwidth of 8 to 50 kHz, and the other 68 are on multiplexed arrays, with a nominal bandwidth of 8 to 39 kHz. The range is located to the west of San Clemente Island, California.

2.2 Alesis recorders

The raw acoustic data are cabled to shore, where they are recorded using a bank of eight Alesis HD24 hard disk recorders. Each Alesis unit records up to 12 channels, with an IRIG time code on the 12th channel. The data are recorded as 24-bit samples at a 96 kHz sample rate using standard IDE hard drives.

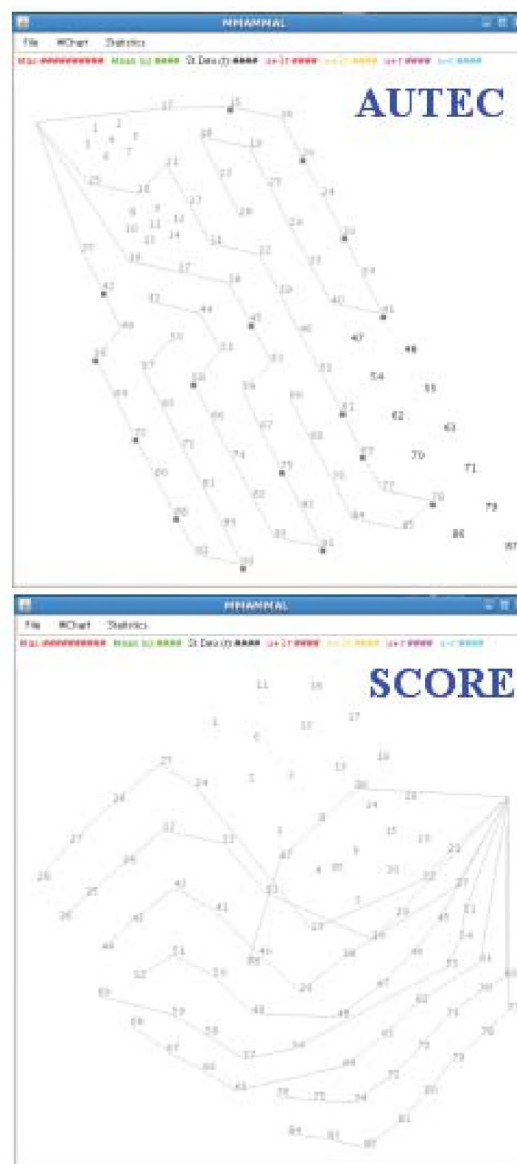


Figure 1. The AUTEc (top) and SCORE (bottom) ranges.

2.3 Species Verification Tests

M3R has conducted species verification tests at each of these ranges with highly trained surface observers. At the AUTEc range, M3R collaborated with the Bahamas Marine Mammal Research Organization and the Woods Hole Oceanographic Institution, and at the SCORE range with the Cascadia Research Collective and Scripps Institution of Oceanography. Animal vocalizations on the range were acoustically monitored in real-time using M3R software. The goal of the species verification tests was to use experienced surface observers to visually verify species that were acoustically detected and monitored on-shore. The shore team monitored the range for animals and directed the

observers to the location of vocalizations of interest. The surface observers found the animals, verified the species present, took photo IDs and recorded behavioral data. During the course of these tests all hydrophones were recorded. The data provided in the workshop were drawn from the closest recordings both spatially and temporally to verified sightings of the species of interest. Though care was taken to provide sound cuts matched to visually verified species, there are cases in which the cuts may be contaminated with vocalizations from other unsighted species in the vicinity of the recorded hydrophones. This is particularly a problem for Risso's dolphin sound cuts from the SCORE range, where the animal density is much higher than at AUTECH.

2.4 Workshop Dataset

Cuvier's and Blainville's beaked whale species are among those most often involved in mass strandings linked to naval mid-frequency sonars. The workshop dataset consists of both training data and test data for a variety of odontocete species, with an emphasis on Blainville's beaked whales (*Mesoplodon densirostris*), as data from multiple verified sightings were available for this species. Sounds from each species comprised a number of "cuts", short segments of continuously-recorded sound.

Training data

Training data sound cuts were provided for the following three species:

1. Blainville's beaked whale (*Mesoplodon densirostris*)
2. Short-finned pilot whale (*Globicephala macrorhynchus*)
3. Risso's dolphin (*Grampus griseus*)

The Blainville's beaked whale and short-finned pilot whale sound cuts were recorded at the AUTECH range, and the Risso's dolphin sound cuts were recorded at SCORE. Most training sound cuts were 2 minutes long, though they varied from 0.5 to 3 minutes in length.

Sixteen sound cuts of Blainville's beaked whale were provided. These data were collected at AUTECH on April 27, 2005 and September 24 and 27, 2005. Nine short-finned pilot whale sound cuts were available for training. These data were recorded at AUTECH on September 24, 26, and 30, 2005. The Risso's dolphin sound cuts were collected at SCORE on August 14, 16, and 19, 2006. Eleven of these sound cuts were provided for training.

Test Data

Nine longer, unidentified sound cuts, approximately ten minutes each, were provided as test data. Six of these files were from verified sightings. Three were unverified, but

were from easily recognizable species (beaked whale and sperm whale). The test files were numbered one through nine and the species identity was withheld. The correct species identifications are shown in Table 1.

Test File	Key
1	Unverified Blainville's beaked whale (<i>Mesoplodon densirostris</i>) & verified short-finned pilot whale (<i>Globicephala macrorhynchus</i>)
2	Blainville's beaked whale (<i>Mesoplodon densirostris</i>)
3	Risso's dolphin (<i>Grampus griseus</i>)
4	Pantropical spotted dolphin (<i>Stenella attenuata</i>)
5	Risso's dolphin (<i>Grampus griseus</i>)
6	Unverified Blainville's beaked whale (<i>Mesoplodon densirostris</i>) & unverified sperm whale (<i>Physeter macrocephalus</i>)
7	Pantropical spotted dolphin (<i>Stenella attenuata</i>)
8	Short-finned pilot whale (<i>Globicephala macrorhynchus</i>)
9	Unverified sperm whale (<i>Physeter macrocephalus</i>)

Table 1: Species identifications for workshop dataset test files.

The test cases were organized into three categories, with the following species in each category:

1. High signal-to-noise ratio (SNR) sound cuts
 - a. Blainville's beaked whale
 - b. Short-finned pilot whale
 - c. Risso's dolphin
2. Sound cuts with multiple species
 - a. Unverified Blainville's beaked whale & verified short-finned pilot whale
 - b. Unverified Blainville's beaked whale & unverified sperm whale
3. Sound cuts with no signal of interest
 - a. Pantropical spotted dolphin
 - b. Unverified sperm whale

High-SNR test cases were presented for the three species for which training data were provided. Alternate species and sound cuts with multiple species were also provided to test and compare the different methods for detection and classification. All the test cases were recorded at AUTECH except for the Risso's dolphin cuts, which were recorded at SCORE.

Dataset Annotations

A spreadsheet was provided listing the sound cut filenames and source files for the sound cuts; the date and location of the recordings; the type of recorder, sample rate and

bandwidth; start and stop times referenced to the local time, Greenwich mean time (GMT), and the Alesis recorder; length (in minutes) and size (in megabytes) of the source files; whether the source files were time-aligned and visually verified; the local sighting time, nearest hydrophone and common and scientific names of the species sighted; notations regarding the presence of clicks, whistles/moans, creaks/buzzes, man-made noise, boat/engine noise, unknown sounds; and more detailed comments describing the contents of the cuts.

Information was also provided describing the naming conventions for the files, the hydrophone locations and depths in meters, and additional sighting data.

Additional Data

Upon request, two additional datasets were provided. One was a Blainville's beaked whale dataset for localization, consisting of 18-minute long sound cuts from five neighboring hydrophones on the AUTEK range. These sound cuts were from a verified sighting at AUTEK on September 27, 2005. The second was an AUTEK noise dataset, consisting mostly of background noise. These were three files, each 30 minutes long, recorded from different parts of the range on April 26, 2005.

3. WORKSHOP DATASET ANALYSIS AND RESULTS

Six participants presented results from the application of their detection and classification algorithms to the test data set. The participants attempted to identify the species present in each of the dataset test files. Following the authors' presentations, the correct species identification for each vocalization test file was released.

The results of the analysis, based on each author's presentation, are provided in Table 2, which summarizes the number of species correctly identified, as well as those omitted, added, or misidentified. The table presents raw classifier outputs. To compile the results, the number of correctly identified species was first tallied. The number of species misidentified was then counted from those remaining, followed by the number omitted, and finally the number added. A direct comparison of the authors' results was difficult. In some cases, the results were open to interpretation, as they were simply a tabulation of the number of times a particular species was correctly identified. Future workshop organizers should provide participants with a clearly defined format for the tabulation of their results. A number of authors recognized that species were misidentified by their classifiers and suggested minor adjustments would improve performance.

Overall, Roch's classifier performed best, with the most correct species and the fewest misidentified ones. Van

IJsselmuide's detector did nearly as well, and the detectors of Gerard and Jarvis had only slightly more errors.

4. CONCLUSION

This third workshop was a success, with a great deal of cross-fertilization of ideas among attendees. The direct comparison of detectors on a common dataset was successful, in part because the workshop dataset was prepared months in advance of the workshop. It is hoped that this workshop series will continue into the future.

5. ACKNOWLEDGEMENTS

The organizers would like to thank the Office of Naval Research and the Chief of Naval Operations for support of this workshop. We would also like to thank the scientific committee and the reviewers for their tireless effort. Finally, we would like to thank all of the workshop participants, particularly those who analyzed the workshop dataset.

	Key	Gerard	Gillespie	Harland	Van IJsselmuide	Jarvis	Roch
Test File 1	Md, Gm	Md, Gg?	Md, Gm	Md	Md, Gm	Md, Gm	Md
Test File 2	Md	Md, Gg?	Md	Md	Md	Md	Md
Test File 3	Gg	Md, Gm?	Gm	Gm	Gg?	Gm	Other
Test File 4	Sa	Md, Gm?	Gm	Gm	Gg?	Gm	Other
Test File 5	Gg	Gg	Gg	Gg	Other	Md	Gg
Test File 6	Md, Pm	Md, Other	Md, Gm	Md	Md, Pm	Md, Pm	Md, Pm
Test File 7	Sa	Gg	Gg	Gg	Md?	Md, Gg	Gg
Test File 8	Gm	Gm?	Other	Md	Other	Other	Gm
Test File 9	Pm	Gg?	Gm	Gm	Pm	Pm	Pm
# Species Correct		6	5	4	7	6	8
# Species Misidentified		5	6	5	4	5	2
# Species Omitted		0	0	0	0	0	1
# Species Added		3	0	0	0	1	0

- Md** *Mesoplodon densirostris* (Blainville's beaked whale)
Gm *Globicephala macrorhynchus* (Short-finned pilot whale)
Gg *Grampus griseus* (Risso's dolphin)
Sa *Stenella attenuata* (Pantropical spotted dolphin)
Pm *Physeter macrocephalus* (Sperm whale)

Table 2: Results for the participants who attempted to identify the species present in the dataset test files.

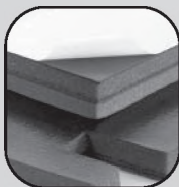


Photo Credit: Bahamas Marine Mammal Research Organisation

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



FEATURE-AIDED TRACKING FOR MARINE MAMMAL DETECTION AND CLASSIFICATION

Odile Gerard^{1*}, Craig Carthel¹, Stefano Coraluppi¹, and Peter Willett²

¹NATO Undersea Research Centre, Viale S. Bartolomeo 400, 19126 La Spezia, Italy

²ECE Department, University of Connecticut, Storrs, CT 06269, USA

* Contact author. E-mail: odileg@free.fr.

ABSTRACT

This paper presents a method to detect and classify odontocete echolocation clicks as well as to estimate the number of animals that are vocalizing. A transient detector using the Page test [1-3] is used to extract the clicks: the click time, the click duration, the click amplitude and the spectral information of the clicks are extracted. A probability distribution over the species is assigned to each click, based on the spectral information of the click. The estimation of the number of animals is done using feature-aided multi-hypothesis tracking (MHT) algorithms. The association is based on the assumptions of slowly-varying click amplitude and intra-click timing [4-5]. This work has been done on the dataset provided by the organizers of the *3rd International Workshop on the Detection and Classification of Marine Mammals using Passive Acoustics*, Boston, July 2007. This dataset consists of training and test data; the training data includes vocalizations of three species: Blainville's beaked whale (*Mesoplodon densirostris*), Risso's dolphin (*Grampus griseus*) and short-finned pilot whale (*Globicephala macrorhynchus*).

SOMMAIRE

Cet article présente une méthode de détection et classification de clics d'écholocation d'odontocètes ainsi que d'estimation du nombre d'animaux vocalisant en même temps. Un détecteur de transitoires utilisant le test de Page [1-3] permet d'extraire les clics : leurs instants, durées et amplitudes ainsi que leurs spectres sont stockés. L'analyse du spectre d'un clic permet de lui affecter une probabilité de distribution parmi les différentes espèces. L'estimation du nombre d'animaux se fait à l'aide d'un algorithme de tracking (multi-hypothesis tracking MHT). L'association des clics est basée sur l'hypothèse que l'amplitude et l'intervalle entre deux clics varient lentement en fonction du temps. Ce travail a été réalisé sur le jeu de données mis à disposition par les organisateurs du *3rd International Workshop on the Detection and Classification of Marine Mammals using Passive Acoustics*, Boston, Juillet 2007. Ce dernier se compose de données d'entraînement sur trois espèces : Mésoplodon de Blainville (*Mesoplodon densirostris*), dauphins de Risso (*Grampus griseus*) et globicéphales (*Globicephala macrorhynchus*) et de fichiers test.

1 INTRODUCTION

The most reliable means to detect echolocating cetaceans is acoustic: one listens for "clicks". It is of interest to detect and classify the clicks automatically, and subsequently to determine how many animals are present.

The process observed from each animal is a sequence of clicks whose inter-event times and whose amplitudes vary slowly. From the observer's point of view there is the superposition of an unknown number of such processes, in addition to spurious measurements, hence both tracking and data association are helpful in determining the number of independent sources.

Numerous approaches exist to the tracking problem. Contact-based approaches are of interest here, since clicks provide contact-level measurement information. These techniques include sequential (scan-based), as well as batch processing techniques. In earlier work we documented our results in the analysis of hydrophone datasets with a variety of approaches; the most effective,

at least at the time being, has proven to be the multi-hypothesis tracking (MHT) based approach. In this work, we further develop this tracker to include click feature information that allows to classify clicks originating from different species of vocalizing mammals.

Section 2 provides a description of the transient detection algorithm. Section 3 describes the assignment of probability over species to each click. Section 4 describes the feature-aided MHT algorithm. Section 5 provides some results. Conclusions are in section 6.

2 TRANSIENT DETECTION ALGORITHM

The transient detection algorithm is a slightly modified version of the algorithm described in [1]. The algorithm is summarized in figure 1.

First, the data is high-pass filtered to remove part of the noise and avoid detection of whistles (Butterworth order 8, cut-off frequency: 15 kHz). The squared time

series of the filtered data is normalized (using an exponential averager) and then submitted to the Page test. The Page test is a sequential detector that provides robustness against unknown signal duration as it detects the start and the end of a signal. At this step, the time, the duration, the amplitude and the spectral information of the click are stored for the next processing steps.

3 PROBABILITY DISTRIBUTION OVER SPECIES

The first step is to have some criteria to distinguish the clicks of the various species.

For the dataset provided for the workshop the maximum sampling frequency is 96 kHz (that means the spectrum is limited to frequencies below 48 kHz). This sampling frequency is not high enough to characterize the entire click spectrum of the different species; because of this limitation the criteria to distinguish the species will be based on the lower portion of the spectrum.

Below we describe some characteristics of the clicks of the three species of interest in the dataset. Based on these characteristics, to each click we identify a normalized likelihood vector that quantifies the goodness of fit of the click spectrum to those of the species of interest. In particular, the four-dimensional likelihood vector includes one element for each species, and one for none of these.

The likelihood vector impacts the track-to-click association scores that are also based on amplitude and Inter-Click Interval (ICI) information, as discussed further in section 4. In particular, the track state includes a probability distribution over the four classes of interest; the inner product between this distribution and the click likelihood vector impacts the track-click association score.

3.1 Blainville's beaked whale

The energy of regular clicks of Blainville's beaked whale is distributed between the -10dB endpoints of about 26 and 51 kHz with a sharp cut-off below 25 kHz and a more gradual cut-off at the high end [6]. The spectra of a few Blainville's beaked whale clicks coming from the training data are illustrated in figure 2; they correspond to the description of [6].

The spectrum of the extracted clicks is not always as nice as the examples of figure 2, as it depends on the quality of the signal, the signal to noise ratio and the quality of the click extraction; what seems important to recognize these clicks is that they have their maximum frequency above 25 kHz and a very sharp cut-off frequency between 20 and 25 kHz. The buzz clicks are different [6] from the regular clicks but no specific criterion to classify them was used.

3.2 Risso's dolphin

For this species the clicks seem to have more spectral diversity; in all the training data files there are some clicks with narrow bursts in their spectrum. These bursts seem to be typical of the Risso's dolphin. They are not always at the same frequency, and there is not always the same number of bursts, but many of these bursts are around 22, 25 and 31 kHz. Some other clicks don't have these bursts at all and are more difficult to characterize. Figure 3 gives an example of some Risso's dolphin clicks spectrum coming from the training data.

3.3 Short-finned pilot whale

The clicks of the pilot whale vary significantly and are not easy to characterize. In the training data, what is often observed is a maximum between 15 and 20 kHz with energy until the end of the band (45 kHz). Some examples in the training data contain clicks with a maximum frequency above 25 kHz. Figure 4 gives an example of some pilot whale clicks spectrum coming from the training data.

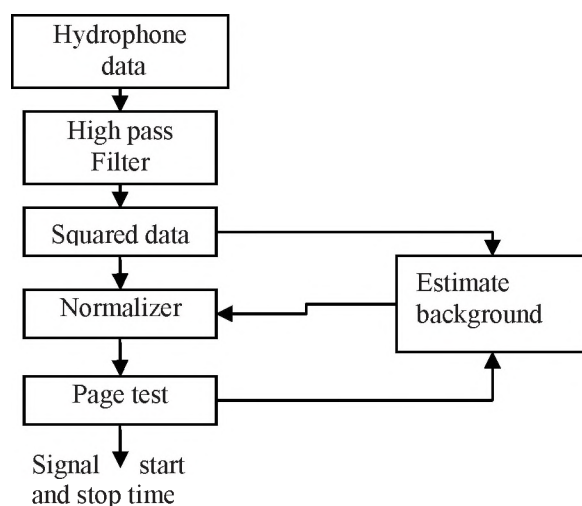


Figure 1: Block diagram of detection scheme.

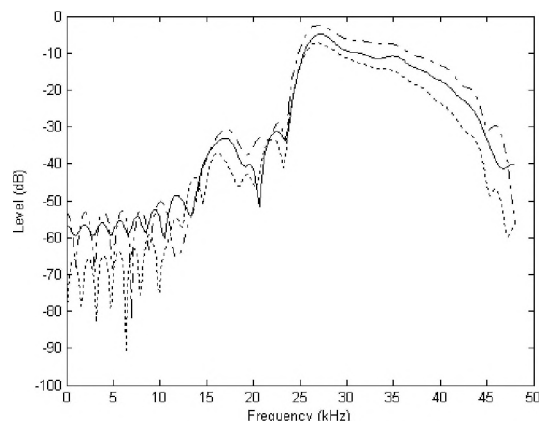


Figure 2: Examples of Blainville's beaked whale regular click spectra (from the training data).

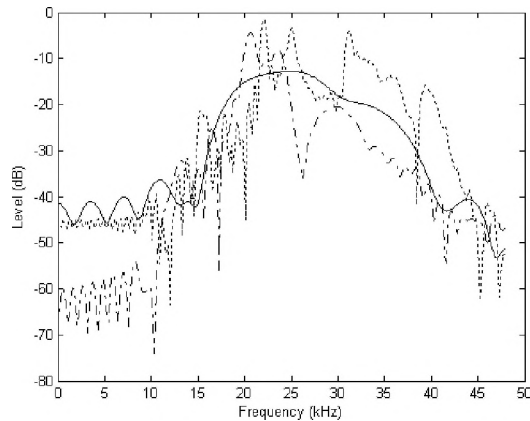


Figure 3: Examples of Risso's dolphin click spectra (from the training data).

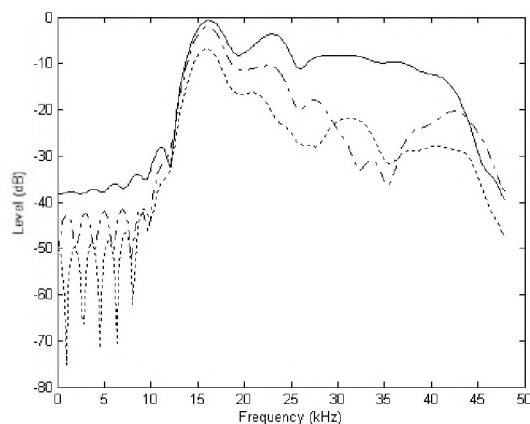


Figure 4: Examples of pilot whale click spectra (from the training data).

4 MHT APPROACH TO CLICK ASSOCIATION

The clicks of interest are echolocation clicks that are emitted by the animals to find prey. The clicks are regular with some pauses; the tracker associates these sequences of clicks. Thus, the track (or associated sequence of clicks) of a single animal will not be contiguous; rather, each animal may generate a number of click sequences separated by lengthy pauses. Our estimate for the number of animals is given by the largest number of tracks that coexist at any time.

Signal processing of hydrophone data results in a single time series of clicks. This time series includes sub-sequences that originate from an unknown number of vocalizing whales, as well as possible spurious clicks. For each marine-mammal originated sub-sequence, we assume the click amplitudes (dB) are slowly varying. Changes in amplitude and intra-click timing are due to animal motion, ambient disturbances, multi-path effects, etc. Animal feeding patterns are another source of change. Each sub-sequence may have missing detections. Our dynamical model for each sub-sequence is the following:

$$20 \log x_{k+1} = 20 \log x_k + w_k, \quad (1)$$

$$(t_{k+1} - t_k) = (t_k - t_{k-1}) + v_k. \quad (2)$$

x_k is the click amplitude of the click at time t_k ; w_k and v_k are noise terms with variance $q_w(t_k - t_{k-1})$ and $q_v(t_k - t_{k-1})$, respectively; the time dependence results from integration of an underlying continuous-time dynamical model.

From equations (1-2), we see that the state of the sub-sequence at time t_k is given by $[x_k \ t_k \ t_{k-1}]^T$. As noted above, the overall observed click sequence is given by the union of the marine-mammal originated sub-sequence, with an additional (unmodelled) spurious false click sequence. In the following we have $X_k = 20 \log x_k$. Equation (1) becomes:

$$X_{k+1} = X_k + w_k, \quad (3)$$

Neglecting transmission loss differences from one click to the next, the model applies to the received signal amplitude.

The identification of the model parameters q_w and q_v requires the use of clean datasets for which each vocalization sequence has few missed clicks and these originate from the same animal. The workshop dataset does not provide the possibility to estimate these parameters because there is not enough data with just one animal vocalizing.

Our past work in MHT tracking has focused on ground and undersea surveillance, the latter based on the use of active sonar; see [7-8] and references therein. Here, we have leveraged the same data association methodology and track management logic, with appropriate modification to kinematic and measurement modeling, recursive filtering, and measurement gating logic. Kinematic modeling is given by (2-3), with parameter settings as noted above. We assume perfect measurements of click times and amplitudes.

The tracker processes the click time series in sequential fashion. At each step, the set of tentative track hypotheses is updated with the current click. With a fixed latency, known as n-scan, track hypotheses are resolved; by this we mean that a single global hypothesis is maintained and all conflicting track hypotheses are pruned. The global hypothesis selection is based on maximization of the sum of track scores, where each track score is a log-likelihood value that includes a track initiation penalty term.

Track hypotheses are generated on the basis of track validation criteria: each click initiates a tentative track. A later click leads to a track update hypothesis if the resulting ICI is low enough, and if the click amplitudes are sufficiently close, based on a chi-squared criterion; a track coast hypothesis is also generated.

Under the hypothesis of two (or more) associated clicks, subsequent track updates require that a chi-squared criterion be met in both amplitude and ICI. Tentative tracks are confirmed with a minimum-click criterion. Tracks that fail to satisfy this criterion are discarded.

In our past work, the tracker did not exploit click feature information beyond click time and amplitude. In the present work, we exploit feature information in the form of the species type probability distribution described in section 3. Thus, the track state includes a species type vector, in addition to the current estimate of ICI and click amplitude.

As the time series click data is processed, the current click is compared against all active tracks, and only those clicks that satisfy the chi-squared criterion previously mentioned (for ICI and amplitude), in addition to a proximity test for species type, are considered as feasible track-click associations. Finally, it should be noted that elements of the species type vector are clipped at each track update: that is to say, each element is bounded away from 1, to avoid insensitivity over time to new data.

5 RESULTS

The results of our automatic-tracking formalism applied on sperm whale clicks can be found in [5-6]. In those efforts, the datasets supported identification of the relevant motion parameter estimates. In the present work, the dataset does not allow the identification of these parameters because it does not provide sequences of clicks long enough from any single animal. Without the estimation of these parameters, the method is challenged when many animals are present. Below we give the result of a very simple case for Risso's dolphin; for Risso's dolphin and pilot whales, we did not estimate the number of animals vocalizing at the same time.

The ICI of Blainville's beaked whales is typically between 0.2 s and 0.4 s [6, 9-10]. With this limitation, even in the presence of many animals, click association is possible even if the right values of model parameters are not known. Some illustrative examples are presented below.

Note that, in the sequence of figures that go with the examples, tracks are plotted alternatively in dashed or dotted line. We do not know definitively whether distinct tracks originate from the same animal. Nonetheless, we estimate the number of vocalizing mammals as the maximum number of co-existing tracks present in a given dataset.

5.1 Training data: Risso's dolphin.

The tracker was applied to a simple case of Risso's dolphin coming from the training data. Figures 5 and 6 give respectively the amplitude of the tracks and the ICI for each track. These results are obtained with the parameters $q_w = 30 s^{-2}$, $q_v = 0.004$. In this case, clicks of one animal are associated leading to an ICI around 0.6 s. Some clicks are not associated into tracks (figure 5): they can come from an animal that is far and whose clicks

are not detected regularly enough to be associated, or possibly they are echoes of the associated clicks.

5.2 Training data: Blainville's beaked whale

For the Blainville's beaked whale data, because of the directivity of the clicks and the fact that these animals have a neck and can move their heads, we have chosen a large value for q_w which allows for large variations in click amplitude. The following parameters are used: $q_w = 50 s^{-2}$, $q_v = 0.01$. Figures 7 and 8 give respectively the amplitude of the tracks and the ICI for each track for one file of the training data. In this case, almost all the clicks are associated. Figure 9 gives the estimated number of whales versus time. It seems that for this file there are often two whales vocalizing simultaneously.

5.3 Test Data

For the test data, only the number of beaked whales is estimated. The tracks are plotted only if the probability for a track to come from a beaked whale is more than 0.5. We will present two examples. In the first one, most of the clicks come from a sperm whale, but there are also some Blainville's beaked whale clicks. Figures 10 and 11 give respectively the amplitude of the tracks and the ICI for each track generated from this data. Many clicks are not associated (those coming from the sperm whales); nevertheless, some clicks are associated at various times (figure 10). Figure 11 is given for a short time window, so as to illustrate how many whales are vocalizing at the same time. In this example, it seems that a maximum of two whales are vocalizing at the same time.

In the second example, all clicks have been identified (by the data provider) as coming from the pantropical spotted dolphin. Nonetheless, at two points in the time series, the tracker associates clicks with tracks having a high probability to be from beaked whales. Figures 12 and 13 give respectively the amplitude of the tracks and the ICI for one of the two tracks generated from this data. The ICI of these clicks matches the expected ICI of the Blainville's beaked whale, and is consistent with the species type probability distribution of the track. Note that the ICI is not used in determining the species type probability distribution, nor is it used in identifying individual animals. Rather, as discussed previously, the species vector is determined through the spectral content information in the clicks.

Figure 14 gives an example of the spectrum of a click coming from a pantropical spotted dolphin (continuous line) and the spectra of two of the associated clicks (dashed and dotted lines). The spectra of the associated clicks really have the typical shape of the Blainville's beaked whale, which is quite different from the other clicks in the time series. Finally, figure 15 gives the temporal signal of these clicks: they too seem to have the typical shape of the Blainville's beaked whale [11]. For

all these reasons, we conclude that these few clicks likely come from a Blainville's beaked whale.

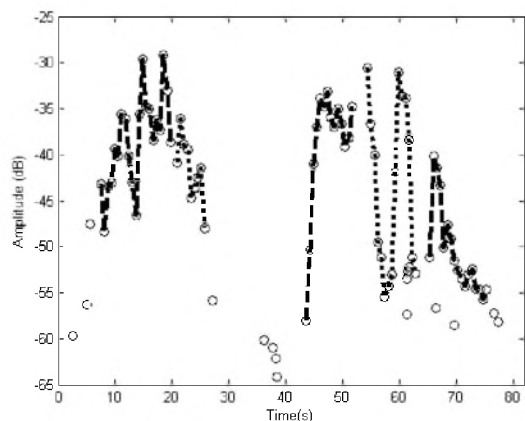


Figure 5: Risso's dolphin click amplitude data (circles) and MHT output (in dashed or dotted lines).

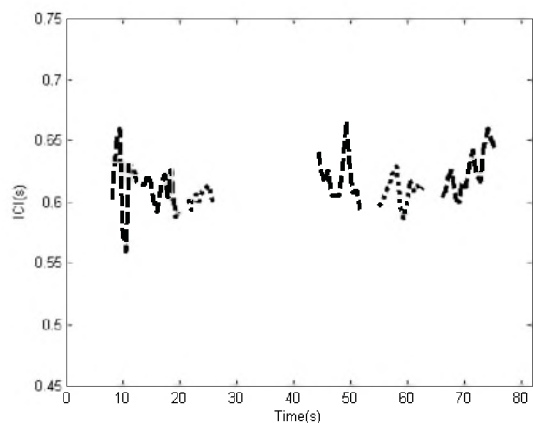


Figure 6: Sequences of Risso's dolphin ICIs for tracks generated by the MHT tracker.

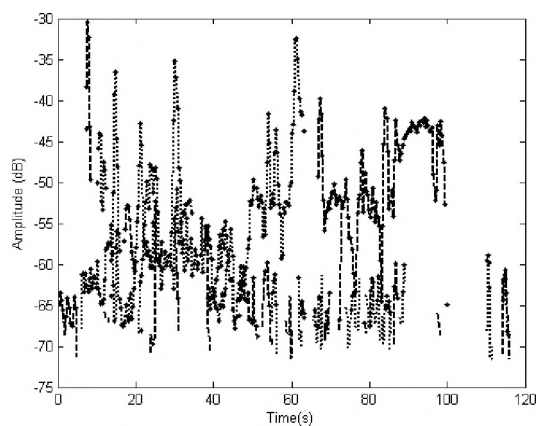


Figure 7: Blainville's beaked whale click amplitude data (dots) and MHT output (dashed or dotted lines).

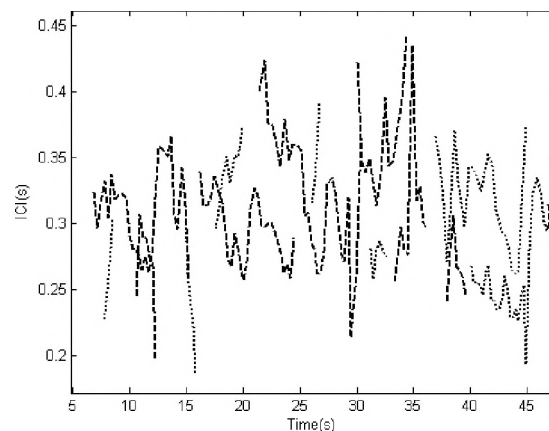


Figure 8: Sequences of Blainville's beaked whale ICIs for tracks generated by the MHT tracker.

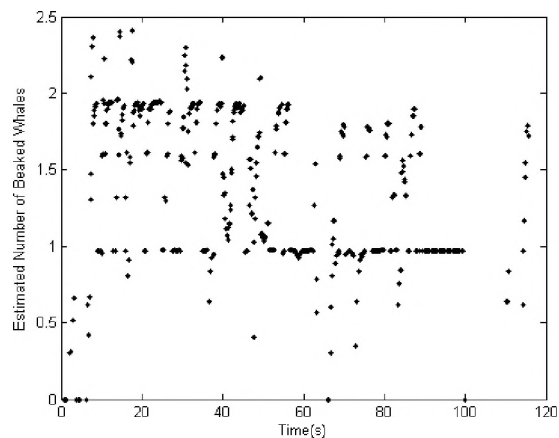


Figure 9: Estimated number of whales vocalizing versus time

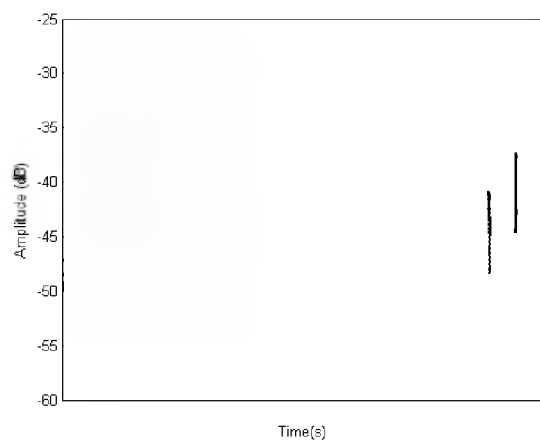


Figure 10: Click amplitude data (dots) and Blainville's beaked whale clicks MHT output (test data, first example).

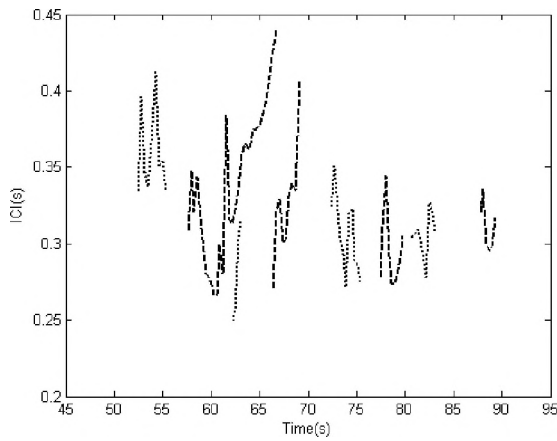


Figure 11: Sequences of ICIs for tracks generated by the MHT tracker.

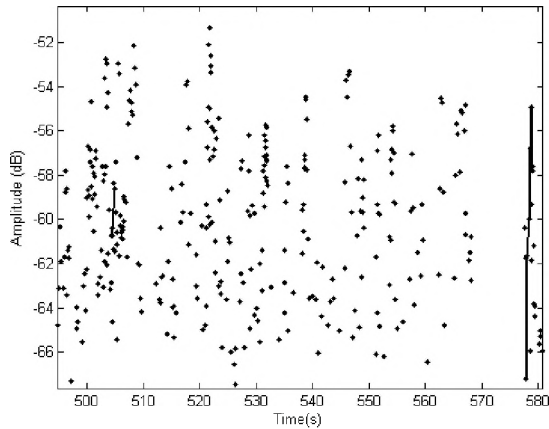


Figure 12: Click amplitude data (dots) and Blainville's beaked whale clicks MHT output (test data, second example).

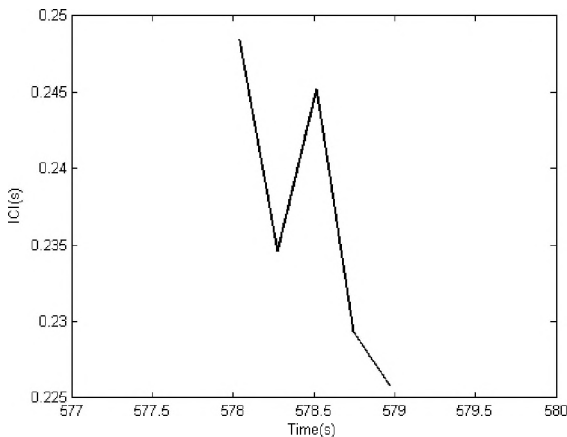


Figure 13: ICIs for a track generated by the MHT tracker.

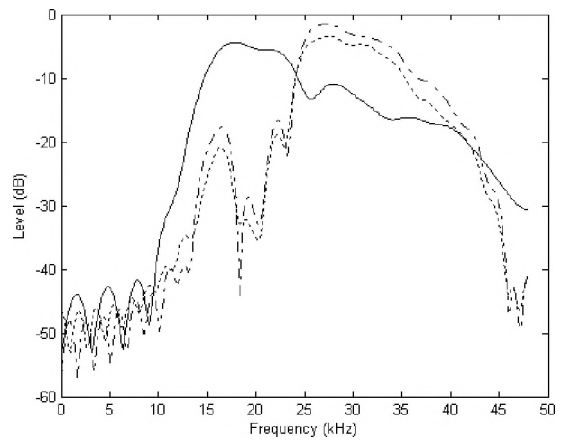


Figure 14: Examples of pantropical spotted dolphin click spectrum (continuous line) and – probably – Blainville's beaked whale click spectrum (dashed and dotted line; test data, second example).

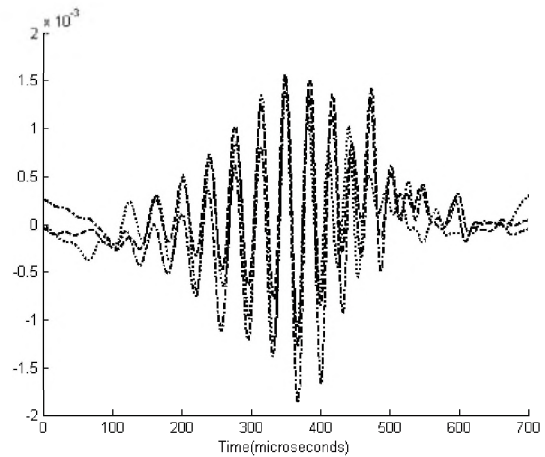


Figure 15: Temporal signal of the associated clicks of – probably – the Blainville's beaked whale (test data, second example).

6 CONCLUSIONS

This paper presents a novel application and extension of target-tracking technology to marine-mammal detection and classification; the paper extends our past work to include feature-aided tracking. The results are promising, and help in classifying beaked whales' clicks as well as to estimate the number of animals present.

Our approach is challenged when many animals are present, especially if they are not beaked whales. To improve the results, it would be helpful to identify features or processing methods to distinguish animals within the same species, and, more generally, to determine an improved methodology to assign the click feature vector. Improved feature vector information would directly support improved tracking performance.

Finally, the parameters in our kinematic modelling of ICI and amplitude dynamics should be species dependent. Thus, a coupled approach to kinematic and classification

filtering has the potential further to improve detection and classification performance.

7 ACKNOWLEDGMENTS

We thank the Woods Hole Oceanographic Institution (WHOI) and the Naval Undersea Warfare Center (NUWC) for providing datasets used in this study. We wish also to thank all the people involved in the organization of the *Third International Workshop on Detection and Localization of Marine Mammals using Passive Acoustics*, Boston, July 2007. And finally we thank the French Navy to have partly funded this study.

8 REFERENCES

1. D.A. Abraham, Passive acoustic detection of marine mammals, SACLANTCEN SM 351, 2000, NATO Undersea Research Centre, La Spezia, Italy.
2. R. Laterveer, S.P. Beerens, S.P. van Ijsselmuide, M. Snyder, J. Newcomb, A high frequency transient detector suitable for monitoring marine mammal activity, in Proceedings of UDT Europe, June 2004, Nice, France.
3. D.A. Abraham, P.K. Willett, Active Sonar Detection in Shallow Water using the Page Test, IEEE Journal of Oceanic Eng. Vol. 27 (1) January 2002.
4. O. Gerard, S. Coraluppi, W. Zimmer, and P. Willett, A Chorus of Whales: Evaluation of Sequential and Batch Approaches to Time-Series Tracking, in Proceedings of OCEANS 2006, September 2006, Boston MA, USA.
5. O. Gerard, C. Carthel, S. Coraluppi, and P. Willett, MHT and ML approaches to marine mammals detection, in Proceedings of OCEANS 2007, June 2007, Aberdeen, UK.
6. M. Johnson, P.T. Madsen, W.M.X. Zimmer, N.A. Aguilar de Soto and P.L. Tyack (2006), Foraging Blainville's beaked whales produce distinct click types matched to different phases of echolocation, The Journal of experimental biology 209 (24), 5038-5050.
7. S. Coraluppi and C. Carthel, Recursive Track Fusion for Multi-Sensor Surveillance, Information Fusion, vol. 5(1), March 2004.
8. S. Coraluppi and C. Carthel, "Distributed Tracking in Multistatic Sonar", IEEE Transactions on Aerospace and Electronic Systems, vol. 41(3), July 2005.
9. M. Johnson, P.T. Madsen, W.M.X. Zimmer, N.A. Aguilar de Soto and P.L. Tyack (2004), Beaked whales echolocate on prey, Proc. Roy. Soc. Lond. B 271, S383-S386
10. P.T. Madsen, M. Johnson, N.A. Aguilar de Soto, W.M.X. Zimmer, and P.L. Tyack (2005), Biosonar performance of foraging beaked whales (*Mesoplodon densirostris*), The Journal of experimental biology, 208, 181-194.
11. M.A. Roch, M.S. Soldevilla, J.C. Burtenshaw, R. Hoenigman, S.M. Wiggins, and J.A. Hildebrand., Comparison of machine learning techniques for the classification of echolocation clicks from three species of odontocetes, in Canadian Journal of Acoustics (current issue).



Photo Credit: Bahamas Marine Mammal Research Organisation

STATISTICAL CLASSIFICATION OF ODONTOCETE CLICKS

Douglas Gillespie and Marjolaine Caillat

Sea Mammal Research Unit, Gatty Marine Laboratory,
University of St. Andrews, Fife, KY16 8LB, Scotland

ABSTRACT

To the best of our knowledge, all odontocetes produce some kind of click like vocalisation, which is used primarily for echolocation but may also play a role in social communication. Characteristics of these echolocation pulses range from the broad band but relatively low frequency clicks of sperm whales to the ultrasonic, narrow-band clicks of harbour porpoise. Although these clicks are often easily detected, it can be difficult to classify them to species, thereby hampering efforts to monitor and study odontocetes using passive acoustics. Candidate clicks from three species were detected using a simple energy trigger, operating in the frequency band of interest. The clicks were then identified to species using two different statistical classifiers to separate beaked whale vocalisations from those of other odontocete sounds. In the first, a number of parameters (peak frequency, mean frequency, sweep frequency, click duration, width of principal spectral peak and the relative energy in different frequency bands) were calculated and a tree classifier was used to separate clicks of different species. In the second, the spectral energy in 32 relatively coarse energy bands 1.5 kHz wide were used as input to a multivariate classifier. Both classifiers were trained and tested using data provided to the 3rd International Workshop on Detection and Classification of Marine Mammals using Passive Acoustics in order to assess the classifiers performance with Blainville's beaked whales, short-finned pilot whales and Risso's dolphin clicks. The methods were also applied to survey data collected using a towed hydrophone deployed from a sailing research vessel in the Bahamas. Some of the towed hydrophone data were collected over the US Navy's AUTEK range where independent confirmation of beaked whale vocal activity was available from bottom-mounted hydrophones.

SOMMAIRE

L'état actuel de nos connaissances nous permet d'affirmer que tous les odontocetes émettent des sons de type impulsifs, aussi appelés clics, destinés surtout à l'écholocation, mais ils peuvent également être utilisés pour la communication. En fonction des espèces, ces clics peuvent couvrir une bande de fréquence plus ou moins large. Le cachalot produit des clics couvrant une large bande de basses fréquences, alors que chez le marsouin, l'écholocation est caractérisée par des clics ultrasoniques couvrant une bande de fréquence étroite. En sélectionnant les clics ayant une puissance supérieure à un certain seuil avec un simple détecteur d'énergie dans la bande de fréquence qui nous intéressait, nous avons collecté les clics de 3 espèces (Baleine à bec de Blainville, globicéphale tropical et dauphin de Risso). Deux méthodes d'analyse nous ont permis de discriminer les sons de la baleine à bec de Blainville de ceux des 2 autres espèces. Pour la première méthode, différents paramètres (pic de fréquence, fréquence moyenne, variation de fréquence, durée du signal, largeur du spectrogramme et énergie relative dans les différentes bandes de fréquences) ont été extraits de chaque clic et utilisés dans un arbre de classification afin de séparer les espèces. Pour la seconde méthode, l'énergie contenue dans 32 bandes de 1.5kHz a servi de données pour une analyse multivariée. Les 2 classificateurs ont été entraînés et testés en utilisant les données de la 3^{ème} commission internationale de détection et de classification des mammifères marins en utilisant l'acoustique passive. L'objectif était de mesurer la performance des classificateurs pour discriminer la baleine à bec de Blainville, le globicéphale tropical et le dauphin de Risso. Ces 2 méthodes ont ensuite été appliquées sur des données collectées au Bahamas à partir d'hydrophones tirés par un voilier de recherche. Quelques données furent collectées au dessus de la zone du réseau d'hydrophone Autek, appartenant à la marine Américaine, permettant d'obtenir une confirmation indépendante de l'activité sonore des baleines à bec via ce réseau sous-marin.

1. INTRODUCTION

Concern over the link between the use of military sonar and strandings of beaked whales has led to much research into the acoustic behaviour of beaked whales in recent years. Archival recording tags (Johnson and Tyack, 2003) attached to Blainvilles Beaked Whales (*Mesoplodon densirostris*) and Cuviers Beaked Whales (*Ziphius cavirostris*) (Johnson *et al*, 2004, Zimmer *et al*, 2005, Madsen *et al*, 2005) show that they produce narrow-band clicks with pulse lengths of around 200 μ s and most of the energy concentrated in the 25 to 40 kHz energy band.

Practical applications for the management of risks to beaked whales require that methods be developed to detect them more efficiently than can currently be achieved using visual observers (Barlow and Gisiner, 2006). Passive acoustic monitoring can potentially be used to assess the distribution and abundance of beaked whales and has also been proposed as a method for detecting animals in the immediate vicinity of vessels using sonar. Real time mitigation requires that beaked whale clicks be detected with a high efficiency, although it may not be necessary to identify to species level. Surveys of abundance do not require that detection efficiency be high, only that it be known, although species identification is more important than it might be for mitigation.

In this paper we demonstrate how beaked whale clicks may be detected and how they may be statistically separated from clicks from short-finned pilot whales (*Globicephala macrorhynchus*) and Risso's dolphins (*Grampus griseus*). Pilot whales produce both tonal vocalisations (whistles) and clicks (Weilgart and Whitehead, 1990). Both the whistles and clicks can be heard by humans since they are at lower frequency than beaked whale clicks. Risso's dolphins on the other hand echolocate at much higher frequencies. Risso's clicks are generally broad band, having energy between 30 and 100 kHz and an average 3 dB bandwidth of 39.7 kHz (Philips *et al*, 2003). The Risso's data analysed in this study were only sampled at 96 kHz and had been low-pass filtered at 38 kHz, and were therefore only acquiring the lower frequency components of the Risso's dolphin clicks.

2. METHODS

Detection and classification algorithms were trained and tested on data provided to the 3rd International Workshop on Detection and Classification of Marine Mammals using Passive Acoustics, which contained clicks from Blainvilles's beaked whales (BBW), short-finned pilot whales (SFP) and Risso's dolphins (RD). All of these data were in the form of wav file recordings, sampled at 96 kHz and containing data from a single bottom-mounted hydrophone either at the Bahamas Atlantic Undersea Test and Evaluation Center (AUTEK) range (BBW and SFP) or from the Southern California Operating Range (SCOR) in San Clemente Island (RD), California, USA.

Data were also available from a towed hydrophone array deployed from a sailing vessel undertaking line transect surveys in the Bahamas. Some of these data were collected at the AUTEK range, where animals were being simultaneously monitored on bottom-mounted hydrophones.

Most of the analysis was done using RainbowClick software (www.ifaw.org/sotw). RainbowClick was originally developed for the detection and analysis of sperm whale echolocation clicks (Gillespie, 1997). As well as containing click detection and classification algorithms, and applying them either to real time data or archived data from file, RainbowClick provides the user with an interactive display where detected clicks may be easily selected and their waveforms and spectra examined by the user. Clicks, or groups of clicks, can be exported via a database for more detailed analysis by other software packages (e.g. Matlab).

2.1. Click Detection

Detection and classification of clicks were conducted as clearly separate stages of the analysis. In the click detection stage, regions of sound files found to have significant energy in the 25 – 40 kHz band were extracted and stored.

Clicks were detected using an algorithm operating on time series data. The algorithm is designed for real time operation, using infinite impulse response filters (IIRF) for efficient data analysis (Lynn and Fuerst, 1989).

Optimal detection of beaked whale clicks is achieved by band-pass filtering the data in the 25 to 40 kHz range. However, the statistical classifiers (see below) require a comparison between acoustic energy within the 25 to 40 kHz beaked whale band and acoustic energy at lower frequencies. The detector therefore contains two separate filters as shown in Figure 1. The use of two filters allows the detector to operate only on signal within the band of interest, but data used in the classification stage can use signal in a wider band. Removal of low-frequency data, particularly at frequencies with wavelengths on the order of or longer than the clip length, is essential to avoid large sidelobes dominating the spectra used in the classifiers. The first filter was a second-order high-pass Butterworth with a cut-off frequency of 7 kHz and the second had a fourth order band-pass 25 kHz to 40 kHz Chebyshev response.

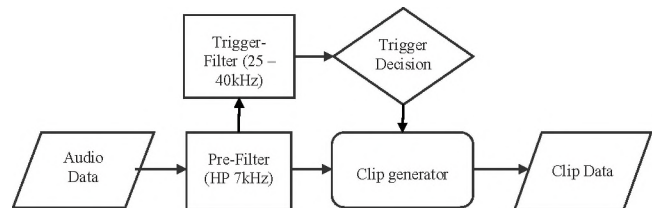


Figure 1. Schematic diagram of click detection process.

Training set extraction

Candidate clicks were detected in all samples of the training data set. To avoid occasional clicks from non-target species, or noise entering the training set, an operator (Caillat) then examined click files using RainbowClick in order to select a sub-sample of clicks which appeared to have consistency of amplitude and inter click interval with other clicks in the data (i.e. appeared to form part of an echolocation click train) and had waveform and spectral properties consistent with published literature for the three species. Numbers of clicks selected to form the training sample were BBW 6399, SFP 1555 and RD 609.

2.2. Click Classification – Method 1

Parameter Extraction

Beaked whale clicks are characterised by having most of their energy in the 25 to 40 kHz band. More detailed spectral analysis using a Wigner-Ville (WV) distribution shows that there is in fact a slight upsweep in frequency during a typical beaked whale click (Johnson *et al.*, 2006 and Figure 2).

Six parameters were measured for each candidate click from the detector. From the power spectrum, the mean frequency, the peak frequency and the ratio of the acoustic energy in the 25 to 40 kHz band compared with that in the 10 to 25 kHz band were measured.

From the WV distribution, the sweep of the click, the maximum ‘bandwidth’ of the click at any point along its length and the length of the click were extracted. This was

achieved by taking the maximum point in the WV distribution and then performing a regional search around it in order to establish a contour 6dB below the energy at the maximum (Figure 2). The ‘ridge’ of maximum acoustic energy along the length of the click was also extracted. The bandwidth was taken as the maximum distance between the lower and upper edges of the contour and the length of the click as the time between its start and its end. The sweep was taken as the difference from the start to the end of the maximum energy ridge. This broadly follows a pattern of parameter extraction found to be useful in detecting and classifying right whale contact calls (Gillespie, 2004).

Classification

The six parameters described above were computed for all clicks in the training data set. A classifier was then realised using a tree classification function (Breiman *et al.*, 1993). Tree classifiers divide data into groups using multiple binary splits. Each split uses a single variable or parameter to divide the data into two groups, the variable and the split value at each node being chosen to maximise the deviance between the two groups. This process is repeated until every group contains perfectly homogeneous data (i.e. clicks of only one type). For practical classifiers, the number of nodes is limited (the tree is ‘pruned’) to avoid problems of over-fitting to training data.

In order to establish which of the six parameters extracted for each click were most useful in classification, 11 different classification models were tested, each using a different sub selection of the six parameters (Table 1). Trees were pruned to five nodes. The training data were split and 2/3 of the

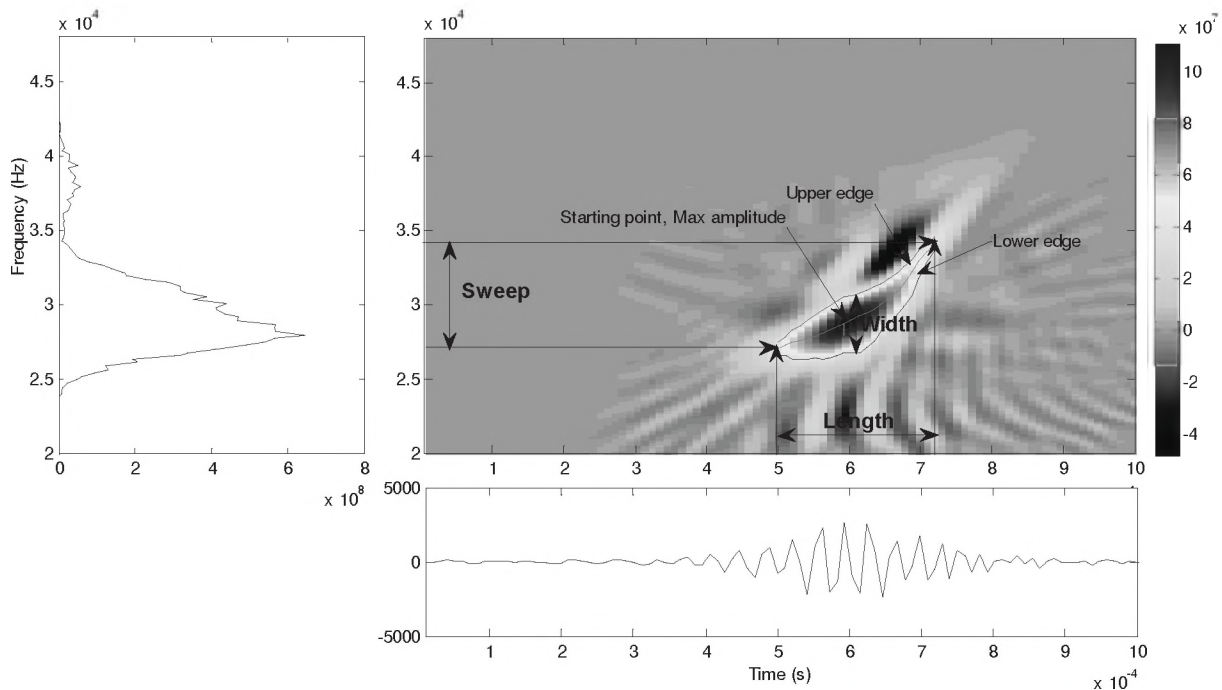


Figure 2 Wigner-Ville distribution, power spectrum and waveform for a typical beaked whale click showing the -6dB contour around the click and the ridge of maximum energy.

Table 1. Parameter selection for eleven different models tested using the tree classifier.

Parameter	Model										
	1	2	3	4	5	6	7	8	9	10	11
Length	√					√		√			√
Width	√	√				√		√			√
Sweep	√	√	√				√	√			√
Mean Frequency	√	√	√	√			√		√	√	√
Peak Frequency	√	√	√	√	√				√		√
Energy Ratio	√	√	√	√	√	√	√	√		√	

data randomly selected for classifier training. Equal numbers of clicks for each species were selected from the remaining 1/3 of the data for testing. Since we had the fewest clicks from Risso's dolphins (609) this meant that 203 clicks from each species were used in the testing samples. Bootstrapping was used to train and test each model 500 times, each time using a different random sample of training and test data and the average error rate from the 500 bootstraps taken.

2.3. Click Classification – Method 2

Parameterisation for the second classification method simply took the power spectrum for each click and divided it into coarse energy bins 1.5 kHz wide. A 512 point FFT was calculated for each click (each click clip either being truncated or padded with zeros to achieve the correct length) and the relative energy in bins 1.5 kHz wide (8 FFT bins) taken, i.e. if $S(\omega)$ is the power spectrum, then the coarse spectrum $S'(\omega)$ is

$$S'(\omega_i) = 10 \log_{10} \left(\frac{\sum_{8i}^{8(i+1)} S(\omega_i)}{\sum S(\omega)} \right) \quad \text{Equation 1.}$$

giving a total of 32 parameters for each click, although only those above 7.5 kHz were used in classification.

Clicks were classified using a one-way Multivariate Analysis of Variance (one way MANOVA) to divide training data into groups (Krzanowski, 1988.). The MANOVA calculates a linear discriminant function chosen to maximise the separation between groups and produces an matrix of eigenvalues which can be used to calculate canonical variables which are simple linear combinations of the parameters (relative energies in 1.5 kHz bands). These canonical variables can then be used to assign clicks to different groups using a relatively small number of variables.

Although such a classifier should work equally well, or better, with finer-scale data, the limited size of the training sample might have made the model over fitted to the available data if all 256 frequency bins had been used as input parameters.

3. RESULTS

3.1. Click Classification – Method 1

Distribution plots for the six parameters extracted for each species are shown in Figure 3. Most of the distributions overlap heavily, particularly for beaked whales and Risso's dolphins. Pilot whale clicks are generally at lower frequency than those of the other species, making them stand out on plots of mean and peak frequency as well as the energy ratio.

Results of the tree classification are shown in Figure 4. When all three species are analysed together, the error rate varies between approximately 20 and 30 % for the pruned tree. If only beaked whale and pilot whale clicks are included in the analysis, the error rate is extremely low.

In both cases, the worst performing model is model 9, which does not include the energy ratio (Table 1) thereby indicating that this is one of the more important parameters. Of particular interest for practical applications are models 4, 5, 9 and 10, which only use parameters extracted from the power spectrum. Computation of the Wigner Ville distribution takes approximately 2000 times longer than the calculation of a power spectrum, so any method that does

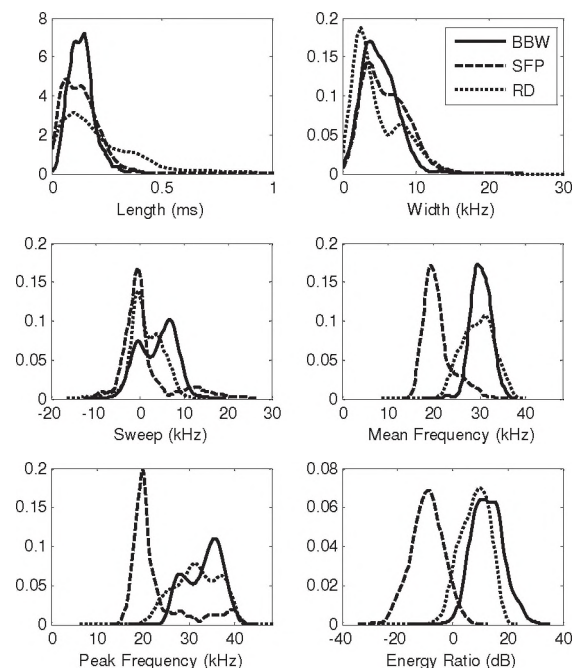


Figure 3: Distributions of the six parameters for each species.

not require those parameters will be much easier to implement in real time systems.

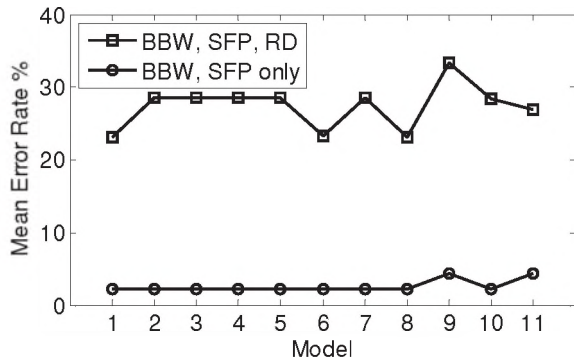


Figure 4: Tree classification error rates for the 11 models.

3.2. Click Classification – Method 2

Figure 5 shows the mean coarse spectra for different click types. Figure 7 shows distributions of the first two canonical variables from the MANOVA analysis of coarse spectra. Clearly, there is good separation between the three species when this method is employed although there is still slight overlap between beaked whales and pilot whales.

4. APPLICATION TO ‘UNKNOWN’ WORKSHOP DATA

The ‘Unknown’ workshop data were analysed using the second classification method since it had better overall performance for all three species than the first. During analysis, it was assumed that only the three species present in the training set were present. If C1 and C2 are the first two canonical variables (Figure 7) then clicks were classified according to the following selection criteria:

```

BBW if C1 > -0.5 and C2 < 2
else
SFP if C1 < -1.5 and C2 < 2
Else
RD if C2 > 3
Else
Click is unclassified.

```

Numbers of clicks of each type in each of the test files are shown in Table 2. Since the classifier was trained only to identify three species, spotted dolphins (*Stenella attenuata*) and sperm whales (*Physeter macrocephalus*) were misclassified as pilot whales. Clicks from both these species have their predominant energy below 20 kHz (Lammers *et al* 2003; Mohl *et al*, 2003) so this misclassification is not surprising.

Click waveforms and spectra were viewed with RainbowClick. Of particular interest are the small numbers of beaked whale clicks in data sets 3 and 4. These beaked

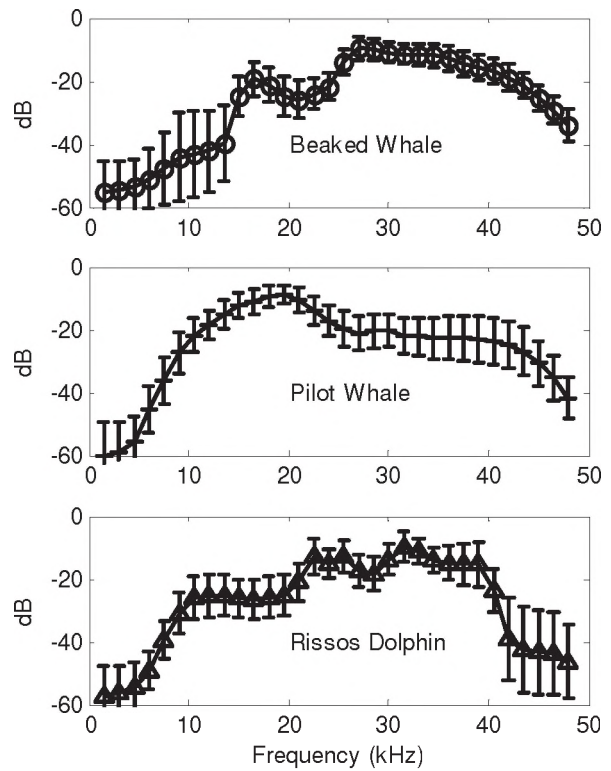


Figure 5: Mean coarse spectra for the second click parameterisation.

whale clicks appeared in short bursts of 3 – 4 clicks and are quite clearly from beaked whales. However, this is not the case for the clicks classified as beaked whales in other files which, on visual inspection of waveforms and spectra, are clearly false classifications.

5. APPLICATION TO BAHAMAS TOWED HYDROPHONE DATA

As well as being tested with the workshop dataset, the detector and classifiers were tested using data collected using a towed hydrophone deployed from the sailing research vessel ‘Odyssey’ while undertaking line transect surveys around the Bahamas in June and July, 2007. Recordings were made at a sample rate of 192 kHz.

The second classification method did not work at all well with the towed hydrophone data, classifying large numbers of false triggers from vessel noise as beaked whales. The tree classifier on the other hand did perform well and picked out a number of beaked whale click trains, some of which were coincident with clicks being detected on the bottom-mounted hydrophones at AUTECH. In all, 172.5 hours of recordings were analysed. Initial processing to detect and classify clicks took approximately one week. It then took an operator (Gillespie) one day to go through the data and confirm beaked whale detections. A typical click from the towed hydrophone is shown in Figure 6.

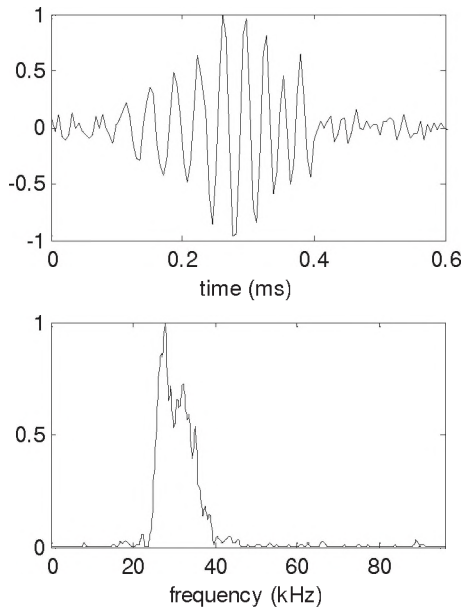


Figure 6: Waveform and power spectrum of a beaked whale click recorded on the towed hydrophone.

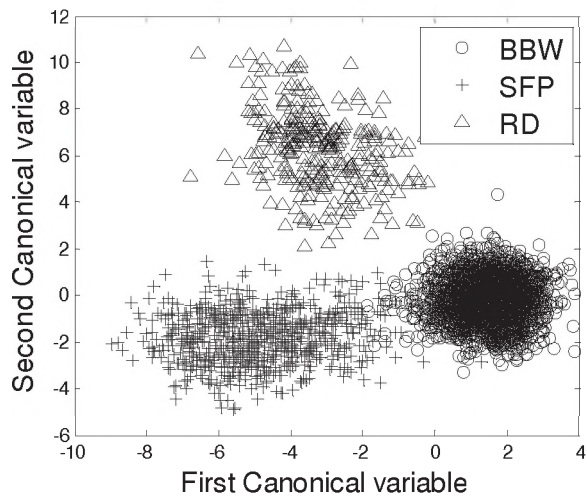


Figure 7: distributions of the first two canonical variables from the MANOVA analysis.

6. DISCUSSION

Clicks from beaked whales, pilot whales and Risso's dolphins can be detected and separated using statistical classifiers operating on parameters extracted from the clicks power spectra. Two methods have been tested, the first of which performed less well with the test data from bottom-mounted hydrophones than the second method, but was more stable when applied to data collected on a towed hydrophone array.

Table 2: Clicks identified within the 'unknown' test data files.

Set	Number of clicks				Truth
	BBW	SFP	RD	No Class	
1	1502	531	5	542	BBW+SFP
2	1566	1000	5	131	BBW
3	11	15200	0	24	Spotted*
4	13	16674	0	28	Spotted*
5	13	462	10879	466	RD
6	947	1217	2	30	BBW
7	154	104	3414	758	RD
8	994	4749	5	2391	SFP
9	0	4909	1	30	Sperm

*Visual inspection of these data following analysis with RainbowClick shows clicks which clearly matched published waveforms and spectra for BBW.

Using the first method, when all three species were analysed together, the error rate was around 20 to 30 % for the pruned tree. Although this appears high, if multiple clicks were analysed together (as they generally are) this might improve. However this depends on whether or not the errors are being caused by random noise (in which case we would expect an improvement) or genuine overlap in click parameters in which case the improvement may be small since adjacent clicks from the same whale are likely to be very similar.

One of the reasons why the first method did not perform particularly well at separating BBW and RD clicks is probably that the RD echolocate at higher frequencies than could be represented in recordings sampled at 96 kHz (Philips et al, 2003), so the workshop data sets only contained the very low edge of the RD spectra. Had higher frequency recordings been available, it is likely that separation of these species would have been as straightforward as the separation of SFP and BBW. Given that many researchers are able to collect data at a sample rate of only 96 kHz, asking whether or not it is possible to separate BBW and RD in such data is a valid question. Unfortunately, however, the RD data had been collected at a different location and heavily filtered above 38 kHz, so if there was a difference, it was lost in these data. Conversely, and perhaps more worrying is the possibility that the detectors somehow 'tuned in' to this fundamental difference in the recordings rather than differences in the sounds themselves.

When applied to the 'unknown' data set, the detector found beaked whale like clicks in a number of recordings. Visual inspection of waveforms, power spectra and inter-click intervals convinces us that some of these are genuine beaked whale clicks. Although no visual sightings of beaked whales were made, these species are notoriously difficult to spot and it is quite possible that some were present, even though they were not seen.

A fundamental problem with classifiers trained to detect only a small number of species is that they will tend to mis-

classify data from any other source (be it noise from a vessel or some other cetacean species). For classifiers of the type presented here to be genuinely useful, they must be trained with data from all species and noise sources likely to be present in the data to which they are to be applied.

Although beaked whales have been successfully detected on bottom-mounted hydrophones, beaked whales echolocate only when undertaking deep foraging dives (Tyack *et al.*, 2006). Since their clicks are produced in a narrow, forward-facing beam (Zimmer *et al.*, 2005), it has therefore been suggested that detection using towed hydrophones close to the surface is unlikely. When applied to data collected using a towed hydrophone, the detector and classifier were able to pick out several beaked whale click trains. This result is extremely encouraging and opens up the possibility for towed hydrophone surveys for beaked whales. However, further work is required to establish with what efficiency this can be achieved as a function of detection range.

As well as separating beaked whale clicks from those of dolphin species (SFP and RD), some applications, such as abundance estimation, may require the separation of species within the beaked whale family. Although Johnson *et al.* (2004), show clear differences between clicks of Blainville's and Cuvier's beaked whales, vocal behaviour of many other beaked whale species remains largely undocumented. Another research priority for the coming years is therefore to obtain broadband recordings from other beaked whale species.

The click detector has been implemented into the PAMGUARD open source software (www.pamguard.org) and work is underway to implement the classification methods.

7. ACKNOWLEDGEMENTS

This work was funded by the Joint Grants Scheme operated by the Defence Science and Technology Laboratory and the Natural Environment Research Council. We wish to thank the organisers of the 3rd International Workshop on Detection and Classification of Marine Mammals using Passive Acoustics for the provision of the training data sets and Dianne Claridge, Jonathan Gordon and other participants in the Bahamas Marine Mammal Research Organisation surveys for the towed hydrophone data. RainbowClick software development was funded by the International Fund for Animal Welfare to promote benign and non-invasive research.

8. REFERENCES

Barlow, J., and Gisiner, R. 2006. Mitigating, monitoring, and assessing the effects of anthropogenic sound on beaked whales. *J. Cetacean Res. Manage.* 7.3, 239–249.

Breiman, L., Friedman, J., Olshen, R., Stone C. 1993. *Classification and Regression Trees*, Chapman and Hall, Boca Raton.

Gillespie D. 1997. An acoustic survey for sperm whales in the southern ocean sanctuary conducted from RSV *Aurora Australis*. Reports of the International Whaling Commission, 47, 897–906.

Gillespie, D. 2004. Detection and classification of right whale calls using an 'edge' detector operating on a smoothed spectrogram. *Canadian Acoustics* 32(2): 39–47

Johnson, M., and Tyack, P. L. 2003. A digital acoustic recording tag for measuring the response of wild marine mammals to sound. *IEEE J. Ocean. Eng.* 28, 3–12.

Johnson, M., P. T. Madsen, W. M. X. Zimmer, Aguilar. de Soto, N. and P. L. Tyack. 2004. Beaked whales echolocate on prey. *Proceedings of the Royal Society of London Series B-Biological Sciences* 271:S383-S386.

Johnson M., Madsen P.T., Aguilar de Soto N., Tyack P. 2006 Foraging Blainville's beaked whales (*Mesoplodon densirostris*) produce distinct click types matched to different phases of echolocation, *J. Exp. Biol.*, 209 5038-5050.

Krzanowski, W. J. 1988, *Principles of Multivariate Analysis*. Oxford University Press.

Lammers, M., Au, W., Herzing, D. 2003. The broadband social acoustic signaling behavior of spinner and spotted dolphins. *J. Acoust. Soc. Am.* 114(3), 1629-1639.

Lynn, P., and Fuerst, W. 1989. *Introductory Digital Signal Processing with Computer Applications*. John Wiley & Sons, New York

Madsen P.T., Johnson M., Aguilar de Soto N. 2005 Biosonar performance of foraging beaked whales (*Mesoplodon densirostris*) *J. exp. Biol.* 208, 181-194.

Möhl, B., Wahlberg, M., Madsen, P.T., Heerfordt, A. & Lund, A., 2003. The monopulsed nature of sperm whale clicks. *J. Acoust. Soc. Am.* 114, 1143–1154.

Philips, J. Nachtigall, P., Au, W., Pawloski, J., Roitblat, H. 2003. Echolocation in the Risso's dolphin, *Grampus griseus*. *J. Acoust. Soc. Am.* 113 605-616

Tyack P.L., Johnson M., Aguilar de Soto N., Sturlese A., Madsen P.T.M.2006. Extreme diving of beaked whales, *J. Exp. Biol.*, 209. 4238-4253

Weilgart, L.S. and Whitehead, H. 1990. Vocalizations of the North Atlantic pilot whale (*Globicephala melas*) as related to behavioral contexts. *Behav. Ecol. Sociobiol.* 26: 399-402.

Zimmer, W., Johnson, M., Madsen, P., Tyack, P. 2005. Echolocation clicks of free-ranging Cuvier's beaked whales (*Ziphius cavirostris*). *J. Acoust. Soc. Am.*, 117(6) 3919-3927

PROCESSING THE WORKSHOP DATASETS USING THE TRUD ALGORITHM

Edward Harland

Chickerell BioAcoustics, Chickerell, Dorset, UK

ABSTRACT

The Transient Research Underwater Detector (TRUD) is designed to search for echolocation clicks from marine mammals. It uses a spectrogram correlation method with a set of reference matrices to search for clicks from multiple species. This paper describes the algorithm and presents the results of processing the workshop datasets from the Third International Workshop on the Detection and Classification of Marine Mammals using Passive Acoustics held in Boston in July, 2007. The work shows that TRUD can detect and classify the target species. Recommendations are made for further improvements to the algorithm.

SOMMAIRE

Le Transient Research Underwater Detector (TRUD) est conçu pour rechercher les clics d'écholocation émis par certains mammifères marins. Il utilise une méthode basée sur la corrélation de spectrogrammes avec un ensemble de matrices de référence pour rechercher les clics de plusieurs espèces. Cet article décrit l'algorithme et présente les résultats de la basée de données proposée par le 3rd International Workshop on the Detection and Classification of marine mammals using Passive Acoustics qui s'est tenu à Boston en juillet 2007. Ce travail montre que TRUD permet de détecter et de classer les espèces de cétacés que l'on recherche. Des recommandations sont formulées pour des améliorations futures de cette approche.

1. INTRODUCTION

There is an increasing need to monitor for the presence of acoustically-sensitive species such as marine mammals. Examples of such a need include as a precursor to the operation of high power sound sources (Tasker, 1998) or as part of a site survey leading to offshore installations such as wind farms or wave/tidal generators (Madsen, 2006).

One option for detecting the presence of marine mammals is to detect, classify, and, if possible, localise their calls. Ideally this process should be completely automatic. This process should have a very low false alarm rate combined with an acceptable probability of detection.

Marine mammals generally make three classes of calls, narrow bandwidth, medium bandwidth, and high bandwidth calls. Narrow bandwidth calls are tonal signals that can be processed using high-resolution FFT techniques to produce a spectrogram. Image processing techniques can then be applied to detect tonals and measure parameters about the signal to allow classification. Medium bandwidth signals are roars or grunts that have a significant instantaneous bandwidth and are prolonged in time. High bandwidth signals are typically the echolocation clicks used by odontocetes. These are very short in duration and occupy bandwidths of several octaves.

The Transient Research Underwater Detector (TRUD) was designed to detect high bandwidth signals only. It is intended to be used in conjunction with other

medium and narrow bandwidth processing such as the MMADS system (Harland and Armstrong, 2004).

When designing such a classification system it is not necessary to classify every single click. Some clicks get distorted by acoustic propagation that results in miss-classification. It is necessary to look at the ensemble of clicks and make a majority decision on the single click classifications. It can be expected that any individual click may pass multiple species classification tests, but with varying degrees of confidence and any classifier must look across all of these outputs to arrive at the final classification decision.

This paper presents the background to the TRUD system, sets out how it operates and then presents the results of processing the workshop dataset for the Third International Workshop on the Detection and Classification of Marine Mammals using Passive Acoustics held in Boston, USA, in July 2007.

2. THE TRUD ALGORITHM

The TRUD algorithm is part of a suite of processing packages designed for the site characterisation role. Other packages process the medium and narrow bandwidth signals and also characterise the ambient noise levels. The whole package is aimed at stand-alone applications such as pre-installation monitoring for wind farms or tidal generators. However, the suite can also be used for real-time monitoring applications and the results presented in near real-time to an operator.

Previous work by the author led to the prototype version of the Porpoise Detector (POD) currently available from Chelonia in the UK (Tregenza, 1998). This prototype system used analogue processing for detection and classification of the harbour porpoise (*Phocoena phocoena*) clicks and suffered from a number of problems inherent in analogue systems including low dynamic range and filter mismatch causing false detections. More recently an attempt was made to produce an improved system using digital processing and this resulted in the Simple Porpoise Underwater Detector (SPUD) (Harland, 2007). SPUD is based on the spectrogram correlation system proposed by Mellinger and Clark (Mellinger and Clark, 2000). The SPUD system was then made more general by using multiple reference matrices and evolved into the TRUD algorithm described here.

Figure 1 shows the block diagram of the TRUD processing chain. The incoming datastream is processed in 128 kilosample blocks using a 64 point FFT with Blackmann-Harris weighting and 75% overlap. These settings were chosen as a good compromise between time and frequency resolution for use with the range of pulse lengths of echolocation clicks. The resulting spectrogram is then searched for clicks using a sparse reference matrix. There is one matrix for each species. The reference matrix of weighting coefficients is cross-multiplied on a cell by cell basis with the spectrogram to form the classification factor at each time increment.

$$clf = \sum_{f_r, t_r} s_{f,t} * w_{f,t}$$

where s is the spectrogram value at f,t and w is the weighting coefficient at f,t within the reference matrix.

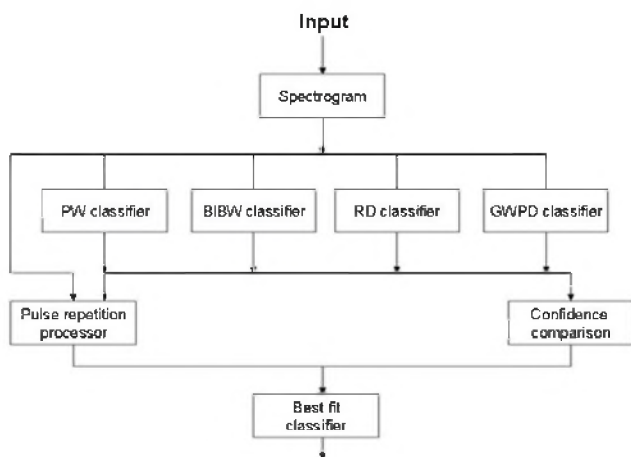


Figure 1 TRUD processing chain

The reference matrix can be up to 16 time samples by 31 frequency bins in size. The reference matrix is then moved along the time axis of the spectrogram, repeating the cross-multiplications at each sample interval to form

the classification factor. By using negative and positive weightings classification occurs whenever the classification factor is positive.

In addition to the individual species reference matrices a General Wideband Pulse Detector (GWPD) is also implemented. This has two functions. It acts as a catch-all detector so clicks from species whose calls are not documented are not missed, and it can also be used as a pre-processor to minimise power consumption. The aim is to process the data initially using only the GWPD and to only search with the reference matrices when cetacean-like clicks are encountered. The GWPD uses energy summation over the frequency range 15-45 kHz and compares this with similar sums 5 samples ahead and behind the summation point. The detection threshold is chosen to be 10dB.

The outputs of each of the individual species classifiers are compared to decide which species any one click originated from. In order to aid this comparison it is better to use a confidence factor rather than the classification factor. The confidence factor is normalised and independent of the amplitude of the input click and reflects the degree of confidence in the single click classification. For the work described here the confidence factor was calculated as the ratio of the classification factor to the same summation carried out across only those cells with positive weighting. The confidence factor varies from 0 to 1 depending on the signal/noise ratio and the degree of match to the reference matrix.

The single click classification is then combined with the pulse train processing output and compared with references sets of expected parameters for individual species to form the classification decision. In a fully implemented system this would then be combined with the outputs from narrow and medium bandwidth classifiers and used with geographic information to arrive at the final classification decision.

TRUD is currently implemented in MATLAB (Mathworks, Release 14) running on a PC.

3. TRAINING TRUD

The initial training of the TRUD algorithm was carried out using a dataset from Blainville's beaked whale (*Mesoplodon densirostris*) recorded on dTAGs from animals off El Hierro in the Canary Islands (Johnson *et al.*, 2005). This dataset is available from the MOBYSOUND website (Mellinger and Clark, 2006). This dataset has the advantage of a high signal/noise ratio and a sampling frequency of 192 kHz. The high sample rate is important as it means the data contains all parts of the click signal. Past experience working with harbour porpoise clicks suggest that the algorithm works best when a sample of the spectrum above the call is available as part of the classification test.

The optimum reference matrix was determined by a manual iterative trial and error method. As may be

expected with such a high signal/noise ratio, 100% detection was achieved with no false alarms.

The reference matrix was then truncated at the lower Nyquist frequency used by the workshop dataset and used to process a selection of the training files from the workshop dataset known to contain calls from Blainville's beaked whale. It soon became clear that this reference matrix was sub-optimal because of the characteristics of the workshop datasets (see below). The reference matrix was then re-optimised using the same manual iterative trial and error method. This resulted in a significant improvement in the detection rate and a lowering of the false positive rate.

An example of a reference matrix is shown in figure 2 for Blainville's beaked whale. The areas shown in black have a positive weighting, while those shown in grey have a negative weighting. Blank cells are ignored and not used in the calculation.

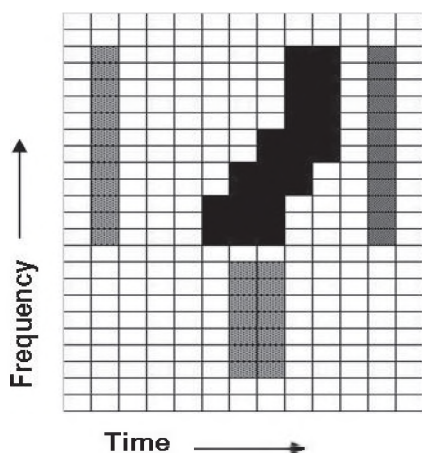


Figure 2 Reference matrix for Blainville's beaked whale

Different regions of the matrix are weighted depending on the frequency response of the system and the characteristics of the target species signal. There is also an additional differential weighting between the black and grey areas to implement the detection threshold. For these tests it was chosen to be 10 dB. The cell by cell weighted cross multiplies are summed together and a detection occurs when the sum is positive.

The same methodology was then applied to generating reference matrices for the other target species: pilot whale (*Globicephalus macrorhynchus*) and Risso's dolphin (*Grampus griseus*). No published information on the clicks of the pilot whale could be found so the reference matrix was initially chosen as a wideband pulse with spectral content from 15-48 kHz. The matrix was then optimised from the training dataset.

The Risso's dolphin click classifier was trained to look for the off-axis clicks. Madsen *et al* (Madsen *et al*, 2004) described the clicks of this species and showed that the off-axis spectral content was significantly different from the on-axis click. The training dataset contains

examples of both types of click, but the off-axis type predominates.

From the testing it became clear that rejection of a non-target species click was as important as accepting the target species click. As an example, the pilot whale reference was adjusted to minimise false classification of some sperm whale clicks. A further round of optimisation was then carried out to improve the false positive rate at the expense of a small reduction in the true positive rate.

The GWPD reference matrix was also tested and partially optimised during this process to improve rejection of non-cetacean clicks.

4. THE DATASET

The dataset consisted of two groups of files. The training files were fully annotated with the sounds in the files, while the test files had no annotation. The aim was to train the classifiers using the training set and then use the trained classifiers to search and classify the test files. All files were sampled at 96 kHz and saved as WAV format single channel files.

During the training process described above it became clear that there were characteristics of the workshop dataset that were impacting operation of the TRUD algorithm and these are described below:

- a) The low sample rate results in loss of the high frequency components of the call, reducing classification performance
- b) The limited available bandwidth does not allow testing of the spectrum above the call which can result in a higher rate of incorrect classifications
- c) System non-linearity introduces artefacts when signal levels are sufficiently high to drive the data collection system into non-linearity. This could be caused by clipping or slew-rate limitations (see figure 4). This can lead to incorrect classification
- d) Different hydrophone channels have different characteristics, requiring a modified reference matrix to maintain optimum performance
- e) Some clicks suffer significant dispersion due to acoustic propagation (see figure 3) which can lead to incorrect classification

The effect of the limited bandwidth and reduced spectral test is to reduce the effectiveness of the TRUD algorithm. Regrettably time did not permit an evaluation of the degree of degradation. The non-linearity was only a problem at high signal levels and resulted in a number of incorrect classifications when the animal was close to a hydrophone.

The reference matrix is optimised for a particular acoustic environment. This can be achieved by either pre-whitening the background or by optimising for a particular hydrophone system. The latter is the simpler method and was chosen for this test. Unfortunately the acoustic background was not consistent across the

different hydrophones. This resulted in a number of incorrect classifications.

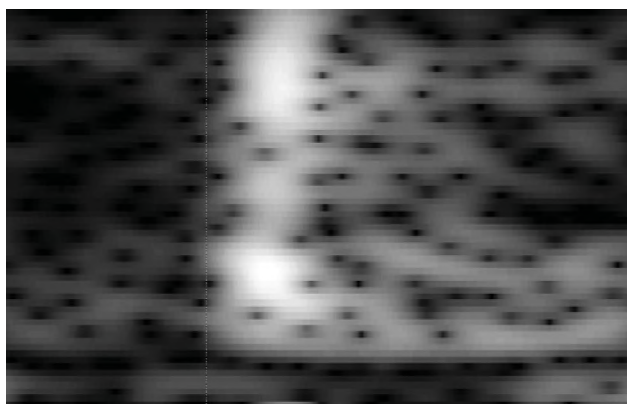


Figure 3 Spectrogram of pilot whale click with dispersion

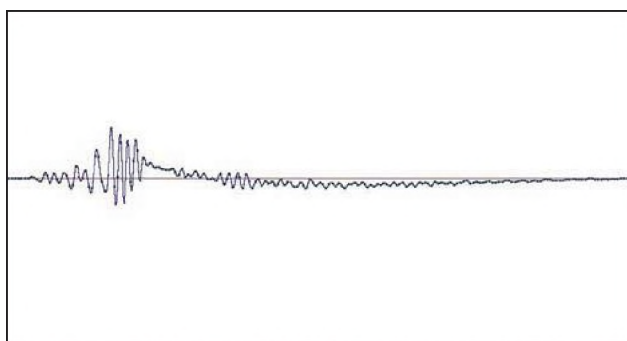


Figure 4 Click with DC shift caused by overloading

Some of the clicks suffered significant dispersion (see figure 3). It was not clear where this was occurring. It was probably predominantly acoustic dispersion but analogue transmission over long cables to the shore may have contributed to the effect. The effect was to turn the wideband clicks of species such as pilot whales into up-sweeps. This effect resulted in false negatives and/or false positives under some conditions.

5. PROCESSING THE DATASET

On completion of the training sequence the TRUD system was then used to process the test dataset. The initial processing suggested that there were other signals in the test files than those declared in the training dataset. This was seen as high GWPD counts but low counts from the target classifiers. Visual/aural checking suggested that the majority of these were sperm whale clicks and this was further confirmed by testing with a previously developed variant of the SPUD system optimised to detect regular sperm whale clicks.

Because of memory size limitations in the computer used to process the data, only five minutes of data from each file were processed. The cumbersome file names in the dataset were discarded and replaced with the simple terminology used in the following tables. The first five files for each of the species as listed in the dataset were used. The nomenclature used is that the column headed

GWPD are the results for the general wideband pulse detector, BBW are the results for the beaked whale detector and PW are the results for the pilot whale detector. In later tables, RD are the results for the Risso's dolphin detector.

The results of processing the training set are shown in the following tables:

Table 1. Processing the pilot whale files

File	GWPD	BBW	PW
Pilot1	424	<i>11</i>	205
Pilot2	645	<i>14</i>	399
Pilot3	3313	<i>12</i>	1495
Pilot4	2023	<i>5</i>	880
Pilot5	1780	<i>508</i>	1140

The counts shown in bold are true-positives, the counts in italics are false-positives. The numbers in the GWB column are a guide to the possible number of clicks in the file. The level of false-positives are well within acceptable levels except for Pilot5. The pulses in Pilot5 are distorted and also suffer dispersion, resulting in the high number of false positive detections of Blainville's beaked whale and an increased number of missed pilot whale detections.

Table 2. Processing the Blainville's beaked whale files

File	GWPD	BBW	PW
BIBW1	91	128	<i>1</i>
BIBW2	244	330	<i>19</i>
BIBW3	603	917	<i>1</i>
BIBW4	1082	1432	<i>2</i>
BIBW5	1003	1429	<i>0</i>

It should be noted that for this species the number of detections by the GWPD is fewer than for the Blainville's beaked whale detector. This is due to the characteristics of the clicks from this species which are not a good match to the GWPD. The level of false positive detections of pilot whales is well within acceptable limits.

Table 3 Processing the Risso's dolphin files

File	GWPD	BBW	PW	RD
Risso1	6340	<i>37</i>	<i>78</i>	86
Risso2	2816	<i>868</i>	<i>255</i>	685
Risso3	5330	<i>501</i>	<i>1480</i>	6004
Risso4	14497	<i>2339</i>	<i>7522</i>	10814
Risso5	297	<i>5</i>	<i>3</i>	69

The Risso's dolphin reference matrix was not available until a few days before the workshop due to time constraints so it was not possible to process all the training set with this classifier. Neither was it possible to complete the optimisation procedure. Nevertheless, the performance is generally satisfactory.

The test files were then processed to give the following results. NP denotes that the data was not processed. The Test1 and Test2 files were processed in two blocks; a is 0-5 minutes and b is 5-10 minutes.

Table 4 Processing the test files

File	GWPD	BBW	PW	RD
Test1a	2902	2466	93	397
Test1b	1037	612	91	436
Test2a	5276	1328	376	100
Test2b	7870	986	200	88
Test3	21007	182	7989	NP
Test4	15213	110	3445	NP
Test5	5576	2663	665	6188
Test6	960	1237	110	NP
Test7	2884	1773	252	2965
Test8	4184	1915	1620	NP
Test9	1673	0	361	5

Test8 has distorted clicks due to both electrical and propagation effects. Test9 appears to be all sperm whale clicks. The PW classifications in Test9 occur on very strong sperm whale clicks with energy to 40 kHz.

A visual inspection of the test files reveals that a number of species are present. If the TRUD classification criteria is such that the species chosen is the classifier with the most outputs then these classifications are compared with the manual results in table 5. UNK are pulses from an unknown species. Note that the sperm whale classifier was not running for these tests so TRUD should not have found these clicks.

Table 5 Comparison of TRUD and visual results

File	TRUD	Visual
Test1a	BIBW	BIBW
Test1b	BIBW	BIBW, Unk
Test2a	BIBW	BIBW, Unk
Test2b	BIBW	BIBW, Unk
Test3	PW	PW
Test4	PW	PW
Test5	RD	RD
Test6	BIBW	SW, BIBW
Test7	RD	RD
Test8	BIBW	PW
Test9	PW	SW

The only incorrect classification is in Test8 where pilot whale calls are classified as Blainville's beaked whale. Table 4 shows that this was a marginal decision and inspection of the file suggests that it is caused by the propagation distortion of the pulses in this file.

A detailed inspection of TRUD operation for the Test1 and Test2 files shows that the unknown pulses were rejected by all three classifiers. The high false-positive count for Risso's dolphin is caused by propagation distortion of the Blainville's beaked whale clicks.

6. PULSE TRAIN TESTING

In addition to the single pulse classification testing, the pulses were associated into trains and the statistical properties measured to aid classification. A simple pulse train follower was written using a two parameter pulse

association test (time and amplitude). The pulse interval was then measured between each successive pulse in the pulse train and the histogram plotted. The training dataset was used to gather the statistics for each of the target species. However, some of the files could not be used because the simple train follower could not cope with the multiply-interleaved pulse trains.

Once the histograms had been built for each of the species declared in the training files, a similar set of measurements were made for the test files and the results compared as shown in figure 5, 6, and 7.

This suggests that Test1 and Test6 contained Blainville's beaked whale, Test2 appears to have two species present, of which one is Blainville's beaked whale, and Test7 is Risso's dolphin. Test8 is a good match to pilot whale. Test6 also contains another species with a low repetition rate as can be seen in figure 7. From table 5 it is likely that these are the sperm whale pulses. Test3, Test4, Test5, and Test9 were not processed.

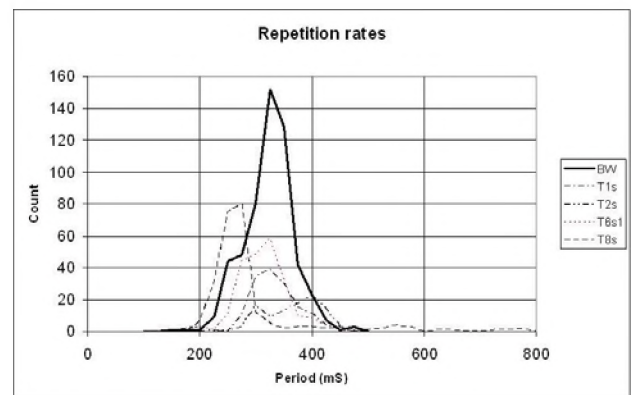


Figure 5 Beaked whale pulse train results

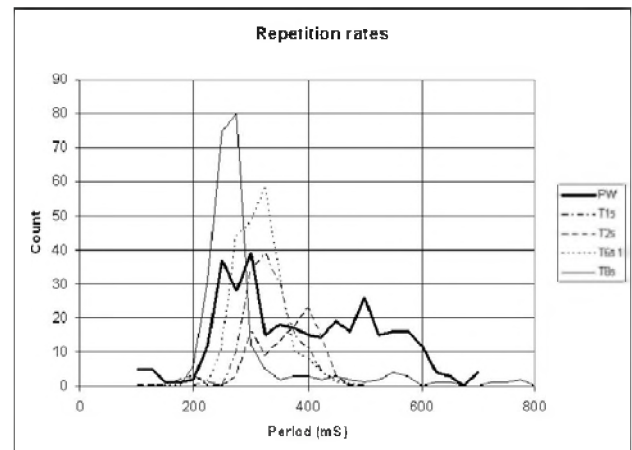


Figure 6 Pilot whale train processing results

These classifications were determined by visual comparison of the curves. Work is still in progress to carry out this comparison automatically.

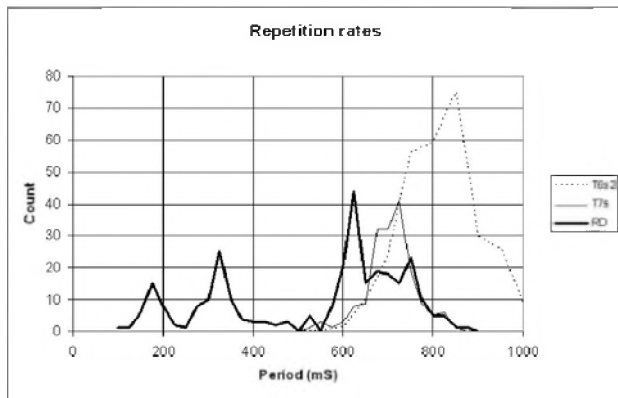


Figure 7 Risso's dolphin pulse train processing results

7. DISCUSSION

It must be emphasised that the results presented here are for the first stages in a multi-stage classification process. It is not possible, except in a very limited number of cases, to design a classifier that will uniquely identify the species that originated any one click. The complete classification process consists of building up a weight of evidence leading to a 'best guess' at the species. The results presented here are based on the classification factor output. Many of the false positives will be eliminated when the confidence factor is used and pulse association processing eliminates multiple classifications on a single pulse.

As an example, many of the incorrect Risso's dolphin classifications listed in the results tables are caused by a low confidence classification immediately before a high confidence beaked whale classification caused by the partial overlap of the reference matrix and the signal. Similarly, a number of beaked whale and Risso's dolphin false classifications are caused by the reverberation tail from a pilot whale click. Both of these would be eliminated by pulse association processing.

The pulse train classification needs a lot more data to fully define the pulse interval reference statistics for each species. The work here shows that it can be a useful classification aid for the specific circumstances of the training and test datasets, but much more work is needed to explore how useful it is across the full acoustic and geographic ranges of an individual species.

The present version of TRUD provides useful initial classification processing stages. The single pulse classification and pulse train statistics are two of the most important factors in the weight of evidence processing and this testing has shown that TRUD has the potential to fulfil this role. However, the testing has also shown that improvement is desirable in a number of areas:

- a) For best performance the incoming data needs to be of a high quality with good linearity and a bandwidth sufficient to allow all of the spectrum tests. When used in its intended role as a stand-alone site monitor

it will generally be possible to provide the requisite high quality input stage.

- b) The Risso's dolphin reference matrix needs to be further optimised to reduce the false alarm rate from distorted Blainville's beaked whale clicks.
- c) For best efficiency the general wideband pulse detector needs to perform better. The aim is to run only the GWPD while searching for clicks and to only activate the more detailed classification processes when candidate clicks are found that pass this lower threshold.
- d) TRUD alone cannot provide a unique classification for many species. It will be able to classify to an acoustic clade level, but to refine the classification decision it will need to be combined with medium and narrow bandwidth classifiers.

8. CONCLUSIONS

This work has shown that spectrogram correlation is a viable classification method for echolocation pulses of the species in the workshop dataset. It has also shown that pulse train processing can aid the classification process for echolocation click sequences.

9. REFERENCES

- Harland, E.J. (2007) The performance of the SPUD algorithm detecting harbour porpoise (*phocoena phocoena*) echolocation clicks. Third International conference on Bioacoustics. April 2007. Institute of Acoustics, Loughborough University, UK.
- Harland, E.J., Armstrong, M.S. (2004) The real-time detection of the calls of cetacean species. Canadian Journal of Acoustics, 32: 76-82
- Johnson, M., Madsen, P., Tyack, P., de Soto, N.A. (2005) A binaural acoustic recording tag reveals details of deep foraging in beaked whales. Journal of the Acoustical Society of America, 117: 2524
- Madsen, P.T., Kerr, I., Payne, R. (2004) Echolocation clicks of two free-ranging, oceanic delphinids with different food preferences: false killer whales *Pseudorca crassidens* and Risso's dolphins *Grampus griseus*. J. Exp. Biol. 207: 1811-1823
- Madsen, P.T., Wahlberg, M., Tougaard, J., Lucke, K., Tyack, P. (2006) Wind turbine underwater noise and marine mammals: implications of current knowledge and data needs. Marine Ecology Progress Series 309: 279-295
- Mellinger, D.K., Clark, C.W. (2000) Recognizing transient low-frequency whale sounds by spectrogram correlation. Journal of the Acoustical Society of America 107: 3518-3529

Mellinger, D.K., Clark, C.W. (2006) MobySound: A reference archive for studying automatic recognition of marine mammal sounds. Applied Acoustics Special Issue: Detection and localization of marine mammals using passive acoustics 67: 1226-1242

Tregenza, N. (1998) Site acoustic monitoring for cetaceans - a self-contained sonar click detector. In: Tasker, M.L., Weir, C. (eds) Proceedings of the seismic and marine mammals workshop. London, 23-25th June 1998. Sea Mammal Research Unit, St Andrews, UK.

Tasker, M. (1998) Guidelines for minimising acoustic disturbance to marine mammals from seismic surveys. JNCC, Aberdeen, UK.

EDITORIAL BOARD / COMITÉ EDITORIAL

ARCHITECTURAL ACOUSTICS: ACOUSTIQUE ARCHITECTURALE:	Vacant		
ENGINEERING ACOUSTICS / NOISE CONTROL: GÉNIE ACOUSTIQUE / CONTROLE DU BRUIT:	Colin Novak	University of Windsor	(519) 253-3000
PHYSICAL ACOUSTICS / ULTRASOUND: ACOUSTIQUE PHYSIQUE / ULTRASONS:	Werner Richarz	Aeroustics	(416) 249-3361
MUSICAL ACOUSTICS / ELECTROACOUSTICS: ACOUSTIQUE MUSICALE / ELECTROACOUSTIQUE:	Annabel Cohen	University of P. E. I.	(902) 628-4331
PSYCHOLOGICAL ACOUSTICS: PSYCHO-ACOUSTIQUE:	Annabel Cohen	University of P. E. I.	(902) 628-4331
PHYSIOLOGICAL ACOUSTICS: PHYSIO-ACOUSTIQUE:	Robert Harrison	Hospital for Sick Children	(416) 813-6535
SHOCK / VIBRATION: CHOCS / VIBRATIONS:	Li Cheng	Université de Laval	(418) 656-7920
HEARING SCIENCES: AUDITION:	Kathy Pichora-Fuller	University of Toronto	(905) 828-3865
HEARING CONSERVATION: Préservation de L'Ouïe:	Alberto Behar	A. Behar Noise Control	(416) 265-1816
SPEECH SCIENCES: PAROLE:	Linda Polka	McGill University	(514) 398-4137
UNDERWATER ACOUSTICS: ACOUSTIQUE SOUS-MARINE:	Garry Heard	DRDC Atlantic	(902) 426-3100
SIGNAL PROCESSING / NUMERICAL METHODS: TRAITMENT DES SIGNAUX / METHODES NUMERIQUES:	David I. Havelock	N. R. C.	(613) 993-7661
CONSULTING: CONSULTATION:	Corjan Buma	ACI Acoustical Consultants Inc.	(780) 435-9172
ADVISOR: MEMBER CONSEILLER:	Sid-Ali Meslioui	Pratt & Whitney Canada	(450) 647-7339

A NOVEL MULTI-CLASS SUPPORT VECTOR MACHINE CLASSIFIER FOR AUTOMATED CLASSIFICATION OF BEAKED WHALES AND OTHER SMALL ODONTOCETES

Susan Jarvis^{2,1} Nancy DiMarzio¹ Ronald Morrissey¹ and David Moretti¹

1- Naval Undersea Warfare Center Division, Newport, RI, 02841

2- Worcester Polytechnic Institute, 100 Institute Rd., Worcester, MA 01609

ABSTRACT

Navy sonar has recently been implicated in several marine mammal stranding events. Beaked whales (particularly *Mesoplodon densirostris*) have been the predominant species involved in a number of these strandings. Monitoring and mitigating the effects of anthropogenic noise on marine mammals are active areas of research. Key to both monitoring and mitigation is the ability to automatically detect and classify animals, especially beaked whales. This paper presents a novel support vector machine based methodology for automated, species level classification of small odontocetes. The new classifier, called the class-specific support vector machine (CS-SVM), consists of multiple binary SVM's where each SVM discriminates between a class of interest and a common reference class. A main objective in the development of the CS-SVM was to realize a robust multi-class SVM whose implementation is simpler than existing multi-class SVM methods. A CS-SVM was trained to identify click vocalization from four species of odontocetes including *Mesoplodon densirostris*. The algorithm processes time series data in a fully automated fashion first detecting and then classifying click events. Results from the application of this automated classifier to the data sets provided by the 3rd International Workshop on Detection and Classification of Marine Mammals Using Passive Acoustics are presented.

SOMMAIRE

Le sonar a été récemment associé à un certain nombre d'événements de mammifère marin immobilisé en eau peu profond. Les Baleines à bec (en particulier le *Mesoplodon densirostris*) ont été les espèces prédominantes impliquées dans un certain nombre d'événements d'immobilisation. La surveillance et l'atténuation des effets du bruit synthétique sur les mammifères marins sont des domaines de recherche actifs. Ce qui est importante de la surveillance et la réduction des effets est la capacité automatiquement de détecter et classifier des animaux, particulièrement les baleines à bec. Cet article présente une nouvelle méthodologie basée sur une machine de support vecteur (SVM) pour automatisé le classification de niveau d'espèces de petits odontocetes. Le nouveau classificateur, appelé le "class-specific support vector machine" (CS-SVM), est composé de SVM binaire multiple où chaque SVM se distingue entre une classe d'intérêt et une classe commune de référence. Un objectif principal dans le développement du CS-SVM était de réaliser une multi-classe robuste SVM dont l'exécution est plus simple que des méthodes existantes de la multi-classe SVM. Un CS-SVM a été formé pour identifier le vocalisation de clic de quatre espèces des odontocetes incluant des *Mesoplodon densirostris*. Les données de série chronologique de processus d'algorithme sont traitées d'une mode entièrement automatisée détectant d'abord et classifiant ensuite des événements de clic. Les résultats de l'application de ce classificateur automatisé fournis par le "Troisième Atelier Internationale de Détection, Localisation, et Classification du Mammifères Marins avec les Acoustiques Passive" sont présentés.

1. BACKGROUND

Until quite recently, little was known about the vocalizations of beaked whale. However, starting with the definitive recording of beaked whale clicks by Johnson, Tyack, et al. (using non-invasive DTAG's) [1,2] and continuing with the visually verified recording of beaked whales and other small odontocete vocalizations at AUTECH [3] there is now sufficient labeled data available to develop automated classification algorithms. To foster exchange of

ideas and classification methodologies, the 3rd International Workshop on Detection and Classification of Marine Mammals Using Passive Acoustics was convened. In addition to providing a venue for scientific exchange in the topic areas, the workshop provided a data set [12] consisting of both labeled training data for 3 species of odontocetes and unlabeled test data. This paper investigates the application of a novel class-specific support vector machine classifier to the classification of vocalizations from beaked whales and other odontocetes specifically using the data set provided by

the Workshop.

At a basic level, a classification system is one that assigns the current input \mathbf{x} membership into one of v known classes according to some set of decision metrics or functions. In general, \mathbf{x} is a multivariate random variable where $\mathbf{x} \sim P(\mathbf{x})$. For example, popular maximum likelihood classifiers assign an input data vector \mathbf{x} membership in one of v possible class hypotheses $\{H_1, H_2, \dots, H_j, \dots, H_v\}$ according to the probabilistic rule $j^* = \arg \max(p(H_j|\mathbf{x}))$. This is equivalently written as $j^* = \arg \max(p(\mathbf{x}|H_j)p(H_j))$ after applying Bayes rule. Theoretically, a maximum likelihood (ML) classifier is optimal in that it offers the lowest probability of error of any classifier [4]. However, in practice, it can be difficult to attain this optimal performance because the multidimensional probability density functions $p(\mathbf{x}|H_j)$ are unknown and must be estimated from training data. The amount of training data required to estimate $p(\mathbf{x}|H_j)$ grows exponentially with the dimension of \mathbf{x} . This is problematic because the collection of labeled training data is usually difficult, time consuming and expensive.

Statistical learning theory [5,6] represents a different paradigm for learning than the classical ML methods presented above. Statistical learning theory advocates solving specific problems directly vice solving more general problems as an intermediate step [5]. That is, if there are limited data available to train a classifier then the best course of action is to estimate a decision boundary directly from the data. This is in contrast to classical ML inference where the data are used to estimate the parameters of density functions and then the PDFs are used to form decision boundaries.

2. DISCUSSION

One of the corner stones of statistical learning theory is the principle of structured risk minimization (SRM). Using the SRM principle, Vapnik developed a bound on the risk of classification error for a given decision function f given the empirical risk (training error) $R_{emp}(f)$ associated with the function, the training set size m , and the capacity h of the hypothesis space in which the decision function resides [6]. This bound (1) is often referred to as the guaranteed risk, and is independent of the underlying distribution of the data. According to the SRM principle, the smallest bound on classification error is achieved by minimizing training error while using the function hypothesis space of the smallest capacity [5,6].

$$R[f] \leq R_{emp}[f] + \sqrt{\left(\frac{1}{m} \left(h \left(\ln \frac{2m}{h} + 1 \right) + \ln \frac{4}{\delta} \right) \right)} \quad [6] \quad (1)$$

Support vector methods (or support vector machines, SVM) are a rich family of learning algorithms based on statistical learning theory. SVM's were originally developed to solve binary classification problems of the following type: Given a set of training data $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_m, y_m)\}$ where each (multidimensional) input example x_i drawn from X is associated with classification label $y_i = \pm 1$, determine the decision function that maps any new \mathbf{x} drawn from X to $y = \pm 1$ that minimizes risk of misclassification [5]. In short,

SVMs implement the SRM principle.

SVM's use the existence of a unique optimal hyperplane which separates the two classes in some feature space (figure 1). The SVM that implements the optimal hyperplane while maximizing the separation (margin) between the two classes will have the lowest risk of test error [5]. This optimal separating hyperplane is realized as

$$f(\mathbf{x}) = \sum_{k=1}^m \alpha_k y_k G(\mathbf{x}, \mathbf{x}_k) + b \quad (2)$$

where G is a kernel mapping and b is an offset. The weights α_k for a "soft" margin SVM classifier [6] are found through the optimization (3)

$$\text{maximize } W(\boldsymbol{\alpha}) = \sum_{k=1}^m \alpha_k - \frac{1}{2} \sum_{k,l=1}^m \alpha_k \alpha_l y_k y_l G(\mathbf{x}_k, \mathbf{x}_l)$$

$$\text{subject to } \boldsymbol{\alpha} \in \mathbb{R}^m, 0 \leq \alpha_k \leq C/m \text{ and } \sum_{k=1}^m \alpha_k y_k = 0.$$

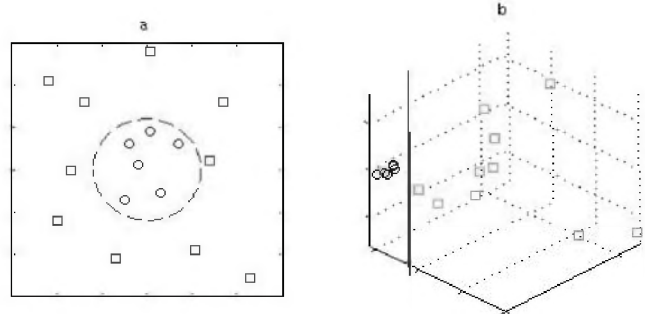


Figure 1: A notional view of a SVM [6]. a) Training data drawn from \mathbf{x} shows two classes. b) A transformation $T(\mathbf{x})$ maps the training data to a higher dimensional space where the optimal separating hyperplane is found. The hyperplane in the higher dimensional space corresponds to a non-linear decision boundary in the input space.

The constant C controls the degree of "slack" in the hyperplane optimization. Large C corresponds to more rigid separation of the classes and less tolerance for class overlap in the training data. Smaller C allows for more class overlap in the training data [7]. Equation (3) can be solved using quadratic programming techniques [6].

While SVM's were originally formulated for binary classification, many real world problems involve more than two classes. As a result, a number of methods have been developed for applying SVM's to multi-class problems. These methods tend to follow one of three basic approaches. The first approach is to form v binary "one-against-the-rest" classifiers (where v is the number of class labels) and choose the class whose decision function is maximized [5]. The second approach is to form all $v(v-1)/2$ pairwise binary classifiers and choose the class whose set of pairwise decision functions are in some way maximized [7]. The third approach is to reformulate the objective function of the SVM for the multi-class case such that the decision boundaries for all classes are optimized jointly [8].

This paper presents a new type of multi-class support vector classifier called the class-specific SVM (CS-SVM). The new classifier consist of v binary SVM's where each

SVM discriminates between one of v classes of interest and a common reference class. The class whose decision function is maximized with respects to the reference class is selected. The CS-SVM extends the concept of exploiting class-specific features as proposed by other researchers for maximum likelihood classifiers [4] and neural networks [9] to the multi-class SVM problem.

Many applications involve the classification of signals which are set in additive noise. In such cases, the problem is not to differentiate between two or more of v signals present at the same time but to differentiate between one of v signals and noise. The input vectors for such problems are actually of the form $x_j = s_j + n$, for $j = 1, 2, \dots, v$. Currently, SVM's are designed assuming the classification problem is to distinguish $x_i = s_i$ from $x_j = s_j$. Any noise in x is assumed to be accommodated by allowing the "slack variables" in the hyperplane optimization [6].

The CS-SVM expressly acknowledges the presence of the noise by treating it as a common reference class. For a single class, the classification problem reduces to a detection problem, a decision as to whether signal s is present or not. That is, $y = \text{sgn}(f(x)) = +1$ when $x = s + n$ and $y = \text{sgn}(f(x)) = -1$ when $x = n$. In the multi-class case, x is assigned membership in the class whose decision function $f_i(x)$ against the reference is maximum, or to the noise-only class when all $f_i(x) < 0$. Note that in acknowledging the presence of a common reference class no assumptions are made about that class. Although it is intuitive to think of the reference class as Gaussian noise, the reference class could be of any arbitrary distribution. This means that the CS-SVM can actually act as a signal detector for signals of unknown distribution set in noise of unknown distribution.

Figure 2 is an notional illustration of the CS-SVM concept for two dimensional data. Optimal separating hyperplanes for each class versus the noise-only reference class are found. Since the optimal hyperplane separating any two classes is unique [5], the optimal hyperplane for class i versus n will be different from the optimal hyperplane for class j versus n . However, both hyperplanes are optimized against a common reference class. The decision function $f_i(x)$ for either signal-present class should reject the noise-

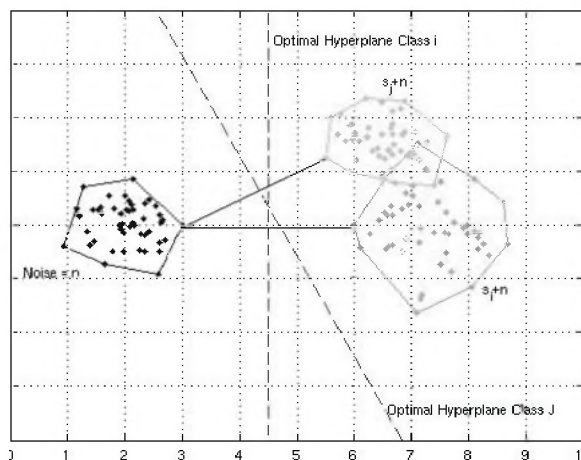


Figure 2: A geometric view of the optimal separating hyperplanes for two CS-SVMs for class i and class j , respectively, versus a common reference class in a 2-D decision space.

only reference case. Further, it is hypothesized that $f_i(x)$ will be greater than $f_j(x)$ whenever x is draw from class i since $f_i(x)$ is optimal for class i and $f_j(x)$ is not.

3. EXPERIMENTAL RESULTS

In the past several years there has been much interest and progress in acoustic monitoring, localization and tracking of marine mammals [3,10]. Acoustic monitoring has a number of benefits over visual monitoring. Chief among them are increased area of coverage and the ability to operate over wider weather conditions and at night. A major drawback of acoustic monitoring is associating species information with the received vocalizations. However, recent field tests combining visual verification and digital recording tags with acoustic monitoring and localization have resulted in sets of labeled acoustic data [3]. One such data set was provided as part of the 3rd International Workshop on Detection and Classification of Marine Mammals Using Passive Acoustics [12]. This section presents the development of a CS-SVM classifier using the Workshop data set.

The data set provided for the Workshop consisted of labeled training data for 3 species as well as unlabeled test data. Training data was supplied for *Mesoplodon desirostris* (Blainville's beaked whale), *Globicephala macrorhynchus* (short-fin pilot whale) and *Grampus griseus* (Risso's dolphin). The training data for each species consisted of five or more .wav files with each file containing 2 to 3 minutes of 16-bit audio data sampled at 96KHz. The test data consisted of nine longer .wav files (each 10+ minutes) also sampled at 96KHz. Within the 9 test files there were examples of the species alone, examples containing a mix of species as well as examples containing none of the 3 species given in the training data. Prior to analysis or processing, all data were passed through a 12 KHz high-pass filter.

3.1 Training Data and CS-SVM Feature Selection

The first challenge in working with the Workshop data was deciding which events and signal features the classifier should be trained to recognize. Design of a classification algorithm generally requires selection a set of distinguishing features q_i to represent the raw data such that the input vector to the classifier is $x = [q_1 \ q_2 \ \dots \ q_n]^T$. Ideally the feature set should be a sufficient statistic for the raw data but it also must be of reasonably low dimension as the amount of training data required grows with the dimension of x .

One goal for the CS-SVM classifier presented here is for it to become incorporated into the acoustic marine mammal monitoring system, M3R [3,10]. This means that the classifier would have to be fully automated and run in real-time. In turn, that required selection of features that can be readily extracted "on the fly". Thus, it was decided that the classifier should classify individual click events rather than attempting to analyze click trains.

In previous experiments [11], the times between consecutive zero crossings were successfully used as features for classifying odontocete clicks. A zero crossing detector is easy to implement and the periods between

crossings capture the time-frequency structure of the signals. Additionally, the envelope shape of the clicks can be captured by using the normalized peak values between crossings (figure 3).

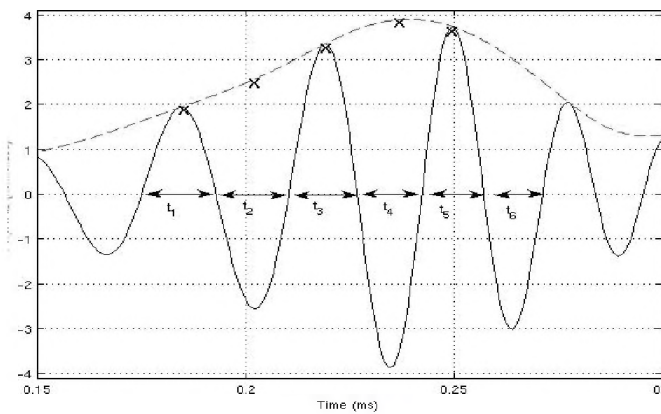


Figure 3: The times between zero-crossings and normalized envelope amplitudes were selected as features for classifying individual clicks.

The next step in the training process was to analyze the training data from each species to select the set of click events to be used in training the CS-SVM. The idea was to select several hundred representative clicks from each species, then to extract the zero-crossing and amplitude envelope features from them. Time-frequency analysis of the training data for *Mesoplodon desirostris* identified two distinct click waveforms, foraging clicks and buzz clicks (figure 4). A large majority of the clicks were the stereotypical foraging clicks, but several buzzes were also detected [13]. Since the foraging and buzz click waveforms are distinct, separate CS-SVM's were trained for the two click types.

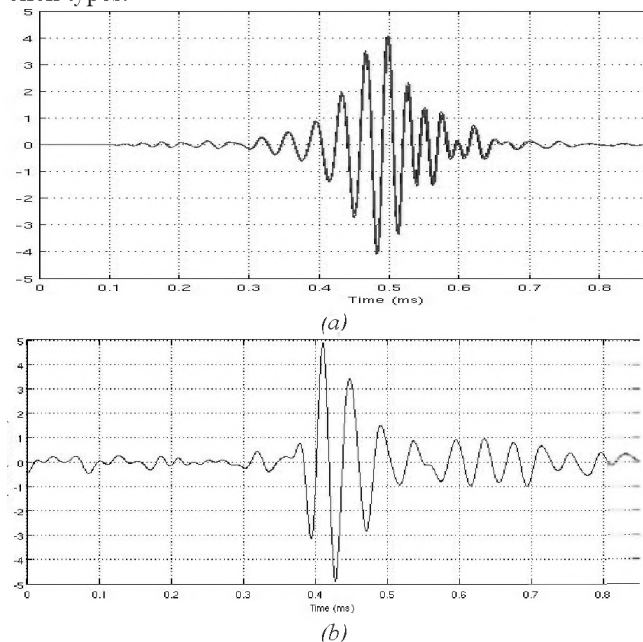


Figure 4: (a) Overlay of time series data for 3 *Mesoplodon desirostris* foraging clicks. (b) Time series of a single buzz click.

Time-frequency analysis of the training data from *Globicephala macrorhynchus* and *Grampus griseus* showed the clicks contained in those files to be highly variable. In contrast to the *Mesoplodon* data where the regularity of click waveforms is almost uncanny, it was difficult to identify representative click waveforms in the pilot whale and Risso's dolphin data (figure 5). After cross correlating all the clicks extracted from the training files, the click from each species most highly correlated with most of the other clicks was selected as a replica. Then, the training clicks which were most highly correlated with the replicas were used to build the training sets.

Nine dimensional feature vectors were formed using the times between 6 zero-crossings about the peak and three normalized envelope amplitude peaks. The resulting feature sets from the *Globicephala* and *Grampus* did not cluster as compactly as those from *Mesoplodon*. There was also significant overlap of the features for *Mesoplodon* and *Grampus* in the feature space (figure 6).

One binary SVM was constructed for each signal class versus an ambient noise reference class. The training set T_j for the j^{th} CS-SVM was defined as

$$T_j = \{(\mathbf{x}_p, y)\} = \{(\mathbf{s}_j + \mathbf{n}, 1), (\mathbf{n}, -1)\} \text{ for } j = 1 \text{ to } 4.$$

The training sets for each of the four class-specific SVM consisted of approximately 250 signal-present vectors and a similar number of noise-only vectors. These training sets were used in the optimization (3) to find the optimal hyperplane $f_j(\mathbf{x})$ for each class. A Gaussian radial basis function was used as the kernel in (2) and (3) yielding

$$f_j(\mathbf{x}) = \sum \alpha_k y_k \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_k\|^2}{2\sigma}\right) + b$$

where $k \in S_j$ the set of support vectors class j .

The four class CS-SVM classifier was then tested using additional clicks extracted from the training data files. These test sets also consisted of approximately 250 signal-present vectors and 250 noise-only vectors (Table 3). The classification performance for the test sets was evaluated using the following metrics, P_{cc} = fraction correctly classified (signal present), P_{miss} = fraction misclassified (signal present) and P_{nsc} = fraction correctly classified (noise-only). The results were encouraging, especially for the Meso-plodon foraging click class. Note that the poorer performance of the buzz click class was attributed to changes in buzz click waveform observed as the inter-click interval decreases during prey capture.

3.2 CS-SVM Results for the Test Data Files

Finally, the CS-SVM classifier was tested using the unlabeled test files. A cursory manual review of the test data prior to processing indicated the presence of sperm whale clicks in multiple test files. Although not part of Workshop training data, a class-specific SVM for sperm whale clicks was trained using additional labeled sperm whale data. The ability to add a class without affecting SVM design for the other classes highlights one of the

strengths of the CS-SVM approach. However, to be consistent with the data conditioning stream for the other classes, the sperm whale data was passed through the same 12 KHz high pass filter. This was known to be a suboptimal processing step as most of the energy in sperm whale clicks is typically below 12 KHz. A better solution for a CS-SVM classifier that includes a sperm whale class would be either to lower the frequency of the high-pass filter or to process in multiple frequency bands.

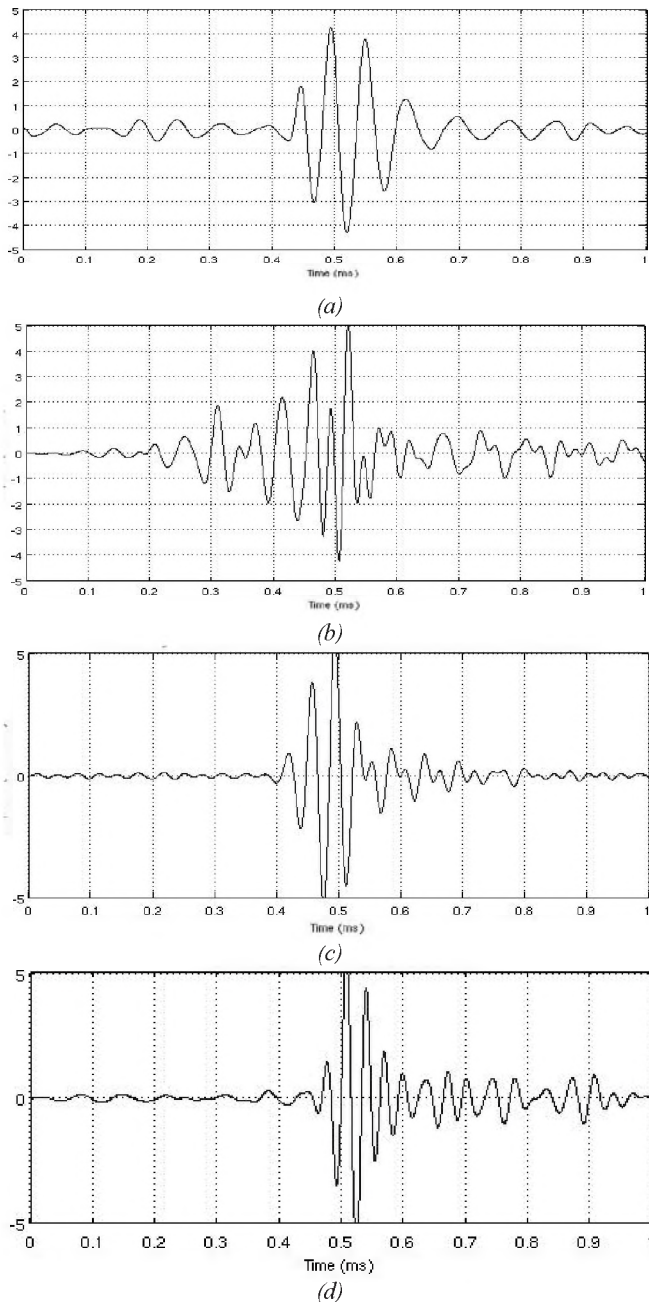


Figure 5: Clicks showing some of the variability in the *Globicephala macrorhynchus* (a-b) and *Grampus griseus* (c-d) training data.

Class	P_{cc}	P_{miss}	P_{nse}
<i>Mesoplodon</i> (forage)	1.0000	0.0000	0.9680
<i>Mesoplodon</i> (buzz)	0.7900	0.2100	0.9600
<i>Globicephala</i>	0.9380	0.0620	0.9682
<i>Grampus</i>	0.9451	0.0549	0.9721

Table 3: Performance of the 4 class CS-SVM on test sets of approximately 250 signal-present vectors and 250 noise-only vectors drawn from the training data files for each class

The nine test data files were processed in a fully automated fashion. The classifier program automatically read the data from the .wav files, filtered it, and performed time domain energy detection to identify click events. Time series data about the energy detector peaks were used to construct feature vectors. The feature vectors x extracted for each click event were then used to evaluate the class-specific decision functions, $f_i(x)$. The click event was assigned membership in the class whose decision function was maximum or to the noise-only reference class when $\max\{f_i(x)\} \leq 0$. Figure 7 shows the output of the 5 class-specific decision functions for data in the neighborhood of a *Mesoplodon* foraging click (from Test File 1). Although $f_4(x)$ associated with Risso's dolphin also peaked, $f_1(x)$ for the foraging click was maximum. Figure 8 shows the class decision output of the CS-SVM for Test File 1. This test file contained clicks from both *Mesoplodon* and *Globicephala*. Table 4 summarizes the performance of the 5-class CS-SVM for all of the Workshop test data.

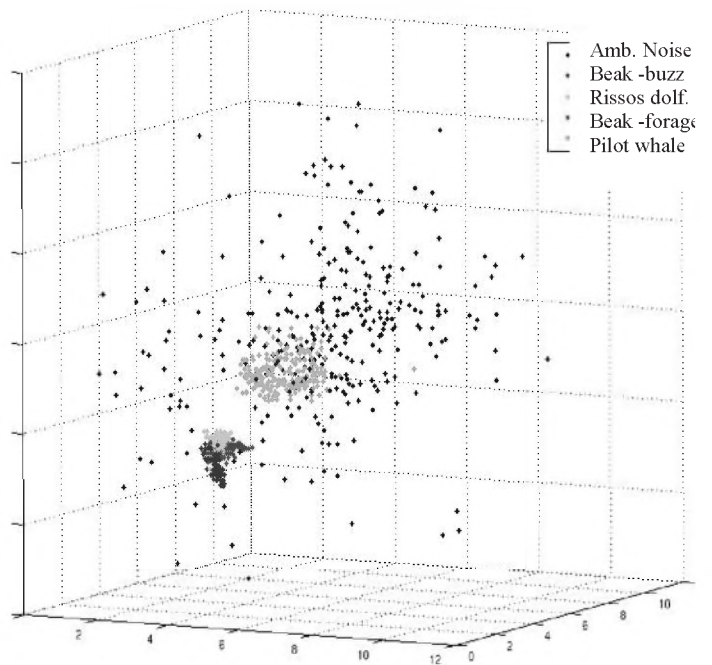


Figure 6: Distribution of the periods between three consecutive zero-crossings for ambient noise, *M. desirostris* (buzz), *Grampus griseus*, *M. desirostris* (forage), and *Globicephala macrorhynchus*.

4. CONCLUSION

This paper has presented a novel multi-class support vector machine classifier, the class-specific SVM. The new classifier consists of v binary SVMs where each SVM discriminates between one of v classes of interest and a common reference class. Test inputs are assigned membership in either the class whose decision function is maximized or the reference class if all decision functions are negative. A five class CS-SVM was created to classify broadband click vocalizations from several species of odontocetes using data provided by the 3rd International Workshop on Detection and Classification of Marine Mammals Using Passive Acoustics. While the CS-SVM's classification performance was quite good for species specific test cases drawn from the labeled training data, its classification performance was not as good for the unlabeled test data files. Some classes, like the *Mesoplodon densirostris* foraging class and the Sperm whale class, performed well on the test files but the performance for the other classes not as reliable.

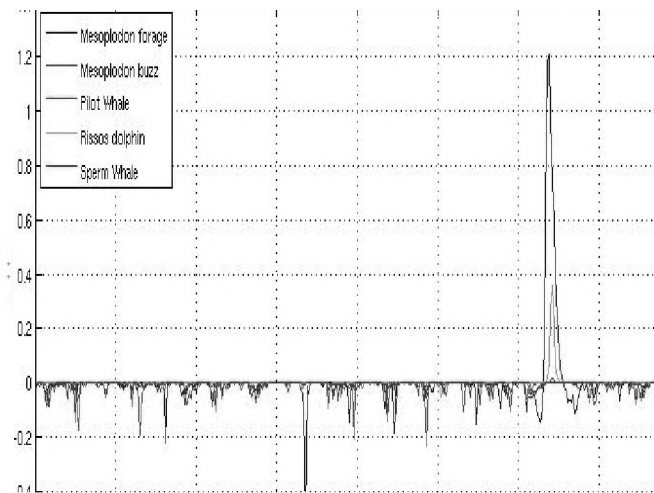


Figure 7: Output of the decision functions $f_i(x)$ vs time for the 5-class CS-SVM processing a data stream containing a *Mesoplodon* foraging click.

This difference in classification performance for the test data files is most likely a reflection of the feature sets chosen. The zero crossing and amplitude features used were very distinctive for the stereotypical *Mesoplodon* foraging clicks, but less distinctive for the other classes. In particular, there was significant overlap in the feature space between the *Mesoplodon* foraging clicks and the *Grampus* clicks. As a result, many *Grampus* clicks were misidentified as foraging clicks. Further analysis of *Grampus* vocalizations and modification of the feature set is recommended. Another issue in classification of the test data was the lack of a “none of the above” designation. The CS-SVM always assigned detected events to one of the 5 classes or to the noise-only class. The addition of a second level of thresholding of the decision functions, such as requiring $\max\{f_i(x)\} > L$, would reduce misclassification of unknown signal-present events.

5. ACKNOWLEDGEMENTS

The authors would like to acknowledge the on-going support of the Office of Naval Research. Additionally, we would like to thank Diane Claridge and her team of observers from the Bahamas Marine Mammal Research Organization, and Dr. Mark Johnson, Dr. Peter Tyack and their DTAG team from Woods Hole Oceanographic Institution. We also would like to thank AUTEK for allowing us access to the range and its assets. A special thanks to the following AUTEK personnel: Jose Arteiro, Marc Ciminello and Tom Szlyk. Finally, a sincere *merci* to Dr. Alex Wyglinski of Worcester Polytechnic Institution for his French translation of the abstract.

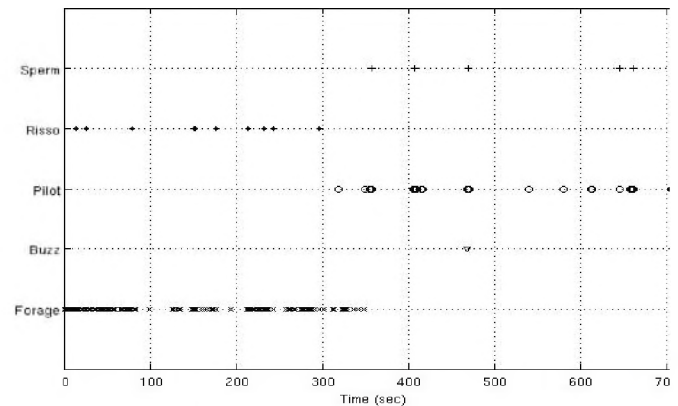


Figure 8: Class decisions from the 5-class CS-SVM for 579 click events from Test File 1.

Test Files	5-class CS-SVM Results
File 1 = Unverified Blainville's Beaked Whale and Short-finned Pilot Whale	Correctly identified as Blainville's beaked whale and Short-finned pilot whale
File 2 = Blainville's Beaked Whale	Correctly identified as Blainville's beaked whale
File 3 = Pantropical Spotted Dolphin	Incorrectly classified as Short-finned pilot whale
File 4 = Pantropical Spotted Dolphin	Incorrectly classified as Short-finned pilot whale
File 5 = Risso's Dolphin	Incorrectly classified as Blainville's beaked whale
File 6 = Unverified Blainville's Beaked Whale & Sperm Whale	Correctly identified as Blainville's beaked whale and sperm whale
File 7 = Risso's Dolphin	Incorrectly identified as a mix of Mesoplodon forage & Mesoplodon buzz (only 11.1% of clicks correctly classified as Risso's dolphin).
File 8 = Short-finned pilot whale	Incorrectly identified as a mix of Mesoplodon forage, Mesoplodon buzz & Risso's dolphin (only 16% correctly classified as Pilot whale).
File 9 = Sperm Whale	Correctly classified as Sperm whale (some confusion with Pilot whale probable caused by 12 KHz HPF)

Table 4: Results from the 5 class CS-SVM for the Workshop Test Data. The contents of the Test Files were not known prior to the Workshop.

6. REFERENCES

1. Johnson, M. and Tyack, P.L. (2003), "A digital acoustic recording tag for measuring the response of wild marine mammals to sound", *IEEE Journal of Ocean Engineering*, vol. 28, pp 3-12.
2. Johnson, M., Madsen, P.T., et al., (2004) , "Beaked whales echolocate on prey", *Proceeding of the Royal Society London, B (Suppl.)*, Vol. 271, S383-386.
3. Moretti, et al, (2005) "Algorithms for passive acoustic detection and localization of marine mammals, including Blainville beaked whales (*Mesoplodon densirostris*)", 2nd Int'l Conf. on Detection and Localization of Marine Mammals using Passive Acoustics, 16-18 November, Monaco.
4. Baggenstoss, P. (2003), "The PDF projection theorem and the class-specific method", *IEEE Trans. on Signal Processing*, vol. 51(3), pp 685.
5. Vapnik, V. N. (1998), *Statistical Learning Theory*, New York, John Wiley & Sons, Inc.
6. Scholkopf, B., Smola, A. (2002), *Learning with Kernels*, Cambridge, MA, MIT Press.
7. Li, Z., Tang, S., and Yan, S., (2002), "Multi-Class SVM Classifier Based on Pairwise Coupling", *Proceedings of the 1st Int'l Workshop, SVM 2002, Niagara Falls, Canada*, pp. 321.
8. Weston, J. and Watson, C., (1998), "Multi-class Support Vector Machines", *Tech. Rep. CSD-TR-98-04*, Royal Holloway, Univ. of London, Dept. of Computer Science.
9. Bailey, A. (2001) *Class-dependent features and multicategory classification*, PhD Thesis, University of Southampton.
10. Moretti, et al, (2005), "Passive acoustic detection and tracking of beaked whales in the Tongue of the Ocean", *Int'l Conf. on Effects of Sound in the Ocean on Marine Mammals*, 2-5 May, Lerici, Italy.
11. Jarvis, et al, (2006), "Automated Classification of Beaked Whales and Other Small Odontocetes in the Tongue of the Ocean, Bahamas", *Proc. of MTS/IEEE Oceans 2006*, 18-21 Sep., Boston, MA.
12. 3rd Int'l Workshop on Detection and Classifications of Marine Mammals Using Passive Acoustics, July 24-26, 2007, Boston, MA, <http://www.rightwhaleweb.org/workshop/>.
13. Johnson, M. et al. (2006) "Foraging Blainville's beaked whales (*Mesoplodon densirostris*) produce distinct click types matched to different phases of echolocation", *Jour. of Exp. Biology*, vol. 209.

Accuracy & Low Cost— Scantek Delivers Sound & Vibration Instruments

Scantek offers two integrating sound level meters and real-time octave-band analyzers from CESVA that make measurements quickly and conveniently. The easy to use SC-30 and SC-160 offer a single dynamic range of 100dB, eliminating any need for range adjustments. They simultaneously measure all the functions with frequency weightings A, C and Z. Other features include a large back-lit screen for graphical and numerical representation and a large internal memory.

The SC-30 is a Type 1 precision analyzer while the SC-160 Type 2 analyzer offers the added advantages of lower cost and NC analysis for real-time measurement of equipment and room noise. Prices starting under \$2,000, including software.

Scantek delivers more than just equipment. We provide solutions to today's complex noise and vibration problems with unlimited technical support by acoustical engineers that understand the complex measurement industry.

Scantek

Sound and Vibration
Instrumentation & Engineering

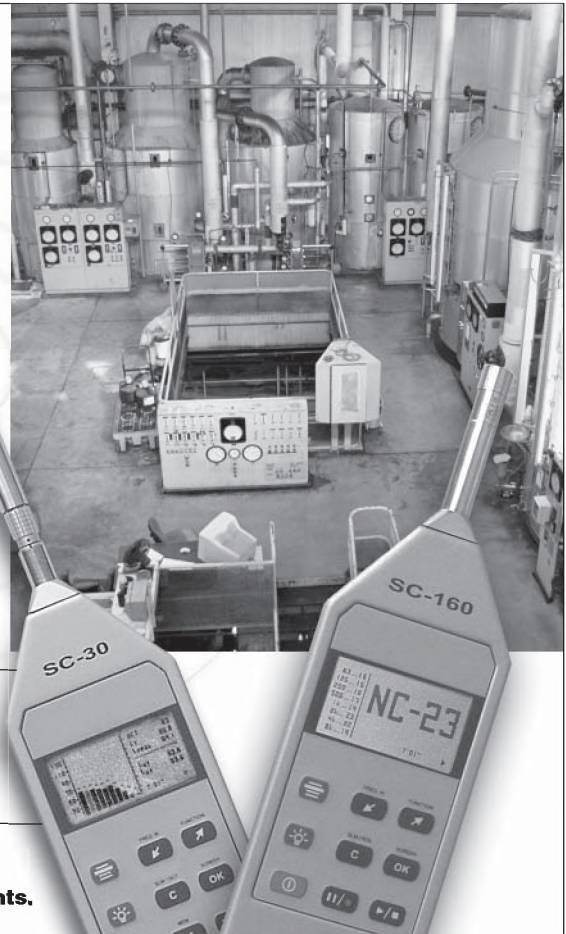
7060 Oakland Mills Road • Suite L
Columbia, MD 21046
800•224•3813
www.scantekinc.com
info@scantekinc.com

SC-30 / SC-160 Applications

- Machinery Noise
- Community Noise
- HVAC Acoustics
- Room Acoustics & Reverb Time
- Noise Criteria (NC) (SC-160)

CESVA

We sell, rent, service, and calibrate sound and vibration instruments.



COMPARISON OF MACHINE LEARNING TECHNIQUES FOR THE CLASSIFICATION OF ECHOLOCATION CLICKS FROM THREE SPECIES OF ODONTOCETES

Marie A. Roch¹, Melissa S. Soldevilla², Rhonda Hoenigman¹, Sean M. Wiggins², and John A. Hildebrand²

¹Dept. of Computer Science, San Diego State University, 5500 Campanile Dr, San Diego, California, 92182-7720 USA

²Scripps Institution of Oceanography, The University of California at San Diego, La Jolla, California 92093-0205 USA

ABSTRACT

A species classifier is presented which decides whether or not short groups of clicks are produced by one or more individuals from the following species: Blainville's beaked whales, short-finned pilot whales, and Risso's dolphins. The system locates individual clicks using the Teager energy operator and then constructs feature vectors for these clicks using cepstral analysis. Two different types of detectors confirm or reject the presence of each species. Gaussian mixture models (GMMs) are used to model time series independent characteristics of the species feature vector distributions. Support vector machines (SVMs) are used to model the boundaries between each species' feature distribution and that of other species. Detection error tradeoff curves for all three species are shown with the following equal error rates: Blainville's beaked whales (GMM 3.32%/SVM 5.54%), pilot whales (GMM 16.18%/SVM 15.00%), and Risso's dolphins (GMM 0.03%/SVM 0.70%).

SOMMAIRE

Ce travail concerne la création d'un système pour identifier trois espèces d'odontocètes par les clics d'écholocation: la baleine à bec de Blainville, la baleine pilote, et le dauphin de Risso. Les clics sont identifiés par l'opérateur d'énergie Teager-Kaiser, et les vecteurs cepstraux sont construits. Dans un travail de détection, on compare les résultats obtenus avec deux modèles différents: le modèle de mélange gaussiens (MMG) et la machine à vecteurs de support (MVS). Les résultats de la détection sont exprimés par les courbes de DET, « *Detection Error Tradeoff* ». Le point sur les courbes de DET où les probabilités de fausses alarmes et manques de détection sont égales est comme suit: la baleine à bec de Blainville (MMG 3,32%/MVS 5,54%), la baleine pilote (MMG 16,18%/MVS 15,00%) et le dauphin de Risso (MMG 0,03%/MVS 0,70%).

1. INTRODUCTION

The use of acoustic information for study of marine mammals is a promising method that is complimentary to visual observations. One use of acoustics is to determine the presence of species of interest, the so called detection problem. In this work, we describe a detection system implemented for the 3rd International Workshop on the Detection and Classification of Marine Mammals Using Passive Acoustics, a conference which brought together multiple groups to work on a common data set containing calls from Blainville's beaked whales (*Mesoplodon densirostris*), short-finned pilot whales (*Globicephala macrorhynchus*) and Risso's dolphins (*Grampus griseus*). Low error-rate detections were achieved for all three species using both Gaussian mixture models (GMMs) and support vector machine algorithms.

2. BACKGROUND

Building an effective machine learning solution is a combination of determining the right set of features to use and an appropriate classifier. Features should be chosen such that they capture the essence of the problem, a

statement that is easy to make and frequently difficult to achieve. Once the feature set is determined, a method of detection or classification must be selected that enables the system to effectively exploit characteristics of the feature set.

2.1 Features

Bioacousticians working on detection and identification problems for odontocetes have traditionally concentrated on extracting features from whistles. Typically, systems identify a variety of measurements of the whistle such as slope, inflection points, frequency, etc. either manually or automatically (e.g. [1, 2]). There has been little effort in the examination of echolocation clicks or burst pulses as providing information that can be used to determine species, and until recently, band limitations of most field recording systems prevented serious consideration of clicks as features for species recognition tasks.

We have noted unique spectral patterns in echolocation clicks of some species of delphinids, notably Pacific white-sided dolphins (*Lagenorhynchus obliquidens*) and Risso's dolphins [3]. Earlier work [4] on an automatic species

identification system showed good results on a species identification problem where whistles, burst-pulses, and clicks were processed in an identical manner. These results have led us to investigate the suitability of clicks as indicators of species. We see this as being a complementary task to whistle-based systems rather than a competing one. Both methods have advantages: whistles propagate farther than clicks [2], but the short duration of clicks makes call separation easier in large population groups, and some species are not known to whistle [5]. In addition, whistle production may be linked to behavioral state and we have observed species which are known to whistle producing only clicks.

A range of techniques have been used to characterize odontocete clicks [6]. In general, signal samples are squared and heuristics or distributional metrics are used to determine the beginning and ending energy. As described later, we use a technique based upon the Teager energy operator which is similar to that proposed by Kandia and Sylianou [7]. Once the click is identified, typical features include the peak frequency, 3 dB bandwidth, inter-click intervals, etc. [8]. These metrics are a very rough approximation of the spectral shape. Most of the work on echolocation has focused on on-axis clicks, but it is well known that off-axis clicks lack the coherence of on-axis ones and have significantly different spectra [9-11]. In addition to inter-species differences, click production is known to vary even in the same individual in source level, peak frequency, and bandwidth, depending upon factors such as activity and environment [8, 10]. The variation in click attributes suggests that an effective species detector needs to be able to learn a variety of click types associated with each given target species.

2.2 Classifiers and detectors

A recent discussion on applications of machine learning techniques to bioacoustics can be found in [4] and includes linear discriminant analysis, neural networks, dynamic time warping, adaptive resonance theory networks, classification and regression trees, hidden Markov models, self-organizing maps, and Gaussian mixture models (GMMs). In this study, we compare the performance of GMMs with that of support vector machines (SVMs). GMMs are well known for their ability to model arbitrary distributions whereas SVMs attempt to model the boundaries between distributions. SVMs have gained in popularity throughout the 1990s in the machine learning community and to our knowledge have only recently been considered in the bioacoustics community [12, 13].

3. METHODS

3.1 Click production of target species

The click characteristics of the three species vary greatly. Digital acoustic recording tag (DTAG) recordings of free-ranging Blainville's beaked whales have shown that they produce two types of click trains [10]. One type has been *Canadian Acoustics / Acoustique canadienne*

observed in prey approach, characterized by a frequency modulated (FM) sweep with inter-click intervals (ICIs) of 100 ms and a median centroid frequency of 38.3 kHz, RMS bandwidth and duration of 6.9 kHz and 271 μ s, respectively. These swept clicks are presumed to be related to foraging activities. As the whales close in on their prey, they have been observed to switch to buzz clicks which have different spectral characteristics from the FM sweep clicks. The buzz clicks have greatly diminished ICIs, a higher median frequency of 51.3 kHz with wider RMS bandwidth (14.6 kHz) and an RMS duration which is about half of the FM sweep clicks (29 μ s).

Analysis of clicks recorded on a ship-deployed hydrophone array [9] show that free-ranging Risso's dolphins produce clicks with ICIs generally between 40-200 ms with short click trains having ICIs of 20 ms. Centroid frequency of on-axis clicks is 75 kHz (out of band for the conference data set) with an RMS bandwidth of 25 kHz and duration of 30-50 μ s. Presumed off-axis clicks from a different population of Risso's dolphins have been shown to have a spectral peak and notch structure [3].

Echolocation clicks of short-finned pilot whales recorded in the Gomera and Canary Islands have been reported [14] to produce clicks with RMS bandwidths of 27 kHz and durations of 8.4 μ s. The mean centroid frequency was 68 kHz (also out of band for the conference data).

3.2 Click detection and feature extraction

Clicks are detected using a two-stage search. In the first stage, spectra are created for 20 ms frames with a 10 ms frame advance that have been windowed using a Hann window. Noise is estimated on a per frequency bin basis over a 5 s average. A frame is said to be a click candidate when frequency bins covering at least 5 kHz exceed the noise floor by 12 dB. After obtaining a set of click candidates, a second pass locates clicks with greater precision in a high pass filtered (10 kHz) signal.

The Teager energy operator [15] is an estimate of the instantaneous energy of a signal and has been shown to be an effective method for detecting echolocation clicks [7]. It is based upon a model of the energy needed to drive a spring-mass oscillator, and measures energy with high resolution:

$$\psi_d(x[n]) = x^2[n] - x[n-1]x[n+1]. \quad (1)$$

A noise floor is set at the 40th percentile of the Teager energy measurements across the interval detected in the previous step. Locations where the Teager energy exceeds the noise floor by a factor of 50 are assumed to be interior to the click and the click onset is found by searching for the point at which the energy dips below 1.5 times the noise floor.

Once the click has been located, cepstral features [16] are computed for a 1200 μ s segment of the signal starting with

the click onset. The log magnitude of the discrete Fourier transform of the segment is computed after windowing with a Hann window. The discrete cosine transform of this result is the cepstrum. We also form an estimate of the cepstral representation of noise in the vicinity of the click and subtract the average noise. This is known as cepstral means subtraction [17] and is a method which normalizes for convolutional noise (e.g. mismatched hydrophones or filtering). Once cepstral features have been generated, they are grouped such that the first click and the last click are separated by no more than 2 s and no click is more than 1 s apart from the previous click.

3.3 Detection

Gaussian mixture models (GMMs) and support vector machines (SVMs) were both used in this study. Due to space constraints, only an outline of each technique is presented, but references to the literature where complete details can be found are provided. For both methods, our experiments are designed to answer the question: Given that we are looking for target species X, was a specific set of clicks produced by this species? This contrasts with an identification task where one attempts to determine which species produced the set of clicks.

Gaussian mixture models

For GMM classifiers, one GMM was trained for each of the three species. GMMs are frequently used to approximate arbitrary distributions as a linear combination of parametric distributions. A set of N normal distributions with separate means μ_i and diagonal covariance matrices Σ_i are scaled by a weight factor c_i such that the sum of their integral across the entire feature space is 1. The likelihood of the cepstral feature vector \vec{x} which represents a click can be computed for model $M = [\{c_i\}, \{\mu_i\}, \{\Sigma_i\}]$ where $1 \leq i \leq N$ by:

$$\Pr(\vec{x} | M) = \sum_{i=1}^N \frac{c_i}{(2\pi)^{\frac{d}{2}} |\Sigma_i|^{\frac{1}{2}}} e^{-\frac{(\vec{x}-\mu_i)^T \Sigma_i^{-1} (\vec{x}-\mu_i)}{2}}. \quad (2)$$

The number of mixtures is typically chosen empirically. Model estimation (training) cannot be accomplished by a straightforward application of the maximum likelihood (ML) principle as the relative contribution c_i of each mixture to the total likelihood is unknown. To address this, the GMM is trained incrementally. A single mixture GMM is estimated from the sample mean and variance. This mixture is then split into two mixtures by dividing the weight in two and forming new mixtures where the means have been slightly perturbed by a small vector $\pm \vec{\delta}$. The resulting model is then refined by an application of the EM algorithm [18] where the current estimate is used to determine the expected values of the mixture weights. With the missing weights estimated, the ML estimator can be

found. This process is executed several times and the model is split again. Once the desired number of mixtures is reached, iteration is performed until a convergence threshold is reached. Convergence is guaranteed and is typically fast (5-15 iterations).

After the models have been trained, the likelihood of click groups are computed and a log likelihood ratio test is used to decide whether each group belongs to each species [19]. We make the simplifying assumption that clicks in a group are independent and compute the group likelihood as the product of the individual click likelihoods normalized for group duration by using the geometric mean. These operations are done in the log domain to prevent machine underflow. Decisions to accept or reject the hypothesis that a click group was produced by the species in question are based upon a log likelihood ratio test. Due to the small number of competing classes, we set the alternative class likelihood to be the likelihood of the highest competitor model as opposed to a background model. The system is implemented using Cambridge University's hidden Markov toolkit (HTK) [20] along with a custom set of programs written in Python and Matlab™.

Support Vector Machines

Support vector machines do not model the distribution of classes, but rather their separation [21]. SVMs find the separating hyperplane that minimizes the risk of a classifier under a 0-1 loss rule. Let $f_{\vec{\theta}}(\cdot)$ be a function parameterized by θ that maps examples to negative and positive class labels $y \in \{-1, 1\}$. As we almost never have access to the actual risk, we can attempt to minimize the empirical risk:

$$R_{emp}(\vec{\theta}) = \frac{1}{N} \sum_{i=1}^N \frac{1}{2} |y_i - f_{\vec{\theta}}(\vec{x}_i)|. \quad (3)$$

Thus, optimizing the parameter vector $\vec{\theta}$ is likely to result in lowering the misclassification rate. For a given family of classifiers, it can be shown that there exists an upper bound on the actual risk with any desired level of certainty [21, 22]. For SVMs, each $f_{\vec{\theta}=[\vec{w}, b]}(\cdot)$ specifies a hyperplane

$\vec{w}\vec{x} + b = 0$ which separates the two classes of linearly separable training data (nonseparable data is discussed later). The hyperplane normal vector \vec{w} and bias b are scaled such that $\vec{w}\vec{x} + b = \pm 1$ holds for the closest positive and negative training example, resulting in an empirical risk of 0. The separating line for a two dimensional synthetic data set and the parallel lines that occur at $\vec{w}\vec{x} + b = \pm 1$ are shown in **Figure 1**. As points on the hyperplane satisfy $\vec{w}\vec{x} + b = 0$, the distance between the closest point of each class and the hyperplane is $\frac{1}{\|\vec{w}\|}$.

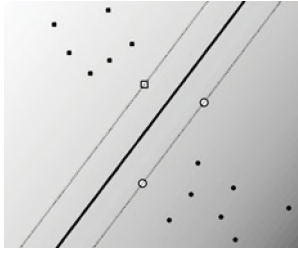


Figure 1 - Separating hyperplane (solid line) between squares and circles that maximizes the distance between the closest vectors (margin). Support vectors lying on $\vec{w}\vec{x} + b = \pm 1$ are outlined.

Consequently, the separation between the two closest points and the hyperplane is $\frac{2}{\|\vec{w}\|}$. This quantity is referred to as the margin and we can learn the appropriate parameters for the SVM by maximizing the margin subject to the constraints of the closest vectors. This is done by minimizing $\|\vec{w}\|$ or equivalently $\|\vec{w}\|^2$ subject to constraints:

$$|\vec{w}\vec{x}_i + b| \geq 1. \quad (4)$$

This is a constrained convex optimization problem, which can be solved by optimizing the dual of the Lagrange multiplier representation [21]. The Lagrange multipliers $\alpha_{1 \leq i \leq N}$ will only be nonzero for training examples which satisfy equality in (4). These vectors are called support vectors. The SVM normal vector \vec{w} can be constructed from the dual solution: $\vec{w} = \sum_i \alpha_i y_i \vec{x}_i$, and b is a more complicated function of the support vectors which we omit. We decide the class of test vector \vec{t} by examining the sign of $\vec{w}\vec{t} + b$, or equivalently in the dual representation:

$$f_{\vec{w}}(\vec{t}) = \begin{cases} 1 & \sum_i \alpha_i y_i \vec{x}_i \vec{t} + b \geq 0 \\ -1 & \sum_i \alpha_i y_i \vec{x}_i \vec{t} + b < 0 \end{cases} \quad (5)$$

The above discussion is for sets that are linearly separable, and can be extended in two ways. The first is to introduce a slack variable $\xi_i \geq 0$ for each training vector which permits support vectors to be on the wrong side of the hyperplane:

$$\begin{aligned} \vec{w}\vec{x}_i + b &\geq 1 - \xi_i & y_i &= 1 \\ \vec{w}\vec{x}_i + b &\leq -1 + \xi_i & y_i &= -1. \end{aligned} \quad (6)$$

When minimizing the risk, a cost factor C is introduced which scales the sum of the slack variables, with high values of C resulting in higher penalties for crossing the margin. Like the linearly separable case, this can also be solved as a constrained optimization problem. The complexity of solving these problems results in selecting strategies such as the sequential minimal optimization algorithm [22] to provide solutions within a reasonable time frame.

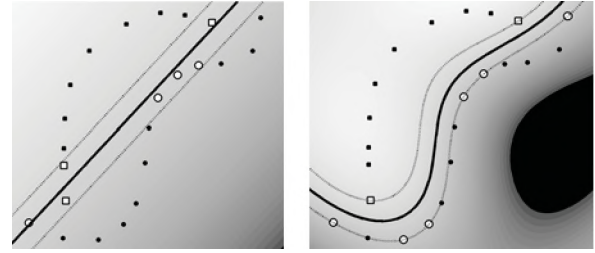


Figure 2 – Squares and circles that are not linearly separable. Hyperplane with dot product kernel (left) versus Gaussian kernel (right).

Typically, the normal vector \vec{w} is not actually constructed, but left as a linear combination of the Lagrange multipliers α_i and their associated training data \vec{x}_i and class y_i : $\vec{w}\vec{t} = \sum_i \alpha_i y_i \vec{x}_i \vec{t}$. A second key element to address nonlinearly separable data is to use a kernel function $K(\cdot, \cdot)$ to transform the data into a different space where linear separation is possible. The examples that we have seen so far use what is known as the dot product kernel $K(\vec{x}, \vec{t}) = \vec{x} \cdot \vec{t}$. While numerous kernels have been proposed [22], we will restrict ourselves to nonlinear Gaussian kernels

$$K(\vec{x}, \vec{t}) = e^{-\frac{\|\vec{x} - \vec{t}\|^2}{2\sigma^2}} \quad (7)$$

where σ is a tunable parameter. Figure 2 shows an example of separating hyperplanes for nonlinearly separable data.

When multiple test vectors are classified as a group, the decision to accept a hypothesis that the clicks are produced by a specific species is based upon the threshold of a statistic of the group's click scores. We use as our statistic the percentage of clicks for which $f_{\vec{w}}(\cdot) \geq 0$. The system is implemented using the Torch machine learning library [23] and custom C++, Matlab™, and Python code.

For both types of classifiers, we used all available training data for the final classifier. During development, training data was jackknifed by recording date so that the system could be evaluated with test data separate from the evaluation test reported in the results section.

3.4 Evaluating results

Results are plotted using the detection error tradeoff (DET) curve [24]. DET curves are similar to receiver operator curves (ROC) except that in the former error rate normal deviates are plotted on both axes, whereas in the latter the correct detection and false alarm probabilities are plotted. When the false alarm and missed detection probabilities are normally distributed, the result is a straight line in DET

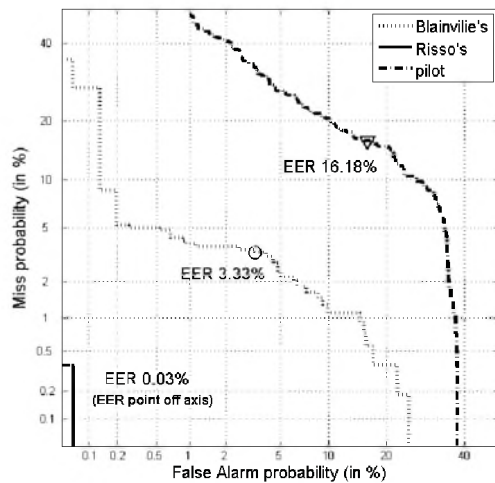


Figure 3 – Detection error tradeoff curves for GMM detector on evaluation data.

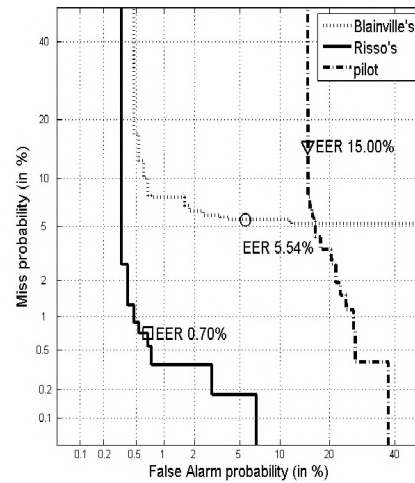


Figure 4 – Detection error tradeoff curves for SVM detector on evaluation data.

space. DET plots are more effective at highlighting differences between similar systems than ROC curves.

4. RESULTS

Mean normalized cepstral features were extracted for all files of the dataset. Tests on the jackknifed training data were used to tune the parameters of each classifier. For the GMMs, 2, 4, 8, 16, 32, and 64 mixture models were created, with 16 mixture models outperforming other parameters. For SVMs, a grid search on the penalty and standard deviation was performed ($C \in \{100, 200, \dots, 600\}$, $\sigma \in \{100, 200, \dots, 1000\}$). Equal error rates (EERs), the point at which a decision threshold results in the same percentage of false alarms (false positives) and missed detections are summarized in **Table 1**. Tests on the last day's training data performed poorly for SVMs, leading to the high overall EERs.

The best performing models from the development data were then used to classify click groups from the nine evaluation files whose content is summarized in **Table 2**. The evaluation dataset contained calls from the three aforementioned species plus an additional two: Atlantic spotted dolphins (*Stenella frontalis*) and sperm whales (*Physeter macrocephalus*).

EER %	GMM	SVM
Blainville's	2.8	21.4
pilot	3.7	21.1
Risso's	2.3	14.7

Table 1– Equal error rates for jackknifed development data with 16 mixture GMMs and $C = 100, \sigma = 200$ SVMs for the best parameter set across all jackknife splits.

File 1 had mixed Blainville's and pilot whale clicks. We manually established "correct" labels for each click group in the file based upon known characteristics of the species and our observations of the calls in the development data. A total of 2040 click groups with a mean of 10.1 clicks per group (min=1, max=103, std dev.=7.2) were classified. DET curves and EERs for all three target species are produced for the GMM and SVM detectors in Figures 3 and 4. The curves show the tradeoff between false alarms and missed detections for various detection thresholds. Note that the thresholds themselves would add a third dimension to the plot and are not reported.

5. DISCUSSION

For both classifiers, the detector performance on Risso's dolphins appears to be nearly perfect in the evaluation data, but the Risso's calls in the conference data were filtered, leading us to suspect that part of the accuracy is due to environment detection as opposed to species detection. It is also worth noting that much of the error on the SVM development set for Risso's dolphins comes from one particular split where the data from August 19th 2006 was used as test data. This was the one day for which the Risso's dolphin data contained clicks with spectra above 40 kHz. The GMM classifier dealt better with this situation, recognizing other similarities in the data. The pilot whale

Species producing calls in the test files					
1	Blainville's + some pilot	4	spotted	7	Risso's
2	Blainville's	5	Risso's	8	pilot
3	spotted	6	Blainville's	9	sperm

Table 2 – Contents of evaluation files 1–9.

detectors had the worst performance on the evaluation data, with the majority of errors being in the 661 out-of-set (species not seen in training) click groups from the spotted dolphins and sperm whales. Using the EER threshold, 42.97% (GMM) and 39.79% (SVM) of the out-of-set click groups were incorrectly identified as pilot whales, indicating that rejection of out-of-set clicks is an area for future work.

For any out-of-set test, the impostor click will most closely fit one of the three distributions, making its GMM likelihood higher than the others. The likelihood ratio between the two highest ranked models may be large, and it is not unexpected that a greater number of errors will occur in this situation. When examining the likelihoods produced by the pilot whale model without the normalizing alternative hypothesis, there is significant overlap. Consequently, setting a threshold based upon the pilot whale model alone would not have improved the results. Adding enough species to the alternative hypothesis to better represent the variability of clicks across species may improve out-of-set rejection. For SVMs, the lack of a distributional approach means that even if a click is far from the target species' distribution, if it lies on the target side of the hyperplane, it will be considered a target, making the need for additional data critical.

It is worth noting that the DET curve for Blainville's beaked whales has a relatively flat slope over much of its length for both detectors. This means that the threshold is not overly sensitive, and we can reduce either the miss or false alarm probabilities significantly with a low impact on the other metric. As an example with GMMs, it is possible to have a very low false detection rate (< 0.2%) and miss no more than 5% of the click groups. While the Risso's dolphin curve has a steep slope, its location in the lower left corner makes this less critical. The shape of the pilot whale curves is more problematic, with small differences in threshold having more significant impact.

When examining what appeared to be off-axis clicks, Johnson et al. [10] were able to distinguish individual pulses by cross correlation with on-axis clicks. They noted that the spectra of the off-axis clicks were "highly featured," lacking the smoothness of presumed on-axis clicks. The spectral irregularities were attributed to possible interference between pulses. We believe this to be a reasonable hypothesis, and one of the major reasons that echolocation based species detection works well. Measurements of the melon taken from CT scans of a deceased Risso's dolphin show a 30 cm length from dorsal bursae to probable signal exit and a 20 cm width at the widest section. While exact propagation paths are beyond the scope of this work, the 1200 μ s window used in this study is adequately long to permit multiple paths to have interfered in constructive and/or destructive manners (assumed sound speed of 1500 m/s), even for the larger species. It is interesting to note that when we used windows smaller than 1100 μ s, detection performance degraded significantly.

6. Conclusions

We have shown that cepstral feature vectors extracted from spectra over a 1200 μ s window starting at the beginning of an echolocation click can be used as the basis for automated species detectors. These detectors are competitive with other state-of-the-art systems for the detection of echolocating marine mammals. It is of particular interest that the system performed well even though the echolocation clicks extended beyond the bandwidth supported by the recording equipment. EERs for this dataset ranged between 0.03% and 16.8% for GMMs and 0.70% and 15.0% for SVMs. Further work is needed on rejecting out-of-set species whose clicks bear a stronger resemblance to the target species than to any of the species used to build the impostor set.

While other explanations may exist, we also believe that the observed degradation of performance when the analysis window was shortened is a strong indicator that interference patterns may play a role in the spectral patterns. Further experiments may help to confirm or reject this hypothesis.

ACKNOWLEDGEMENTS

The authors would like to thank Heather Pettis for her tireless work in organizing DCMMPA 2007, Dave Moretti and the NUWC M3R program for providing the data set, part of which was collected with the participation of Cascadia Research. We are especially appreciative of Ted Cranford and Megan McKenna for sharing the CT measurements with us and helpful discussions about click production as well as Simone Baumann and the anonymous reviewers for their thoughtful comments on the manuscript. This work was supported by US Navy Chief of Naval Operations N45 – Frank Stone and Ernie Young and HPWREN/NSF grant 0426879. SVM plots were produced with Steve Gunn's SVM Toolbox, and DET plots with NIST DETware.

REFERENCES

- [1] L. E. Rendell, J. N. Matthews, A. Gill, J. C. D. Gordon, and D. W. Macdonald, "Quantitative analysis of tonal calls from five odontocete species, examining interspecific and intraspecific variation," *J. Zool., Lond.*, vol. 249, pp. 403-410, 1999.
- [2] J. N. Oswald, S. Rankin, J. Barlow, and M. O. Lammers, "A tool for real-time acoustic species identification of delphinid whistles," *J. Acous. Soc. Am.*, vol. 122, no. 1, pp. 587-595, 2007.
- [3] M. S. Soldevilla, M. A. Roch, S. M. Wiggins, J. Calambokidis, and J. A. Hildebrand, "The use of delphinid echolocation click spectral properties for species classification," *J. Acous. Soc. Am.*, submitted.
- [4] M. A. Roch, M. S. Soldevilla, J. C. Burtenshaw, E. E. Henderson, and J. A. Hildebrand, "Gaussian mixture model classification of odontocetes in the

- Southern California Bight and The Gulf of California," *J. Acous. Soc. Am.*, vol. 121, no. 3, pp. 1737-1748, 2007.
- [5] T. Morisaka, and R. C. Connor, "Predation by killer whales (*Orcinus orca*) and the evolution of whistle loss and narrow-band high frequency clicks in odontocetes," *J. Evol. Biol.*, vol. 20, no. 4, pp. 1439-1458, 2007.
- [6] W. R. Elsberry, "Interrelationships between intranarial pressure and biosonar clicks in bottlenose dolphins (*Tursiops truncatus*)," Wildlife and Fisheries Sciences, Texas A&M, College Station, TX, 2003.
- [7] V. Kandia, and Y. Stylianou, "Detection of sperm whale clicks based on the Teager-Kaiser energy operator," *Appl. Acous.*, vol. 67, no. 11-12, pp. 1144-1163, 2006.
- [8] W. W. L. Au, *The sonar of dolphins*, New York: Springer-Verlag, 1993.
- [9] P. T. Madsen, I. Kerr, and R. Payne, "Echolocation clicks of two free-ranging, oceanic delphinids with different food preferences: false killer whales (*Pseudorca crassidens*) and Risso's dolphins (*Grampus griseus*)," *J. Exp. Biol.*, vol. 207, no. 11, pp. 1811-1823, 2004.
- [10] M. Johnson, P. T. Madsen, W. M. X. Zimmer, N. A. de Soto, and P. L. Tyack, "Foraging Blainville's beaked whales (*Mesoplodon densirostris*) produce distinct click types matched to different phases of echolocation," *J. Exp. Biol.*, vol. 209, no. 24, pp. 5038-5050, 2006.
- [11] W. M. X. Zimmer, P. T. Madsen, V. Teloni, M. P. Johnson, and P. L. Tyack, "Off-axis effects on the multipulse structure of sperm whale usual clicks with implications for sound production," *J. Acous. Soc. Am.*, vol. 118, no. 5, pp. 3337-3345, 2005.
- [12] S. Fagerlund, "Bird species recognition using support vector machines," *EURASIP J. Adv. Sig. Proc. (online journal)*, doi:10.1155/2007/38637, [Accessed 2007].
- [13] S. M. Jarvis, N. A. DiMarzio, R. P. Morrissey, and D. J. Moretti., "A novel multi-class support vector machine classifier for automated classification of beaked whales and other small odontocetes," *Canadian Acoust.*, submitted.
- [14] T. Götz, A. Boonman, U. Verfuss, and H.-U. Schnitzler, *Echolocation Behaviour of Five Sympatric Dolphin Species off la Gomera/Canary Islands [abstract A-14]*, 2005. [<http://www.univ-lr.fr/labo/lbem/ecs2005/Abstract%20book.pdf>, Accessed December 2007].
- [15] T. F. Quatieri, *Discrete-time speech processing: principles and practice*, Upper Saddle River, NJ: Prentice Hall PTR, 2002.
- [16] J. W. Picone, "Signal modeling techniques in speech recognition," *Proc. IEEE*, vol. 81, no. 9, pp. 1215-1247, 1993.
- [17] H. Hermansky, "Exploring temporal domain for robustness in speech recognition," in *The 15th Intl. Congress on Acoustics*, Trondheim, Norway, 1995. pp. 61-64.
- [18] X. Huang, A. Acero, and H. W. Hon, *Spoken language processing*, Upper Saddle River, NJ: Prentice Hall PTR, 2001.
- [19] F. Bimbot, J. F. Bonastre, F. C. G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-Garcia, D. Petrovska-Delacretaz, and D. A. Reynolds, "A tutorial on text-independent speaker verification," *EURASIP J. Appl. Sig. Proc.*, vol. 2004, no. 4, pp. 430-451, 2004.
- [20] S. Young, G. Evermann, T. Hain, D. Kershaw, X. A. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland, *The HTK book, version 3.4*, 2006. [<http://htk.eng.cam.ac.uk>, Accessed March, 2007].
- [21] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Min. Knowl. Disc.*, vol. 2, no. 2, pp. 121-167, 1998.
- [22] N. Cristianini, and J. Shawe-Taylor, *Support vector machines and other kernel-based learning methods*, Cambridge, UK: Cambridge University Press, 2000.
- [23] R. Collobert, S. Bengio, and J. Mariéthoz, *Torch: A modular machine learning software library*, IDIAP-RR 02-46, IDIAP, Valais, Switzerland, 2002.
- [24] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The DET curve in assessment of detection task performance," in *Eurospeech*, Rhodes, Greece, 1997. pp. 1895-1898.

DETECTION OF CLICKS BASED ON GROUP DELAY

Varvara Kandia¹ and Yannis Stylianou^{1,2}

1- Institute of Computer Science, FORTH, Crete, Greece

2- Computer Science Dept. University of Crete, Greece

ABSTRACT

In this paper we present a novel approach for the automatic detection of clicks from recordings of beaked whales based on the phase characteristics of minimum phase signals and especially using the group delay function. Group delay is estimated through the and first derivative of the Fourier Transform of a signal. A major advantage of the proposed approach is its robustness against additive noise while it doesn't require the definition of ad-hoc or adaptive thresholds for the detection of clicks. This method works on raw recordings which are usually quite noisy as well as on click enhanced recordings (after band-pass filtering or using operators like the Teager-Kaiser energy operator). Moreover, a click is just detected by searching the positive zero crossings over time of the slope of the phase spectrum. To evaluate the effectiveness of the proposed approach in detecting clicks, a one-minute recording has been manually marked providing a test set of about 320 clicks. Results show that the proposed approach was able to detect 71.37% of the hand labelled clicks within an accuracy of 3 ms.

SOMMAIRE

Dans cet article, nous présentons une nouvelle approche pour la détection automatique de clics sur des enregistrements de baleines à bec exploitant les caractéristiques de signaux à phase minimale notamment via l'utilisation de la fonction de retard de groupe. Le retard de groupe est estimé à partir de la transformée de Fourier d'un signal et de la dérivée de celle-ci. L'approche proposée est robuste vis-à-vis du bruit additif et ne requiert pas la définition ad-hoc ou adaptative de seuils pour la détection de clics. Elle permet de traiter aussi bien des enregistrements bruts fortement bruités que des enregistrements rehaussés (après filtrage passe-bande ou à l'aide d'opérateurs tels que l'opérateur d'énergie de Teager-Kaiser). De plus, un clic est simplement détecté en recherchant un passage par zéro sur la partie croissante de la pente du spectre de phase. Pour évaluer l'efficacité de l'approche proposée à détecter des clics, une minute d'enregistrement a été annotée manuellement, fournissant ainsi un ensemble de test d'environ 320 clics. Les résultats montrent que l'approche proposée parvient à détecter 71.37% des clics marqués manuellement avec une précision de 3 ms.

1. INTRODUCTION

Beaked whales are deep-diving toothed whales and are the least known family of all marine mammals [1]. Two genera of beaked whales, *Ziphius* and *Mesoplodon*, are not as well known as other genera of beaked whales such as *Berardius*. Acoustic monitoring of the sound activity of these animals may help to study their habitats, which is of considerable conservation value since these whales are very difficult to sight. Moreover, there has been a growing concern about the sensitivity of these animals to human-made sounds [2]. Acoustic analysis of the sounds they produce may help in understanding this sensitivity.

Although some whales (i.e., sperm whales) produce sound pulses in the range of human hearing, which is below 20kHz, beaked whales emit short directional ultrasonic clicks (with significant energy above 20 kHz). An analytic report on recordings using acoustic recording tags attached on *Ziphius* and *Mesoplodon* beaked whales may be found in [1] and [3]. Since clicks produced by beaked whales (as well as from other toothed whales) are highly directional, there is a difference in the properties of the signals if they are recorded off or on the acoustic axis of the whale [1]. In the case where hydrophones are used

for the recordings, the intensity of the clicks will vary a lot over short periods of time. This makes the detection of clicks harder using energy-based criteria. It modifies the frequency content as well, which complicates the detection problem for the frequency or time-frequency based detectors [4]. Moreover, recordings are usually very noisy which makes the detection task even more difficult. Band-pass filtering is widely used for the enhancement of clicks. However, since the frequency properties of clicks change over time, a time-invariant band pass filter may cancel some of the clicks. More appropriate methods for click enhancement have been suggested in the literature, like in the Rainbow Click detector [5] and the Teager-Kaiser energy operator [6], [7]. To overcome the variability in energy levels of clicks the above approaches need to define adaptive energy thresholds, increasing the complexity of the detection system without significantly improving the detection score.

Therefore, new techniques for the automatic detection and classification of clicks generated from beaked whales are urgently needed to study their behavior and habitat use, and to identify risk factors for exposure of these animals to noise [1]. In this paper time-domain and frequency domain techniques for the automatic detection of the high-

frequency clicks of beaked whales will be considered. Although time domain techniques are widely used for detecting clicks, they are not robust in low Signal to Noise Ratio (SNR) conditions as was mentioned earlier. Frequency domain techniques are mainly based on the cross-correlation function defined on the magnitude spectrum of the sounds, ignoring therefore any information provided by the phase spectrum. In this paper we focus on clicks produced by Blainville's beaked whales (*Mesoplodon densirostris*) and we suggest a click detector that combines a time domain technique with frequency domain information based on the slope of the phase spectrum. Specifically, we suggest the use of the Teager-Kaiser energy operator [6] as a click enhancement tool followed by a high resolution group delay function obtained from short-time phase spectra. The group delay function has found important application in numerous signal processing areas, such as speech processing [8]. Clicks are then easily detected by locating the zero crossings of the slope of the phase spectrum (referred to as *phase slope function*) computed as the average of the group delay function. This makes the proposed detector insensitive to variations in sound source level. The proposed detection algorithm performs a frame-by-frame (short-term) analysis. In each analysis window, the slope of the phase spectrum corresponding to the center of the analysis window is computed as the average of the group delay function. Frame (step) size defines the resolution capability of the proposed approach. The algorithm has been tested both on raw recordings of beaked whales and after the application of click-enhancement tools. Results show that the proposed approach is capable of detecting clicks in raw data as well as in pre-processed data (enhanced). For the evaluation of the detector, a one-minute recording was manually labeled providing a set of 317 clicks. Note that in this recording more than one animal is present since the click rate is much higher compared to the mean click rate reported in the literature [1].

The paper is organized as follows. Section 2 describes the time-frequency characteristics of clicks of *Mesoplodon* beaked whales and tools to improve the SNR in recordings (click enhancement). In Section 3 we present different ways to compute the group delay function and the properties of this function for minimum phase signals are discussed. To motivate the use of the group delay for the detection of clicks, the group delay of synthetic signals similar to a sequence of regular clicks is computed and extensively discussed. Details on the application of group delay for click detection using real click recordings are discussed in Section 4. To evaluate the effectiveness of the proposed approach, clicks have been labeled manually. The database which has been used for the evaluation of the proposed detection system is described in Section 5. A summary of the obtained results and future work concludes the paper.

2. CLICKS OF BEAKED WHALES

Beaked whales are difficult to study and they are mostly known from strandings. They are deep-diving animals, they echolocate on prey [1] and they react to human-made sounds. In [1] two genera of beaked whales, *Ziphius* and

Mesoplodon have been tagged making orientation and sound recordings. The tagged whales started clicking at an average depth of 400-500 m. Both species produce regular clicks with an inter-click-interval (ICI) of about 0.4s for *Ziphius* and between 0.2 and 0.4s for *Mesoplodon*. Regular clicks usually terminate with a buzz sound (rapid increase in click rate, 250 clicks per second [1]). The average duration of the clicks were measured at 175 μ s and 250 μ s for *Ziphius* and *Mesoplodon*, respectively. For both species, the energy of their sounds is mostly distributed at high frequencies, i.e., in the 30kHz-40kHz range.

2.1 Time-frequency information

In Fig. 1(a) 13 seconds of a recording from beaked whales (*Mesoplodon densirostris*) are depicted¹. Sounds were digitized at a sample rate of 96 kHz, with 24-bit resolution. From this figure, it is not easy to detect clicks by inspecting the time domain signal. In Fig. 1(b) the time-frequency distribution of the signal is displayed, computed via the Short-Time Fourier transform using a Hanning window of 1000 samples (10.4 ms) with an overlap of 500 samples, and a frequency resolution of 2048 bins. In Fig. 1(b), looking at frequency bands above 20 kHz, some wideband signals may be detected (for instance, around 4, 5, and 8 seconds), indicated the presence of clicks. The high frequency content of the clicks is expected after the results presented in [1]. It is worth noting that by comparing the two figures, it is obvious that the detection of clicks is easier in the time-frequency domain. This is the motivation for using such a representation for the detection of clicks with software products such as Ishmael [4].

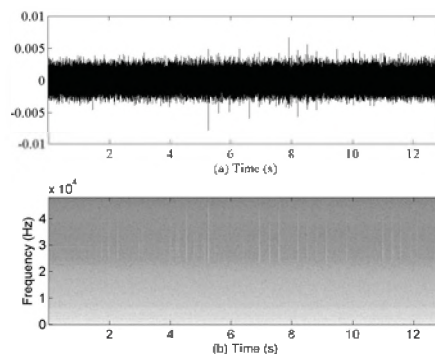


Figure 1. Typical recording from a beaked whale (a) time-domain signal, sampled at 96 kHz, 24 bits, and (b) short-time Fourier transform (2048 frequency bins, Hanning window of 1000 samples, with 50% overlap).

2.2 Click-enhancement

The signal depicted in Fig. 1(a) is very noisy and it is not easy (if not impossible) to detect any click by visual inspection. Therefore, a click enhancement tool could

¹ This recording is part of the recordings made available by the Naval Undersea Warfare Center (NUWC) Marine Mammal Monitoring on Navy Undersea Ranges (M3R) program.

possibly reveal the clicks and improve the SNR. For this purpose, we will use 2 enhancement tools: one is based on the Teager-Kaiser energy operator [6] and the other is based on modulation and downsampling.

Teager-Kaiser energy operator

For a discrete time signal $x[n]$, it is shown in [9] that the Teager-Kaiser (TK) energy operator is given by

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

where n denotes the sample number. An important property of the TK energy operator in (1) is that it is nearly *instantaneous* given that only three contiguous samples are required in the computation of the output at each time instant. More details on the TK operator as applied to click sounds may be found in [6, 7].

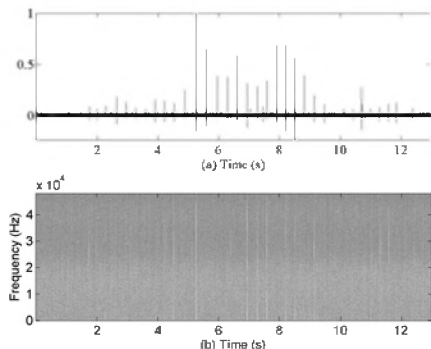


Figure 2. After applying the Teager-Kaiser operator (a) the output from the Teager-Kaiser (TK) operator and (b) the short-time Fourier Transform of the TK output using the same setup as in Fig.1.

Fig. 2(a) shows the results from the application of the TK operator on the original recording, while Fig. 2(b) displays the associated Short-Time Fourier transform. Contrary to Section 2.1, the time-domain signal now carries more information about the time occurrence of clicks than its time-frequency representation. Using an adaptive energy threshold, detection of most of the clicks would be possible. Note that the sampling frequency of the signal has not been changed.

Modulation and down-sampling

Low SNR makes it difficult to detect clicks. Moreover, since most of the energy of clicks is distributed to frequencies higher than 20 kHz, it is also hard to detect them aurally. Therefore, clicks cannot also be detected aurally. Based on the frequency characteristics of the emitted clicks we suggest to modulate the amplitude of the signal and then down-sample it appropriately. Let's denote the original signal by $x[n]$. Then the output signal, $y[n]$ from the above operations is given by:

$$\begin{aligned} v[n] &= x[n] \cos(\pi n) \\ w[n] &= v[\downarrow 2] \\ y[n] &= w[n] \cos(\pi n) \end{aligned}$$

where $\downarrow 2$ denotes the downsampling operation by two after applying an anti-aliasing lowpass filter. The last modulation in the above equations is required in order to re-establish the order of the frequency information. The time domain signal ($y[n]$) and the time-frequency distribution of the signal are depicted in Fig.3. As for the Teager-Kaiser operator, the clicks are clearly seen in the time domain signal. Although some of the narrow band signals are also present in the time-frequency distribution, not all of the clicks are visible. Therefore, again in this case, one would prefer the time domain representation to detect the clicks.

Although, after the application of the above enhancement tools, the "click structure" was revealed in the previous examples, it is still difficult to detect a great number of clicks because of the variability of the click intensities. We recall that the intensity is a function of the position of the whale relative to a hydrophone. Since the whale is moving the intensity changes quickly because of the high directional characteristic of clicks. Taking into account the possibility that other beaked whales are present in the area and may also emit clicks at different distances from the hydrophone, the click intensity can vary quickly over a short period. In this case, a sophisticated time-adaptive system of thresholds should be used for click detection. To overcome this, we suggest to use the slope of the phase spectrum, computed as the average of the group delay function, for click detections. This will make the click detector insensitive to the plethora of different click intensities.

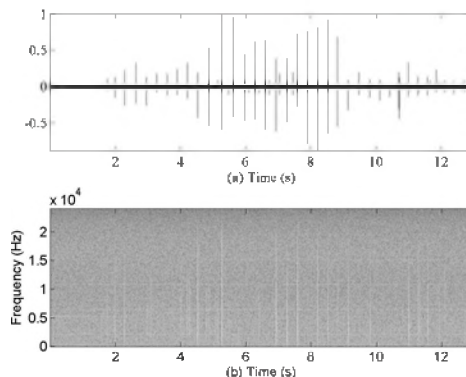


Figure 3. After modulation and down-sampling (a) time-domain signal, sampled at 48 kHz, 16 bits, and (b) short-time Fourier Transform (1024 frequency bins, Hanning window of 500 samples, shifted by 250 samples).

3. GROUP DELAY FUNCTIONS

3.1 Motivation

Consider a delayed unit sample sequence $x[n] = \delta[n - n_0]$ and its Fourier Transform $X(\omega) = e^{-j\omega n_0}$. The group delay is defined as [10]:

$$\tau(\omega) = -\frac{d\phi(\omega)}{d\omega} \quad (2)$$

so the group delay for the delayed unit sample sequence is $\tau(\omega) = n_0 \forall \omega$, since the phase spectrum of the signal is $\phi(\omega) = -\omega n_0$. The average over ω of $\tau(\omega)$ provides n_0 which corresponds to the negative of the slope of the phase spectrum for this specific signal and to the delay of the unit sample sequence. An example of a delayed unit sample sequence with $n_0 = 200$ samples as well as the associated group delay function are depicted in Fig. 4(a) and (b), respectively. In this example the Fourier Transform has been computed considering the center of analysis window to be at $n = 0$. When the window center is moved to the right (closer to the instant $n = n_0$), the slope of the phase spectrum increases (the average of the group delay function decreases) reflecting the distance between the center of the analysis window and the position of the impulse. When the center of the analysis window is at $n = n_0$ then the slope is zero. Continuing moving the analysis window to the right the slope will continue to increase (while the average of the group delay will decrease). In this way, the slope of the phase spectrum is a function of n . Note that the location of the zero-crossing of this function will provide the position of the non-zero value of the unit sample sequence independently of the amplitude value of the impulse. Filtering the unit sample sequence by a minimum phase system, results in a minimum phase signal with the same delay as the input unit sample sequence. The group delay function will still provide information about this delay value as well as information about the poles of the minimum phase system. In Fig. 4(c), (d) the output of the minimum phase signal and the associated group delay are depicted. The slope function will have a similar behaviour to this described earlier for the unit sample sequence.

By creating a periodic sequence of minimum phase signals as the one displayed in Fig. 4(c), a sequence similar to a train of regular clicks may be obtained. Defining an analysis window of length proportional to the period of the sequence (it will be referred to as *long window*), a frame-by-frame analysis is performed. In each frame the slope of the phase spectrum of the windowed signal is computed and it is associated at the center of analysis window. By setting the analysis step size at one sample (moving the analysis window by one sample at a time), the obtained phase slope function (signal) has the same time resolution as the original recording. The window length may have a duration shorter than the period of the signal (it will be referred to as *short window*). In Fig. 5(a) the periodic sequence of the minimum phase signal is displayed along with the phase slope function using long (dashed line) and short (dash-dotted line) window. As it was expected based on the description provided before, the positive zero crossings of the slope function provide the position of the "clicks". Of course, the detection of clicks using a simple energy criterion will provide the same detection score as the proposed approach, in this example. In Fig. 5(b), the same sequence of "clicks" is repeated but now the energy of the minimum phase signals linearly decreases as time increases. In the same figure the phase slope function is also displayed

using, as in Fig. 5(a), the same types of lines for long and short analysis window. It is obvious that a simple energy criterion will not work as well as before and an adaptive energy criterion should be used. Instead, using the slope of the phase function, the position of "clicks" are still easily detected. This is expected since the phase information is not related to the total energy of a signal but rather to the distribution of the signal energy over time.

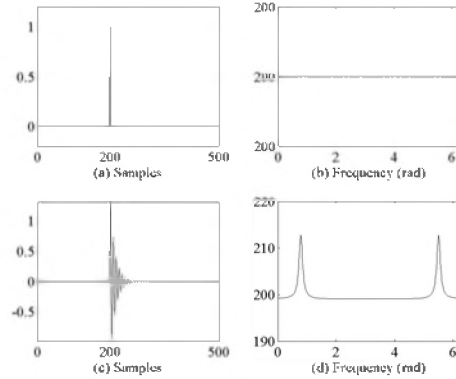


Figure 4. (a) A delayed by 200 samples unit sample sequence. (b) The group delay function of the signal in (a). (c) A minimum phase signal with an oscillation at $\pi/4$. (d) The group delay function of the signal in (c).

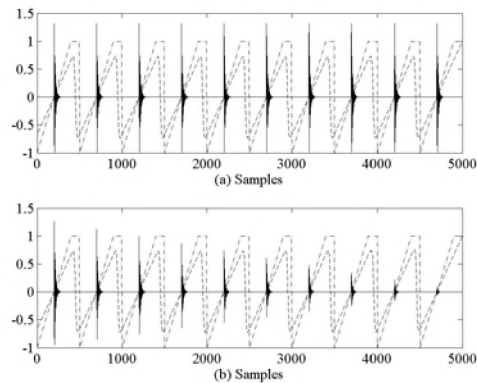


Figure 5. (a) A sequence of impulses of constant amplitude and the associated phase slope function using long (dashed line) and short (dash-dotted line) window (b) A sequence of impulses with linearly time varying amplitudes and the associated phase slope function using long (dashed line) and short (dash-dotted line) window.

3.2 Computing group delay

To compute the group delay of a signal or the average slope of the phase spectrum we need to compute the unwrapped phase spectrum. This is necessary because phase is computed modulo 2π and any attempt to compute the slope using wrapped phase data will produce erroneous results. Usually phase unwrapping is performed by adding appropriate integer multiples of 2π to the principal phase values, as to remove discontinuity (jumps of 2π radians) in the phase curve. Unfortunately, phase unwrapping is not always successful. Therefore, we suggest to compute the slope of the phase function

through an alternative to Eq. (2) computation of the group delay [10]:

$$\tau(\omega) = \frac{X_R(\omega)Y_R(\omega) + X_I(\omega)Y_I(\omega)}{|X(\omega)|^2} \quad (3)$$

where

$$X(\omega) = X_R(\omega) + jX_I(\omega)$$

$$Y(\omega) = Y_R(\omega) + jY_I(\omega)$$

are the Fourier Transforms of $x[n]$ and $nc[n]$, respectively. Using Eq. (3) we avoid the computation of the unwrapped phase. The phase slope is then computed as the negative of the average of the group delay function.

4. DETECTION ALGORITHM

Clicks from beaked whales are highly directional and of very short duration. They can therefore be seen as realizations of impulse responses of minimum phase systems. For the application of the phase slope function to the detection of clicks, we set the length of the analysis window as a function of the average inter-click interval. According to Johnson et al. [1] the average inter-click interval for Blainville's beaked whales is about 0.3s. So for the experiments shown below, we used a hanning window of 0.5s (long window). In this section, the example of the recording shown in Fig.1 will again be considered. Comparing the original recording and its enhanced versions with the ideal train of pulses presented earlier, it is expected that the first step before the computation of the phase slope will be the application of an enhancement tool. In the upper panel of Fig.6 the output from the Teager-Kaiser operator on the original recording is repeated, while in the middle panel the associated phase slope function is depicted. By detecting the positive zero crossings of the phase slope function, the location of clicks is obtained. This is shown in the lower panel of Fig.6, where a unit sample sequence is generated according to the positive zero crossings. It is worth to note the high correlation between the train of clicks and the unit sample sequence. Similar results are obtained if the modulated and downsampled version of the original recording is used. The associated results are shown in Fig.7.

Finally, we have applied the phase slope function on the original recording without the application of any enhancement tool or any other pre-processing step. To our surprise, the structure of clicks is clearly revealed! We believe that this result merits further investigation. Results are depicted in Fig.8. It is worth to note the similarity of results with and without the use of enhancement tools. By comparing closely the detection results in these three cases, we found that there are some differences in detecting the clicks which is more noticeable in the noisy areas of the signal. When the SNR is high² (for instance between 5 and 8 s, in Fig.6), the obtained results are very similar.

² Although SNR is not so meaningful for the original recording, since none of the clicks are easily detected by visual inspection.

To evaluate a click detector, hand labeled data are required. Part of the signal shown in Fig.7 and specifically between the 5th and 8th s, is depicted in Fig.9(a) along with a series of hand labels shown as little triangular marks. It is worth to note the presence of very low intensity clicks in addition to the clicks with relatively high intensity. Labels for the low intensity clicks are also shown. We assume that these clicks are recordings by a conspecific made off the acoustic axis of the whale. To also detect these low intensity clicks using the phase slope function, the window length has to be short enough. For this example the window length was set to 0.1s. In Fig.9(a) the phase slope function is also displayed by a dashed line. It is worth to note the high correlation between the labels and the positive zero-crossings of the phase slope function. If we were to use a longer window, for instance a window of duration 0.5s, then only the high intensity clicks would have been detected. In Fig. 9(b) the detection of clicks is indicated by a sequence of unit samples. For this example, the mean absolute error between the manually labeled and the automatic detected click instances is 1.1ms. Finally, there is deletion of one

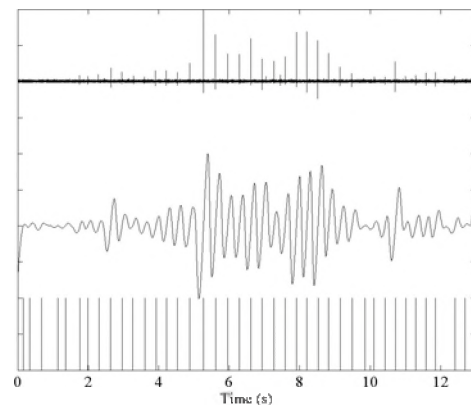


Figure 6. Upper panel: the output from the Teager-Kaiser operator. Middle panel: the associated phase slope function. Lower panel: detection of clicks based on the positive zero crossings of the phase slope function.

click at around 6.7s and one extra click detected at around 7.8s. The mean absolute error in this case is low, however, there are some cases where the detection is not so accurate. The degree of accuracy mainly depends on the SNR. Thus, despite that the structure of clicks can be revealed under very low SNR conditions using the phase slope function, comparisons with hand labeled data shows that using an enhancement tool improves the accuracy. As we have seen, the modulation and the downsampling process improves the SNR of the original recordings. Besides, this is the signal that a human will use to mark the clicks. We suggest to improve further the SNR of that signal by applying the Teager-Kaiser operator on the output after the downsampling. For the short example discussing above, this shows an improvement in accuracy. Indeed, using the TK operator the mean absolute error was decreased to 0.03 ms. However the number of deleted or inserted clicks remained the same as before (one click is deleted and one is inserted). The proposed click detection system is shown in Fig.10.

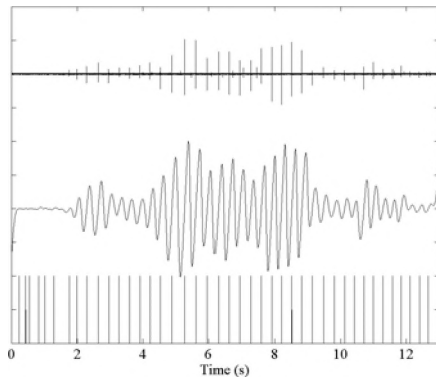


Figure 7. Upper panel: the output from the modulation and down-sampling operations. Middle panel: the associated phase slope function. Lower panel: detection of clicks based on the positive zero crossings of the phase slope function.

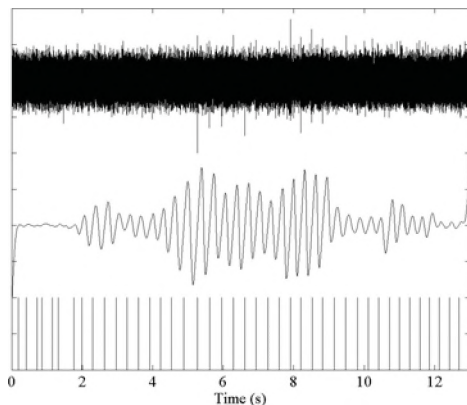


Figure 8. Upper panel: the original recording. Middle panel: the associated phase slope function. Lower panel: detection of clicks based on the positive zero crossings of the phase slope function.

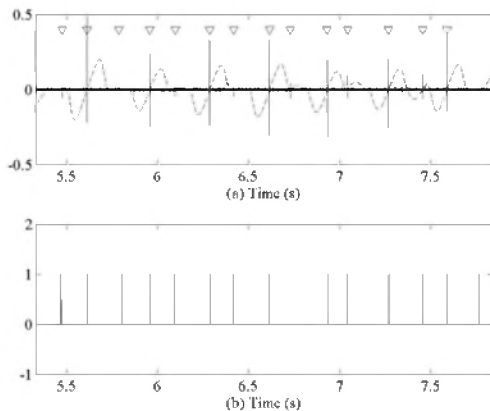


Figure 9. (a) the original recording after modulation and downsampling (solid line), the manual labels (triangle), and the phase slope function (dashed line). (b) Positive zero crossings of the phase slope function indicated by a sequence of unit samples.

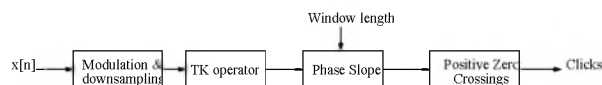


Figure 10. Block diagram of the proposed click detection system.

5. APPLICATION

The proposed click detection system has been evaluated on the training dataset provided by the organizers of the 3rd International Workshop on Detection and Localization of Marine Mammals using Passive Acoustics (Boston, MA, USA, 2007) and were recorded by the Naval Undersea Warfare Center. More specifically, we used recordings of Blainville's beaked whales recorded at a sample rate of 96 kHz, with 24 bits accuracy, from Set1, Alesis number 2, hydrophone H18 at AUTECH (one minute) to manually labeled and from Set3 and 4, Alesis 3 and 6, hydrophones H27 and H76 for visually control the click detection results.

5.1 Database

One minute of recording³ was manually labeled providing in total 317 clicks. These labeled clicks will be referred to as *ds1* dataset. The recording contained only regular clicks. However, the presence of more than one animal was evident in this recording because of the density of clicks and the variability in their intensity. From a visual inspection of the original recording, it was not possible to detect any of these clicks. The initial recording was modulated and downsampled to 48kHz and it was listened through a closed type headphone in a quite office room. Marking was facilitated by using the Sound Forge software. To improve the accuracy of the marking the playback speed of the sound was considerably lowered in some cases (i.e., to about 2kHz). This facilitates the auditory and visual inspection of fast moving acoustic events. It is worth noting that some of the clicks were not easily detectable visually, while only a trained person could hear them. The decision was made to remove some of the clicks, creating therefore a second dataset for the evaluation of the system. This set contained 248 clicks and it will be referred to as *ds2* dataset.

In addition to the manually labeled dataset, the recordings mentioned above from hydrophones H27 and H76, were analyzed and inspected visually. The phase slope function and the automatically detected clicks (positive zero crossings of the phase slope function) were displayed along with the TK output of the modulated downsampled signals. In this way, the correlation between the positive zero crossings of the phase slope function and the clicks could be checked relatively quickly.

5.2 Results

The proposed system was tested on the manually labeled data as well as other recordings contained in the provided training dataset. For all tests, a Hanning analysis window of duration 0.15s was used. To speed up the computation of the slope phase function, a step size of 30 samples was used. Undetermined values of the slope function were computed by linear interpolation. A click was assumed to be detected if the absolute difference between the time-instances of the manually and the automatically detected

³ Filename: Set1-A2-092405-H18-0000-0030-1008-1038loc_0000-0100min.wav

clicks was within 3ms. If this difference was between 3 ms and 20 ms, the click was assumed to be missed (deletion), while a difference over 20 ms indicated a click insertion. For the evaluation of the system two criteria were used; the detection rate (referred to as *Det*) and the corrective rate (referred to as *Corr*). The detection rate is defined as:

$$Det = \frac{\text{Number of clicks correctly detected}}{\text{Total}} \times 100$$

where *Total* is the total number of manually labeled clicks, and the corrective rate is defined as:

$$Corr = \frac{\text{Total} - \text{Deleted} - \text{Inserted}}{\text{Total}} \times 100$$

where *Deleted* referred to as the number of clicks that were considered to be missed (deleted) and *Inserted* refers to the number of extra clicks that have been inserted by the proposed system. It is worth noting that the phase slope function shows occasionally oscillations of very low amplitude about zero which are mostly associated to the noise. On the contrary, for clicks we observe high-amplitude oscillations about zero. Therefore, by subtracting a constant (dc component) from the phase slope function, most of the erroneous clicks associated with noise or very low intensity clicks embedded into noise were eliminated. Such a subtraction was done in order to use the same phase slope function for the two different sets of labels: *ds1* and *ds2*. For *ds1* and *ds2*, the constant was set to 0.1% and 1% of the maximum value of phase slope function, respectively, eliminating 24 erroneous clicks for *ds1* and 7 erroneous clicks for *ds2*.

Table 1 summarizes the detection results for the two datasets, *ds1* and *ds2* (we recall that *ds2* is a subset of *ds1*). As expected, results are better for the second dataset. The number of clicks detected by the system was comparable to the number of manually detected clicks; 321 for *ds1* and 253 for *ds2*. The detection score as well as the corrective rate were mostly affected by the missed clicks. For instance for *ds1*, by increasing the lower threshold (tolerance) to 6ms, the detection score is 85.64% and the corrective rate is 78.12% (compare to 64.03% and 63.72%, respectively). Most of the missed clicks were clicks closely located and for their discrimination a shorter analysis window was needed.

The system was also tested on the training test data mentioned above which contained recordings from Blainville's beaked whales. By visual inspection it has been found that the positive zero-crossings of the phase slope function corresponded to clicks in most cases while some erroneous clicks were introduced from occasional oscillations about zero of the slope phase function as discussed above. Again, by subtracting a constant (dc component) from the phase slope function most of these erroneous clicks were eliminated. Also some clicks were missed because the analysis window was not short enough for their detection (closely spaced clicks).

6. SUMMARY AND CONCLUSIONS

In this paper we present a new technique for detecting clicks from beaked whales based on group delay. More specifically we use the slope of the phase spectrum which is computed as the average value of the group delay function of the input signal. The approach is insensitive to

the intensity of clicks and it is robust against very low SNR conditions.

	Clicks	Corr (%)	Det (%)
ds1	317	63.72	64.03
ds2	248	71.37	72.98

Table 1. Detection results using a tolerance of 3ms.

Combined with the Teager-Kaiser operator, a click detection system was developed and evaluated using recordings from Blainville's beaked whales (*Mesoplodon densirostris*). The proposed system was tested on recordings with manually labeled clicks as well as on unlabeled recordings. Results show the effectiveness of the proposed system in detecting clicks. Mainly, the only design parameter of the system is the length of the analysis window and dc offset. Window length controls the details of the detection. We plan to use the proposed system on clicks from sperm whales and compare its performance with frequency domain and energy based click detectors. Alternative ways in computing the slope of the phase spectrum will be considered.

7. ACKNOWLEDGMENTS

The authors would like to thank the Naval Undersea Warfare Center (NUWC) for making the data used in this paper publicly available, and the organizers of the 3rd International Workshop on Detection and Localization of Marine Mammals using Passive Acoustics (Boston, MA, USA, 2007), for providing access to these data.

8. REFERENCES

1. M. Johnson, P.T. Madsen, W. M. X. Zimmer, N. A. de Soto and P. L. Tyack. "Beaked whales echolocate on prey." Proc. Royal Soc. Biology Letters, 271:383-386, 2004.
2. Frantzis. "Does acoustic testing strand animals?" Nature, 329(29), 1998.
3. P. T. Madsen, M. Johnson, N. A. de Soto, W. M. X. Zimmer and P. L. Tyack. "Biosonar performance of foraging beaked whales (*Mesoplodon densirostris*)." The Journal of Experimental Biology, 208:181-194, 2005.
4. D. K. Mellinger. "Ishmael 1.0 Users Guide", NOAA. NOAA/PMEL/OERD, 2115 SE OSU Drive, Newport, OR 97365-5258, 2001. Technical Memorandum OAR PMEL-120.
5. D. Gillespie. "An acoustic survey for sperm whales in the Southern Ocean sanctuary conducted from the R/V Aurora Australis", Rep. Int. Whal. Comm. 47:897-908, 1997.
6. V. Kandia and Y. Stylianou. "Detection of creak clicks of sperm whales in low SNR conditions." In CD Proc. IEEE Oceans, Brest, France, 2005.
7. V. Kandia and Y. Stylianou. "Detection of sperm whale clicks based on the Teager-Kaiser energy operator." Applied Acoustics, 67(11-12):1144-1163, 2006.
8. R. Smits and B. Yegnanarayana. "Determination of instants of significant excitation in speech using group delay function." IEEE Trans. on Speech and Audio Processing, 3(5):325-333, 1995.
9. J. F. Kaiser. "On a simple algorithm to calculate the 'Energy' of a signal." In Proc. IEEE ICASSP, pages 381-384, Albuquerque, NM, USA, 1990.
10. A. V. Oppenheim, R. W. Schaffer and J. R. Buck. "Discrete-Time Signal Processing." Prentice Hall, 1998.

A NEURAL NETWORK FOR CLASSIFYING CLICKS OF BLAINVILLE'S BEAKED WHALES (*MESOPLODON DENSIROSTRIS*)

David K. Mellinger

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

ABSTRACT

Beaked whales are difficult to detect visually, and researchers have thus proposed using acoustic detection and classification. Because of the large data volumes often involved in acoustic detection and classification, automatic methods are often used. Here a neural network classification method is investigated. Using backpropagation, a feedforward neural network with one hidden layer was trained to classify clicks of Blainville's beaked whales and other odontocetes recorded in the Bahamas. Training and testing data consisted of approximately 1600 Blainville's beaked whale clicks and 3100 clicks from other odontocetes. Networks with 2-10 hidden units were trained and tested, with performance curves (ROC curves) calculated at several levels of signal-to-noise ratio. Results for most networks were quite good when compared with previous classification efforts, with less than 3% errors in both the wrong-classification and missed-call categories. Future work includes testing the network on sounds recorded in different noise backgrounds and from other populations of Blainville's beaked whales, and combining it with a detector and evaluating the joint performance.

SOMMAIRE

Mésoplodons sont difficiles à voir et chercheurs ont proposé d'employer la détection et la classification acoustique pour en trouver. Face à la quantité de données produites par détection et classification acoustiques, méthodes automatisées sont souvent utilisées. Ici on présente une méthode de réseau neuronal pour classifier. Un réseau neuronal à rétropropagation non récurrent avec une seule couche cachée a été formé pour classifier des clics des Mésoplodon de Blainville et autres odontocètes enregistrés aux Bahamas. Les données de formation se sont composées d'environ 1600 clics de Mésoplodon de Blainville et 3100 clics d'autres odontocètes. Réseaux avec 2-10 unités cachées ont été formés et examinés par courbes caractéristiques d'opération du récepteur (ROC curves) calculés à plusieurs niveaux du ratio signal/bruit. Résultats pour la plupart des réseaux étaient tout à fait bons en comparaison avec des efforts précédents de classification avec moins de 3% d'erreurs chez les clics incorrectement classifiés ou manqués. Travaux à suivre sont essais du réseau avec les enregistrements venant d'autres niveaux de bruit de fond et d'autres populations de Mésoplodon de Blainville, et en combinaison avec un détecteur, une évaluation d'exécution commune.

1. INTRODUCTION

Beaked and bottlenose whales – members of the family Ziphiidae, including the genera *Ziphius*, *Mesoplodon*, *Berardius*, *Hyperoodon*, and others – are among the most cryptic and least known of all mammal species. They inhabit deep-water regions (MacLeod and Zuur 2005), which are mostly distant from land and thus relatively difficult to study. They spend much of their time submerged, making it difficult to see them (Barlow et al. 2006). Indeed, visual line-transect studies have found narrower effective strip widths and lower detection probabilities on the trackline for beaked whales than for most other cetaceans (Barlow et al. 2006).

Despite being difficult to see, beaked whales have had notable interactions with humans: they have stranded in places and at times associated with anthropogenic sound use [Frantzis 1998; NMFS 2001; Fernández 2005; Aguilar de Soto 2006], and have attracted intense interest from management agencies, conservation organizations, and the public. A basic first step in preventing harm to beaked whales is to detect when they are present in an area of concern. Because of the difficulty of detecting beaked whales visually, acoustic detection and classification methods have been suggested as a tool for aiding in mitigation of the effects of human activities. Beaked whales are known to produce both echolocation clicks (Johnson et al. 2004) and whistles (Dawson et al. 1998). However,

whistles appear to be relatively rare among all species of beaked whales, making clicks a potentially more useful type of sound for an acoustic detection and classification system.

A wide variety of methods have been used for detection and classification of cetacean sounds. A method that has worked well for a number of species is a neural network (Ghosh et al. 1992; Potter et al. 1994; Kundu and Chen 1997; Murray et al. 1998; Deecke et al. 1999; Houser et al. 1999; Mellinger 2004). Neural networks combine a design phase, in which the structure of the network is chosen, a training phase, in which the parameters of the network's units are configured, and a testing and use phase, in which the network is operated with the parameters fixed. Here a method is presented for acoustic classification of clicks of Blainville's beaked whale (*Mesoplodon densirostris*) and other odontocetes using a neural network. Although a simple automated detector is used for finding clicks of other odontocetes to use in training and testing, the focus here is on classification of the clicks.

2. METHODS

2.1 Classification method

An input sound signal is used to compute a spectrogram, to which conditioning steps – spectrum level equalization, rectification, and normalization – are then applied. The conditioned spectrogram is then used as input to the neural network, resulting in a classification value indicating the certainty that a Blainville's beaked whale click is present.

In more detail, the spectrogram is calculated using a frame length of 0.000667 s (64 samples at a sample rate of

96 kHz), overlap of 50%, and a Hann window. This short frame size was chosen because of the brief nature of beaked whale clicks (Johnson and Tyack 2005), and indeed other known odontocete clicks (e.g., Au 1993). At this frame length, time resolution is relatively good, while the spectrogram filter bandwidth is a relatively poor 6.0 kHz. Nevertheless, the upsweeping nature of these clicks can still be resolved in these spectrograms (Fig. 1). The logarithm of each spectrogram cell is used, compressing spectrogram values to a range typically in the range of ± 10 .

After calculation of the spectrogram, the next step is spectrogram level equalization, rectification, and normalization. This is similar to the method described by Mellinger (2004), and will be explained only briefly here. Equalization is performed by subtracting the time-averaged spectrum (Van Trees 1968) from each spectrogram frame; that is, the spectrum for each spectrogram frame is multiplied by a small positive constant α near zero and added to the product of the long-term average spectrum and $1-\alpha$. Rectification consists of hard-limiting the minimum value in the spectrogram with a (constant) floor value to remove small and negative values. Normalization consists of subtracting the floor value from each spectrogram cell, so that the minimum spectrogram value becomes zero. 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, a floor value is applied, and the floor constant is subtracted so that the minimum value in the resulting spectrogram is 0. The time constant used for equalization here was 0.02 s, while the floor value of the (logarithmic) spectrogram was 0.3 (this is equivalent to $e^{0.3} \approx 1.35$ as a raw FFT value).

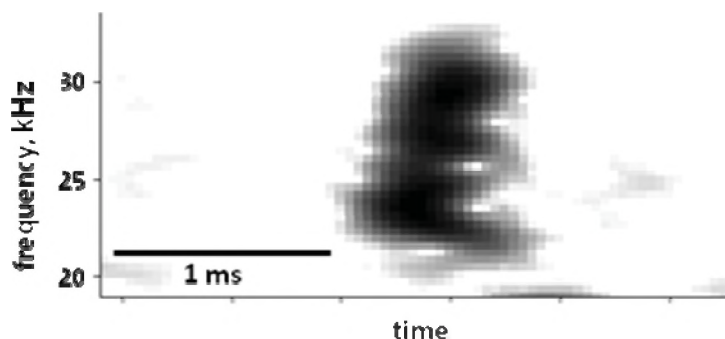


Fig. 1. An example click of a Blainville's beaked whale showing the upsweeping nature of these clicks. Spectrogram parameters: frame size 0.000667 s (64 samples), FFT size 128 samples, hop size 1/16 frame, Hann window, for a filter bandwidth of 6.0 kHz.

A neural network (Hagan et al. 1996) was designed with 192 input elements, a variable number of hidden units, and 1 output unit. The network was strictly feedforward,

i.e., without any backward loops. Each hidden unit consisted of a weighted sum with bias followed by an arc-tangent nonlinearity. The output unit was linear, consisting of just a

weighted sum. The number of hidden units was varied between 2 and 10 to estimate what the optimal number would be. The network was trained using the data set described below; batch training in each epoch was used to remove any bias in order of presentation. The training method was backpropagation (Rumelhart et al. 1987), so that network weights were adjusted according to a back-propagated error function, and a momentum term was used to prevent the network from getting 'stuck' in local maxima.

2.2 Data set

The data set consisted of recordings made at the Atlantic Undersea Test and Evaluation Center (AUTEK) in the Bahamas that contained clicks of Blainville's beaked whales. The whales were visually identified in the field by trained observers; the visual sightings coincided with the acoustically localized positions of the clicks (Moretti et al. 2006). Recordings were made at a sample rate of 96 kHz.

The recordings were manually scanned to detect clicks of Blainville's beaked whales. Manual scanning was used to remove the possibility of bias in detection of clicks; automated methods were not used to detect sounds for use in training and testing, as the methods themselves may introduce bias. The recordings were annotated to indicate the time and frequency bounds of each Blainville's beaked whale click. A total of 1595 Blainville's beaked whale clicks were found, and are henceforth called the BBW clicks.

The AUTEK recordings were also scanned to find clicks, presumably echolocation clicks, of other odontocetes. Known species on these recordings included Risso's dolphins (*Grampus griseus*) and long-finned pilot whales (*Globicephala macrorhynchus*). This scanning was done automatically, using a simple detector that found energy in the 20-38 kHz range of the Blainville's beaked whale clicks (Moretti et al. 2006): a ratio of the long-term to short-term averages was calculated, and when this ratio exceeded a threshold, a click was registered. These clicks were annotated similarly to the beaked whale clicks, with a total of 3096 clicks found. These clicks were named the 'other' clicks.

2.3 Training and testing

Conditioned spectrograms of the annotated clicks, both BBW and 'other', were calculated and used for training and testing the neural network. Only a portion of the spectrogram was used, namely the portion from 15 kHz to 38 kHz, as this frequency band contained most of the energy of beaked whale clicks present in these recordings (Moretti et al. 2006). Also, it was important to exclude frequencies below 14 kHz, as some of the recordings were filtered with a high-pass cutoff at this frequency. Using the entire bandwidth of such recordings would provide an unrealistic cue to the neural network for distinguishing BBW and other clicks. This frequency range contains 16 bands of the spectrogram.

For each click, a conditioned spectrogram centered on the click and lasting 0.004 seconds was used. For BBW clicks, the center was defined as the midpoint of the annotated sound; for 'other' clicks, the center was defined as the peak of the summed energy in the 20-38 kHz range. The 0.004-second spectrogram comprised 12 spectrogram frames, for a total of $12 \times 16 = 192$ cells. It was these cells for each click, arranged into a 192-element vector, that were used as input to the neural network.

The click data set was randomly divided into training and testing data. The BBW data were divided such that 9/10 of the clicks were used for training, with the remaining 1/10 used for testing; the 'other' clicks were divided similarly. The network was trained using the two datasets, with target outputs of +0.5 and -0.5 for the BBW and 'other' clicks, respectively.

2.4 Performance evaluation

Performance was measured using the one-tenth of the BBW and 'other' clicks reserved for testing. Testing was done by calculating the output of the network for the two sets of test data – typical output values were between -1 and 1, though other values occurred too – and applying a set of thresholds. For each threshold, the fraction of wrong classifications (false positives) and missed clicks (false negatives) was calculated; as the threshold was increased, there were fewer false classifications but more missed clicks. Varying the threshold and calculating the fractions of wrong classifications and missed clicks for each threshold yielded a parametric curve, the Receiver Operating Characteristic curve (Fawcett 2006). Training and testing was done five times, and the ROC curve calculated five times, for each number of hidden units in the network, and the five ROCs were averaged to produce the final results.

The signal-to-noise ratio (SNR) of any sound, including a click, is a key parameter in evaluating performance. Nearly all methods work well when the SNR is high, while only some work well at low SNR. Thus it is important to distinguish differing levels of SNR in describing performance of a classification method. Here SNR is measured by calculating the energy ratio of the signal in the 20-38 kHz band in a time period ± 0.01 s around each click; that is, the average band-limited energy of the click is measured and is divided by the average band-limited energy of the background noise in this time period. Separate ROC curves were calculated in 5-dB increments of SNR level, i.e., SNRs of less than 10 dB, 10-15 dB, 15-20 dB, and more than 20 dB.

3. RESULTS

ROC curves for the neural network are shown in Fig. 2. Because of the large range of values, the curve was plotted on a logarithmic scale. The left plot shows the ROC curves for varying numbers of hidden units, with the number of units indicated next to each curve. The right plot shows performance for various SNRs for the best network.

A single-point measure of performance was assessed as well: at the 1% wrong-classification rate, a total of 0.6% of all BBW clicks are missed.

4. DISCUSSION

Some of the better ROC curves for the neural network are entirely less than the 3% error bounds in both dimensions, wrong classifications and missed clicks. This performance is very good compared with previous detection methods, including neural networks, that were applied to baleen whale vocalizations (Mellinger 2004, Mellinger et al. 2004). Part of the reason for the good performance is that the training and testing data were drawn from the same recordings of presumably the same whales, so the signals to be detected were probably very similar between the training and testing data sets. However, this was also true for the data sets in the baleen whale detection studies. Another reason may be that the clicks studied here are more stereotyped than the moans of baleen whales, so that a

network trained to detect clicks in the training data works well on other, adjacent clicks in the testing data. In addition, Blainville's beaked whale clicks do not travel very far (Moretti et al. 2006), and so must have been produced closer to the hydrophone than the baleen whale vocalizations. They would therefore have been affected less by variability in ocean acoustic propagation. However, successive baleen whale vocalizations – some used for training, some for testing – should have been affected by essentially the same propagation conditions, so if they were produced in a highly stereotyped manner, they should have arrived at the hydrophone with very similar structure, and should have been detected equally well. It is also possible that the reason is timing: adjacent beaked whale clicks are closer to each other in time – they are typically less than a second apart – while adjacent baleen whale sounds are tens to hundreds of seconds apart, so that the propagation conditions varied more between adjacent baleen whale vocalizations than they did between adjacent beaked whale clicks.

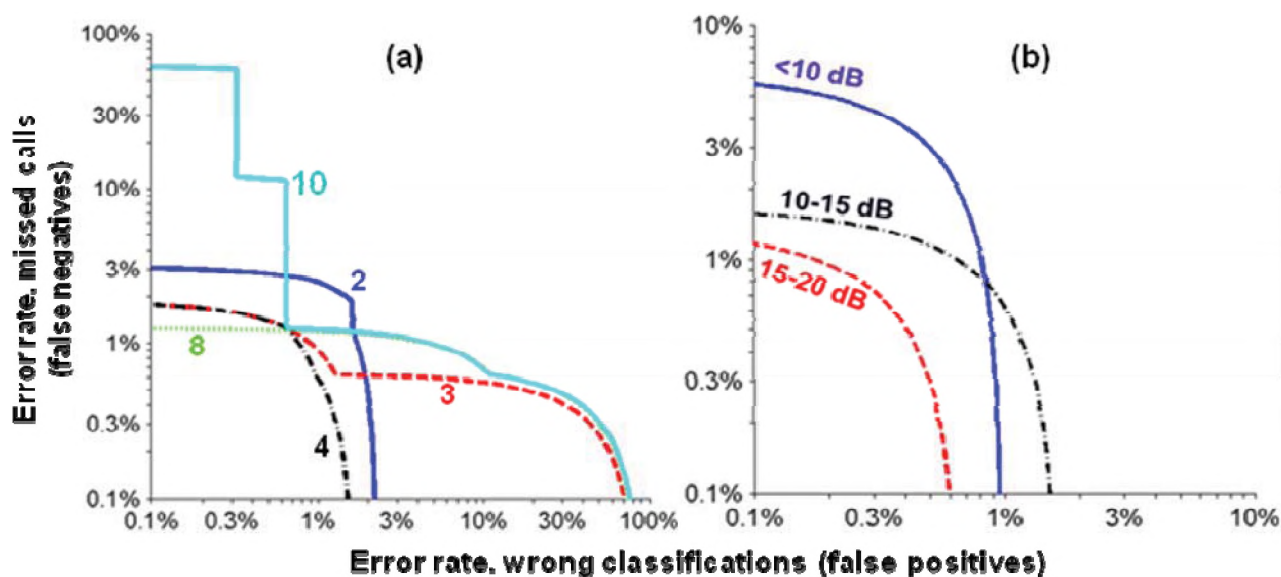


Fig. 2. Receiver Operating Characteristic curves for the neural network detector applied to the training data. Values near the lower left corner, representing smaller numbers of false detections and missed calls, are better. (a) ROC curve for different numbers of hidden units in the neural network. [The 8-hidden-unit curve is hidden by the 10-hidden-unit curve to the right of the 1% false positive point.] (b) ROC curves for the 4-hidden-unit network, with the curve for each 5-dB SNR bin plotted separately. The bin with SNR greater than 20 dB had no missed calls, and so could not be plotted on a logarithmic scale.

It appears that the curve for the 4-hidden-unit network performed best over much of the range, with the 8-hidden-unit network best over the remainder (Fig. 2). However, re-running the training procedure on another 4-hidden-unit network resulted in a performance curve somewhat worse than this one, closer to the 10-hidden-unit curve shown here to the right of the 1% false positive mark. So the superior performance of this network cannot be attributed solely to the number of hidden units.

It is possible that the performance shown here is due to over-fitting. For instance, the network with 4 hidden units

has $192 \times 4 + 4 = 772$ weights, which were trained using a data set of 90% of the whole – i.e., 1436 BBW clicks and 2786 ‘other’ clicks, or 4222 data points in total. This is about 5.5 training clicks per weight, which could be insufficient. The networks with fewer hidden units are less likely to have suffered over-fitting, with e.g. 11 data points per weight for the 2-hidden-unit network, and vice versa – the network with 10 hidden units had only 2.2 data points per weight.

Future work includes testing this network on sounds recorded from Blainville's beaked whale populations elsewhere in the world and in different noise conditions.

One might expect a neural network to perform poorly when confronted with different background noise. However, there is some hope that this one will do well, as the spectrogram conditioning steps reduce the influence of stationary or slowly-varying noise – indeed, of any noise source that is stationary on roughly the time scale at which the spectrum equalization occurs, 0.02 s.

Also, this classifier needs to be combined with a detector and the two evaluated together so that they can be useful for detection of beaked whales in the field, and can be used to mitigate the effects of human activities, including anthropogenic noise, upon these cryptic and little-understood animals.

ACKNOWLEDGEMENTS

Thanks to Dave Moretti, Nancy DiMarzio, Ron Morrissey, and Jessica Ward for providing the data used here. Thanks also to Sara Heimlich and Sharon Nieu Kirk for the many hours spent annotating beaked whale clicks; you hung in there. Thanks to Kate Stafford for the French translation. The work was supported by Navy awards N00014-03-1-0099 from the Office of Naval Research (thanks to Bob Gisiner), and by N00014-03-1-0735, N00244-06-P-1870, and N00244-07-1-0005 from the Office of the Chief of Naval Operations and the Naval Postgraduate School (thanks to Frank Stone, Ernie Young, and Curt Collins). This is PMEL contribution #3147.

REFERENCES

- Aguilar de Soto, N., M. Johnson, P.T. Madsen, P.L. Tyack, A. Bocconcelli, and J.F. Borsani. Does intense ship noise disrupt foraging in deep-diving Cuvier's beaked whales (*Ziphius cavirostris*)? *Mar. Mamm. Sci.* 22:690-699.
- Au, W.W.L. 1993. *The Sonar of Dolphins*. Springer-Verlag: New York.
- Barlow, J., M.C. Ferguson, W.F. Perrin, L. Balance, T. Gerrodette, G. Joyce, C.D. Macleod, K. Mullin, D.L. Palka, and G. Waring. 2006. Abundance and densities of beaked and bottlenose whales (family Ziphiidae). *J. Cetacean Res. Manage.* 7:263-270.
- Dawson, S., J. Barlow, and D. Ljungblad. 1998. Sounds recorded from Baird's beaked whale, *Berardius bairdii*. *Mar. Mamm. Sci.* 14:335-344.
- Fawcett, T. 2006. An introduction to ROC analysis. *Patt. Recogn. Lett.* 27:861-874.
- Fernández, A., J.F. Edwards, F. Rodríguez, A.E. de los Monteros, P. Herráez, P. Castro, J.R. Jaber, V. Martin, and M. Arbelo. 2005. "Gas and Fat Embolic Syndrome" involving a mass stranding of beaked whales (family Ziphiidae) exposed to anthropogenic sonar signals. *Veterinary Pathology* 42:446-457.
- Frantzis, A. 1998. Does acoustic testing strand whales? *Nature* 392:29.
- Hagan, M., H. Demuth, and M. Beale. 1996. *Neural Network Design*. Brooks/Cole: Pacific Grove.
- Johnson, M., P.T. Madsen, W.M.X. Zimmer, N. Aguilar de Soto, and P.L. Tyack. 2004. Beaked whales echolocate on prey. *Proc. Royal Soc. London B Supplement* 6, *Biology Letters*:S383-S386 (DOI 10.1098/rsbl.2004.0208).
- Johnson, M. and P. Tyack. 2005. Measuring the behavior and response to sound of beaked whales using recording tags. National Oceanographic Partnership Program Report: Award Number OCE-0427577.
- MacLeod, C.D., and A.F. Zuur. 2005. Habitat utilization by Blainville's beaked whales off Great Abaco, northern Bahamas, in relation to seabed topography. *Mar. Biol.* 147:1-11.
- NMFS. 2001. Bahamas marine mammal stranding event of 15-16 March 2000. Joint Interim Report, National Mar. Fish. Serv., Washington, DC. 66 pp.
- 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.
- Ghosh, J., L.M. Deuser, and S.D. Beck. 1992. A neural network-based hybrid system for detection, characterization, and classification of short-duration oceanic signals. *IEEE J. Oceanic Engr.* 17:351-363.
- Houser, D.S., D.A. Helweg, and P.W. Moore. 1999. Classification of dolphin echolocation clicks by energy and frequency distributions. *J. Acoust. Soc. Am.* 106:1579-1585.
- Kundu, A., and G.C. Chen. 1997. An integrated hybrid neural network and hidden Markov model classifier for sonar signals. *IEEE Trans. Sig. Process.* 45:2566-2570.
- Mellinger, D.K. 2004. A comparison of methods for detecting right whale calls. *Can. Acoust.* 32:55-65.
- Mellinger, D.K., S. Heimlich, and S. Nieu Kirk. 2004. A comparison of optimized methods for detecting blue whale calls. *J. Acoust. Soc. Am.* 116:2587(A).
- Moretti, D., N. DiMarzio, R. Morrissey, J. Ward, and S. Jarvis. 2006. Estimating the density of Blainville's beaked whale (*Mesoplodon densirostris*) in the Tongue of the Ocean (TOTO) using passive acoustics. *Proc. IEEE Oceans '06*. 5 pp.
- 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.
- 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., J.L. McClelland, and the PDP Research Group. 1987. *Parallel Distributed Processing*. MIT: Cambridge.
- Van Trees, H.L. 1968. *Detection, Estimation, and Modulation Theory*, Vol. I. John Wiley Sons: New York.

PASSIVE ACOUSTIC DETECTION AND LOCALIZATION OF MESOPLODON DENSIROSTRIS (BLAINVILLE'S BEAKED WHALE) VOCALIZATIONS USING DISTRIBUTED BOTTOM-MOUNTED HYDROPHONES IN CONJUNCTION WITH A DIGITAL TAG (DTAG) RECORDING

Jessica Ward¹, Ronald Morrissey¹, David Moretti¹, Nancy DiMarzio¹, Susan Jarvis¹, Mark Johnson², Peter Tyack², and Charles White³

1 - Naval Undersea Warfare Center Division Newport, 1176 Howell St., Newport, R.I., USA

2 - Woods Hole Oceanographic Institute, Woods Hole, MA., USA

3 - University of Rhode Island, Narragansett, R.I., USA

ABSTRACT

Click data from a tagged *Mesoplodon densirostris* was compared with broadband acoustic recordings from an 82 hydrophone wide-baseline array located in the Tongue of the Ocean, Bahamas. Two detectors, a Fast Fourier Transform (FFT) based detector and matched filter, were evaluated in white noise and with the acoustic recordings from the array for performance detecting *M. densirostris* clicks. The matched filter performed the best, allowing 92% of the tagged animal's clicks to be detected on at least one hydrophone. Time Difference of Arrivals (TDOAs) between the DTag and the surrounding hydrophones were computed. These TDOAs were used to compute a three-dimensional hyperbolic localization track of the tagged animal. A maximum detection range of 6500 m from the tagged animal to the recording hydrophone was observed. Offset aspect angles were determined from the DTag heading information and the bearing to the receiving hydrophone. Clicks within ± 30 degrees were detected at the farthest ranges, while clicks were detected at all off-set angles at closer ranges.

SOMMAIRE

Les données de « clics », obtenues pour un *Mesoplodon densirostris* marqué, ont été comparées à des enregistrements acoustiques à large bande obtenus à l'aide d'un vaste réseau de référence à 82 hydrophones situés dans la Langue de l'Océan (Bahamas). Deux détecteurs, soit un détecteur à transformée de Fourier rapide (FFT, de l'anglais Fast Fourier Transform) et un filtre adapté, ont été évalués en bruit blanc et avec les enregistrements acoustiques obtenus grâce au réseau pour détecter les « clics » des *M. densirostris*. Le filtre adapté a réalisé la meilleure performance, ayant permis de détecter 92 % des clics de l'animal marqué au moins avec un des hydrophones. La différence entre les temps d'arrivée entre le DTag et les hydrophones avoisinants ont été calculées. Les différences ainsi calculées ont été utilisées pour effectuer un suivi tridimensionnel et hyperbolique des déplacements de l'animal marqué. Une plage de détection maximale de 6 500 m entre l'animal marqué et l'hydrophone enregistreur a été observée. Les angles correcteurs ont été déterminés à partir de l'information du DTag et le relèvement géographique a été effectué par rapport à l'hydrophone récepteur. Les clics se trouvant à l'intérieur des angles de ± 30 degrés ont été détectés aux distances les plus éloignées, alors que des clics se trouvant à l'intérieur de tous les angles correcteurs ont été détectés à des distances plus courtes.

INTRODUCTION

On October 23, 2006, a Woods Hole Oceanographic Institute (WHOI) DTag was placed on one individual in a group believed to consist of four *Mesoplodon densirostris* in the Atlantic Undersea Test and Evaluation Center (AUTC) underwater tracking range. The tag remained attached for approximately 19 hours over which time seven deep dives were recorded. Whale vocalizations were simultaneously monitored using 82 AUTC bottom-mounted hydrophones. The hydrophones are at depths of ~ 2 m and are separated by ~ 4 km baselines. Two detectors are evaluated for use in *M. densirostris* click detection: a FFT based detector and matched filter. Detection performance is first evaluated in the presence of Gaussian white noise, and

then with hydrophone recordings corresponding to the tagging event.

2. METHODS

2.1 DTag Data Description

The DTag was attached to a probable female *M. densirostris* in a group believed to consist of two mother-juvenile pairs at 11:37:38 a.m. (± 5 seconds) local time on October 23, 2006. The tagging GPS location was $24^{\circ} 30.412' N$, $77^{\circ} 35.320' W$. A male *M. densirostris* may have been in the vicinity. The DTag recorded stereo audio at a 192 kHz sampling rate with an audio sensitivity of -171 dB re

1.0/ μ Pa. The pitch, roll, heading, and depth of the whale were determined from the accelerometer, magnetometer, and pressure sensors sampled at 50 Hz. The DTag measurements were processed using the methods described in Johnson and Tyack [1] resulting in orientation and depth data with 5 Hz resolution.

2.2 AUTECH Hydrophone Array

Prior to and for the duration the tag was attached to the whale, audio data from the 82 bottom mounted hydrophones of the AUTECH tracking range were simultaneously recorded digitally on multiple Alesis HD24 hard drive recorders at a 96 kHz sampling rate. Each recorder can accommodate 12-channels of data with the last channel recording an IRIG-B modulated time signal.

The hydrophones are mounted 4-5 meters off the sea floor with an upward, roughly hemispherical, beam pattern. There are 68 wideband hydrophones with a usable bandwidth from 50 Hz to approximately 45 kHz. There are an additional 14 hydrophones with a bandwidth from roughly 8 kHz to over 50 kHz installed in two 7 hydrophone arrays. Hydrophone data is digitized at a sampling rate of 96 kHz. This is a standard audio rate that allows for Nyquist sampling of the wideband hydrophones. The upper 2 kHz of the 14 wider bandwidth hydrophones is aliased. This folding has not been found to have a significant effect on the determination of click arrival times.

2.3 Detection

M. densirostris produce echolocation clicks with a frequency modulated upswEEP. The peak source level of similarly sized delphinids has been estimated at 220 dB re. 1 μ Pa [2]. Tag data from another species of beaked whale (*Ziphius cavirostris*) indicate a pronounced beam pattern with a 3 dB beam width of 6° [3]. Due to the narrow beam width, determination of the arrival time of a specific click on multiple hydrophones in a widely spaced array such as at AUTECH is a challenge. The hyperbolic localization technique used requires a minimum of three hydrophones for a two dimensional position to be determined. Improved detector performance is critical in order to maximize the probability that a given click will be detected on enough hydrophones to produce a position. Accordingly, two detection methods, a FFT based detector and a matched filter, were compared with respect to detection performance. Detection performance was first compared in Gaussian white noise, and then on recorded data in the vicinity of the tagged animal.

DTag Click Detector

Clicks recorded on the DTag were classified as belonging to the tagged whale based on two features. The attachment of the tag to the whale results in a low-frequency energy component that is not present in clicks from conspecific

whales [3]. Second, the angle of arrival for clicks from the tagged whale is close to zero between the two hydrophones on the tag, while it varies as the whale moves for clicks from conspecifics [4].

FFT based detector

A multi-stage FFT based energy detector has been successfully used for detection of clicks from a variety of echo-locating odontocetes, including sperm whales [5] and beaked whales [6]. A 2048 point FFT with a 50% overlap is used for this analysis. At the 96 kHz sampling rate this provides a frequency resolution (per bin) of 46.875 Hz and a time resolution (per FFT) of 10.67 ms. Each bin of the FFT is independently thresholded against an exponentially decaying time average of the data in that bin as given in Eq. 1:

$$n>0: NVT[n] = (1 - a)bin[n] + aNVT[n - 1] \quad Eq. 1a$$

$$n=0: NVT[n] = 0 \quad Eq. 1b$$

where, the parameter a has been chosen empirically to provide a time constant of 0.2 seconds.

The binary output of the thresholding process is combined into a single detection report. If any of the 1024 bins have passed threshold, the first stage declares a detection and passes the detection report on to the next stage.

The output of the first stage of the detector is then examined to determine whether the event was triggered by a beaked whale. Since clicks are broadband events, the detection report may be broadly classified as a click by counting the number of FFT bins which triggered. Assuming a click event is declared, the frequency content of the thresholded detection report is examined. A set of five frequency bands roughly conforming to species of interest (cut off by the hydrophone response) have been selected, where beaked whales comprise band 2 (Table 1).

Table 1: Frequency bands

Band	Low Frequency (kHz)	High Frequency (kHz)
1	45	48
2	24	48
3	12	48
4	1.5	18
5	0	1.5

A ratio of bins above threshold to the total number of bins in each band is computed. If band two is selected, then the detection is tentatively classified as a beaked whale. Due to the fact that many of the bands overlap, a second check is performed by examining the number of bins set out-of-band. If this exceeds 10%, then the detection is reclassified as a dolphin as they are more likely to have significant spectral energy below 24 kHz.

Matched filter detector

A linear matched filter can be shown to be the optimal detector for known signals in white gaussian noise [7]. A high signal to noise ratio *M. densirostris* click extracted from the data set was used as the match template. The instantaneous output of the filter is then compared to an exponentially decaying time average of the filter output with a time constant of 0.1 seconds. If the instantaneous output exceeds the time average by a specified threshold, a detection is declared.

False alarm statistics

Both the FFT detector and the matched filter have been implemented as constant false alarm rate (CFAR) detectors. A direct comparison of detection performance between the two requires normalizing the false alarm rates. False alarm statistics have been computed in the presence of white Gaussian noise using the Box-Mueller pseudorandom noise generation algorithm from the GNU Scientific Library (GSL). The false alarm rate was then computed by dividing the number of false detections by the total run time for each threshold.

Two sets of results were compiled for the FFT detector. The first set indicates the performance of the first stage of the detector. As can be seen, this stage runs with a high false alarm rate.

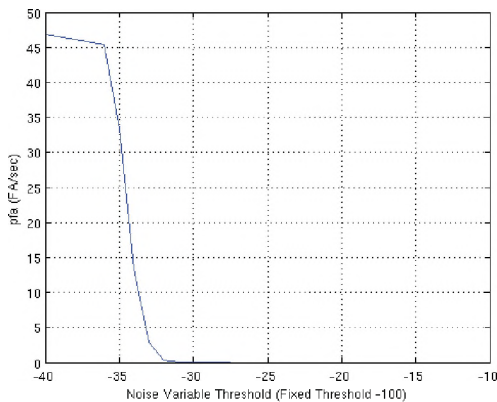


Figure 1: FFT Detector first stage false alarm curve

The false alarm rate drops dramatically at the output of the second stage. The main parameter determining performance is the click threshold used to determine when a sufficient number of bins have been detected to declare a click event.

The matched filter curve is typical and indicates a false alarm rate dropping exponentially with an increase in the threshold.

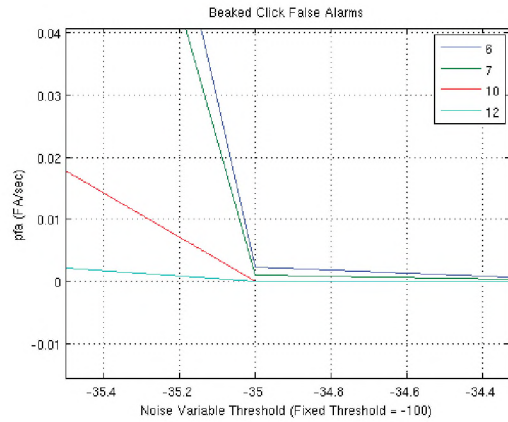


Figure 2: FFT Detector False Alarm Curve after filtering for beaked whales

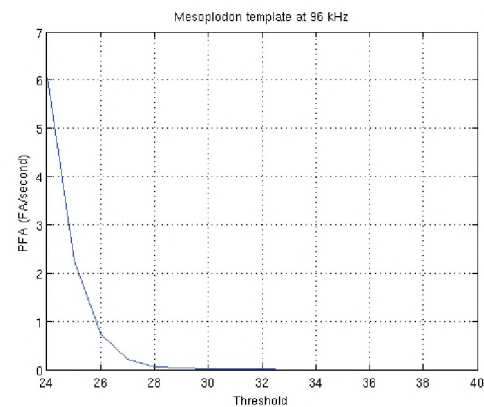


Figure 3: Matched filter false alarm curve

A false alarm rate of 1×10^{-3} was chosen as the test criterion. To achieve this rate, the following thresholds were chosen based on the false alarm curves:

FFT Detector: -35.1432
Matched Filter: 28.7009

Probability of Detection

Probability of detection statistics were compiled by creating a series of test data sets consisting of a high SNR click identified as *M. densirostris*. The click was scaled by a specified constant to achieve a desired signal level and repeated 827 times at a regular interval of 4/second. White Gaussian Noise was added to the signal to obtain the desired SNR.

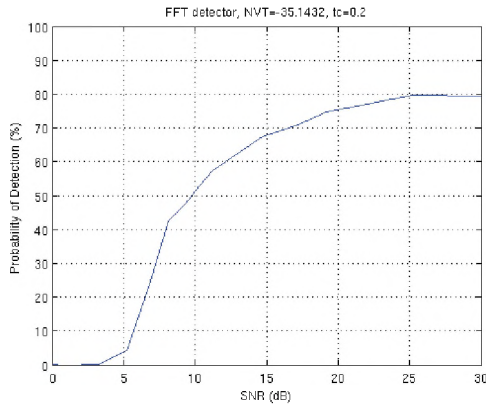


Figure 4: FFT detector probability of detection

The FFT detector tops out at approximately 80% probability of detection. Visual examination of the data indicates that some clicks are not present at the output of the first stage of the detector. The performance deficit at high SNR is therefore most likely linked to the choice of time constant for the exponential noise filter. A lower time constant may improve performance for regular clicks, however this remains to be investigated.

The matched filter provides the expected behavior when the match template exactly matches the signal present in the data set. This is never the case in practice. To estimate the effect of using an arbitrarily chosen high SNR template, a click from a completely separate data set (also collected at AUTEK) was used as a second match template. This is plotted in the rightmost curve in Figure 5. Using an arbitrarily chosen click degrades detector performance by approximately 2-3 dB. In either case probability of detection is at least 95% by 0 dB SNR.

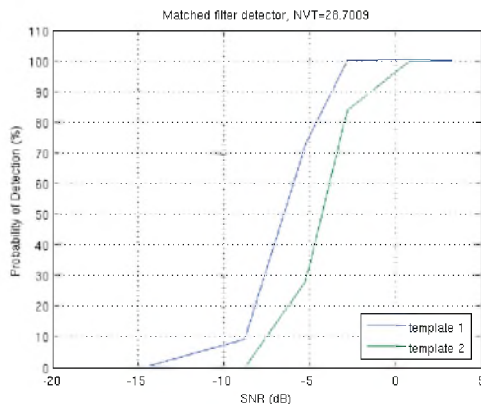


Figure 5: Matched filter probability of detection. Two match templates are plotted. First, template matching the click used to generate the dataset. Second template for high SNR click from a separate dataset.

Conclusion: 'Optimal' Detector

The matched filter significantly outperformed the FFT detector on the test data sets. This was true even when the

click used as the match template was not the same as the click used to generate the data sets. The probability of detection data shows a performance gain of at least 25 dB for the test cases studied. This suggests that the click structure is relatively constant with reasonably low variance between clicks from different individuals.

High SNR clicks are typically chosen as match templates. Due to the narrow beam width emitted by the animal, it is expected that most high SNR clicks received will be received when the animal has the ensonified receiver directly in the beam. The structure of these clicks may not be representative of off-axis clicks. In this case the matched filter will be sub-optimal at any aspect angle other than the one at which the match template was tuned for. However, in a widely spaced array such as AUTEK, it is possible to enhance detection and association significantly by improving detection performance at hydrophones which are farther away from the animal, but still in the beam. In this instance the matched filter may be employed to significant advantage.

2.5 Data Association

Clicks originating from the tagged animal have been identified on the surrounding bottom mounted hydrophones by matching inter-click interval patterns [5,8]. These patterns have been found to be an effective means of associating patterns of detections among hydrophones for sperm whales (*Physeter macrocephalus*). A fundamental assumption is that each animal exhibits its own unique pattern of clicks. The unique pattern is used as a template for a comb sieve that is correlated against the beaked whale clicks detected on the surrounding hydrophones [5]. The window with greatest number of correlations between the template and the hydrophone is assigned as the TDOA between the DTag and the hydrophone. After the comb sieve is complete, the probability density function of the TDOAs for each hydrophone is calculated in one minute windows. TDOAs that are significantly above the noise level in each window are passed on for use in localization and considered valid. The remaining TDOAs are considered invalid and not used further.

2.6 3D Hyperbolic Localization

Positions are computed from the valid TDOA sets using a hyperbolic multilateration positioning algorithm developed by Vincent [9]. Two 2500-ft depth XBT profiles were collected on 23 October 2006. These profiles were combined with a standard deep water profile and converted to sound speed by AUTEK. The sound speed profile nearest to the tagging location was used for calculating the direct path effective sound velocity [9]. TDOAs are required between at least four hydrophones and the DTag to compute a 3-D position. Due to the directional nature of the clicks, there were very few instances when an individual click was correlated on 4 hydrophones. Therefore, the TDOA for each hydrophone was interpolated using a piecewise cubic

Hermite interpolating polynomial function. The four hydrophones with the greatest number of valid time-difference of arrivals, 37, 43, 44, and 50, were used to create a Time of Arrival (TOA) matrix. The time of emission, x, y and z position of the sound source were estimated using the TOA matrix in the hyperbolic multilateration algorithm.

3. DISCUSSION

All results discussed in this paper are for the first deep dive recorded on the DTag, from approximately 34 to 91 minutes after tagging. The whale began vocalizing 6 minutes into the dive at 567 m depth, and continued to vocalize for approximately 36 minutes between 567 and 1049 m depth (Figure 6).

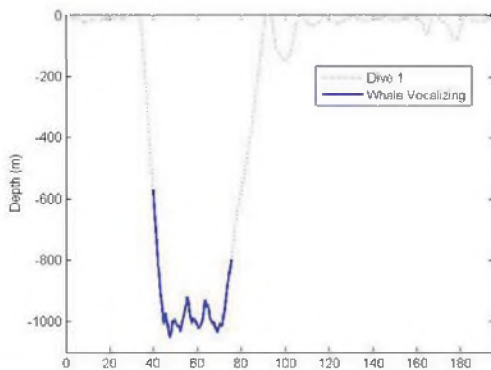


Figure 6: DTag Dive Profile Depth (m) vs. Time (minutes)

3.1 Detection Efficacy

Detection efficacy was assessed by evaluating how many of the clicks emitted by the tagged whale were successfully detected and associated on the nearby hydrophones. For each method, hydrophones 36, 37, 38, 42, 43, 44, 49, and 50 were processed through each detector and the beaked whale filter. These hydrophones were chosen by visually evaluating thresholded spectrogram data over the entire range for the presence of beaked whale clicks for the duration of the tagging event. Whale click times from the DTag were used as the template to search for correlation with the resulting TOAs produced using the FFT and matched filter detectors. Using the association algorithm detailed in Section 2.5, TDOAs were calculated between each hydrophone and the DTag.

During the 36 minute first dive, 5797 clicks were produced by the whale. Approximately 97% of these clicks were foraging clicks with an Inter-Click Interval (ICI) between 0.15 and 1 sec [3]. The mean foraging click ICI was 0.31 sec (std=0.05). This is in agreement with Johnson [4], who also observed a regular click ICI 0.37 seconds for a *M. densirostris* in the Canary Islands.

The FFT detector, implemented with a noise variable threshold of 34, was able to detect 49% of the clicks on at least one hydrophone. The matched filter detector,

implemented using a threshold of 28.7 was able to detect 92% of the clicks on at least one hydrophone. The filter template was a *M. densirostris* click recorded on an AUTECH hydrophone from a previous year. On each hydrophone, the matched filter detector performed significantly better than the FFT detector (Figure 7).

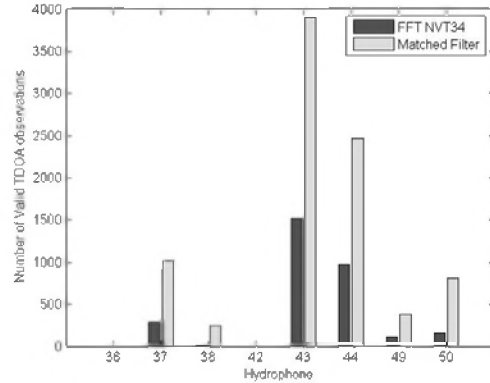


Figure 7: Detection efficacy per hydrophone

3.2 Localization

While the majority of the clicks were detected on at least one hydrophone using the matched filter, three dimensional localization was still difficult due to the need for at least 4 TDOAs between the DTag and hydrophone as input. Of the 5767 clicks associated with the hydrophones, only 1% were detected on four or more hydrophones. More commonly, the clicks were detected on only one hydrophone (44%), two hydrophones (36%), or three hydrophones (11%). Only 8% of the clicks were not detected at all. As a result, the TDOA trends were interpolated as discussed in section 2.6 prior to input into the hyperbolic multilateration algorithm. In addition, depth from the DTag was also used as an initialization parameter to provide better convergence of the solution.

The 577 localizations estimated using at least 3 measured TDOAs and only one interpolated value, were used to “ground-truth” the track estimated by the DTag alone [1] (Figure 8). Usually, the DTag “tag on” and “tag off” positions are known and can be used as absolute start and end positions. However, in this case the tagging vessel had to leave the range due to the presence of range operations and the “tag off” position is unknown. To “ground truth” the DTag track, small user-chosen sections of the track were individually fitted to time-synchronized 3D localizations by adjusting the swim speed to a least-squares match (Figure 9).

3.3 Detection Range

The 3D localizations created using the matched filter detection data were used to determine the range from the whale to the hydrophone for each detection method. Clicks from each hydrophone determined to be valid on the basis

of their TDOA with the DTag were used to estimate range and bearing to the hydrophone. The maximum detection range for both methods was approximately 6500 m, significantly greater than previously estimated [10]. The whale was traveling generally in a north-east direction, but was observed to turn at various times in all directions (Figure 10). The off-axis aspect angle between the caudal-rostral axis of the tagged whale and the receiving hydrophone was determined by subtracting the bearing angle from the whale to the hydrophone from the heading measured by the DTag. The detection range as a function of off-axis aspect angle is depicted in Figure 11 and Figure 12. The majority of the clicks detected at far ranges were within ± 30 degrees. With decreasing range, a greater number of clicks were detected further off-axis. While a -3 dB beam width of 6° has been suggested for *Z. cavirostris* by Zimmer [3], *M. densirostris* may be less directional due to their smaller body size and potentially smaller source aperture [10]. From these figures, it is evident that the matched filter detector significantly outperforms the FFT detector at longer ranges.

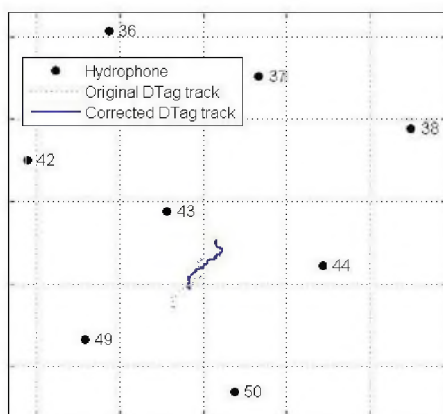


Figure 8: Original DTag Kalman-filtered track and corrected DTag track, each grid square is 2 km x 2 km.

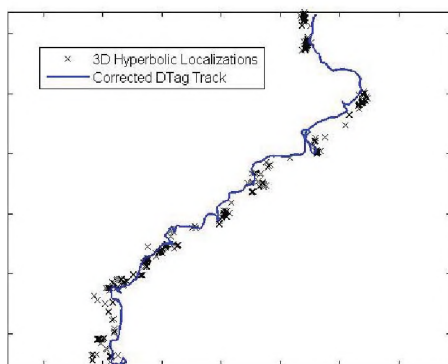


Figure 9: DTag track corrected based on 3D hyperbolic localizations, tick marks at 200-m increments

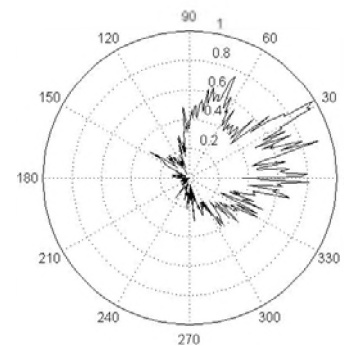


Figure 10: DTag heading for dive 1, Due North = 0, the radius axis is probability density (%)

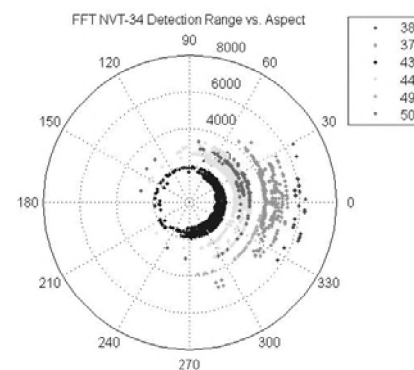


Figure 11: Detection range vs. aspect from the whale's head: FFT (NVT34) detector

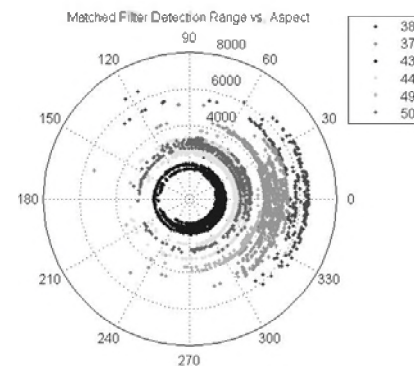


Figure 12: Detection range vs. aspect from the tagged whale's head: Matched filter detector

4. CONCLUSIONS

The matched filter detector performed significantly better than the FFT detector for *M. densirostris* foraging clicks. Using the matched filter detector, the wide-baseline AUTECH hydrophones were able to detect the tagged whale at up to 6500 m range. The off-axis aspect angle from the tagged whale to the hydrophone indicates the ability to detect signals significantly off-axis at lesser ranges and out to far

ranges when close to the axis. For Dive 1, 92% of the clicks produced by the tagged whale were detected on at least one hydrophone within the array. This combination of long detection ranges and increased probability of detection with the matched filter indicates that wide-baseline, broadband arrays, such as at AUTECH, provide an excellent opportunity for long-term monitoring of beaked whale populations and successful passive acoustic based mitigation.

5. ACKNOWLEDGEMENTS

We would like to acknowledge our sponsor, Dr. Frank Stone at N45, as well as the Office of Naval Research. We would also like to thank the Bahamas Marine Mammal Research Organization who provided observer support on the water. Finally we would also like to acknowledge the NUWC Division Newport Independent Laboratory Innovative Research program manager, Richard Philips, for providing the initial funding to study the data association methods presented here on sperm whales, and the AUTECH range staff for access to their considerable infrastructure.

6. REFERENCES

1. M. P. Johnson and P. L. Tyack, "A digital acoustic recording tag for measuring the response of wild marine mammals to sound." IEEE Journal of Oceanic Engineering, 28(1), pp 3-12, 2003
2. P. T. Madsen, M. Johnson, N. Aguilar de Soto, W. M. X. Zimmer and P. Tyack, "Biosonar performance of foraging beaked whales (*Mesoplodon densirostris*)", Journal of Experimental Biology 208, 181-194, 2005
3. Walter M. X. Zimmer, Mark P. Johnson, Peter T. Madsen, and Peter L. Tyack, "Echolocation clicks of free-ranging Cuvier's beaked whales (*Ziphius cavirostris*)", J. Acoust. Soc. Am., Vol. 117, No. 6, June 2005, pp.3919-3927
4. M. Johnson, P. T. Madsen, W. M. X. Zimmer, N. Aguilar de Soto, and P. L. Tyack, "Foraging Blainville's beaked whales (*Mesoplodon densirostris*) produce distinct click types matched to different phases of echolocation," The Journal of Experimental Biology (209), pp.5038-5050, 2006.
5. J. A. Ward, Sperm whale bioacoustic characterization in the Tongue of the Ocean, Bahamas (U). NUWC-NPT TR 11,398. Naval Undersea Warfare Center, Division Newport, RI, 20 September 2002 (Unclassified).
6. D. Moretti, N. DiMarzio, R. Morrissey, J. Ward, and S. Jarvis, "Estimating the density of Blainville's beaked whale (*Mesoplodon densirostris*) in the Tongue of the Ocean (TOTO) using passive acoustics," Oceans 2006, pp 1-5, September 2006.
7. Smith, "The Scientist and Engineer's Guide to Digital Signal Processing", California Technical Publishing, San Diego, CA, 1997, pp. 307-310
8. R. P. Morrissey, J. Ward, N. DiMarzio, S. Jarvis, D. J. Moretti, "Passive acoustic detection and localization of sperm whales (*Physeter macrocephalus*) in the tongue of the ocean", Applied Acoustics 67, pp 1091-1105, 2006.
9. Vincent, H. "Models, Algorithms, and Measurements for Underwater Acoustic Positioning," Ph.D. Dissertation. University of Rhode Island, Kingston, R.I., 2001.
10. P. L. Tyack, M. P. Johnson, W. M. X. Zimmer, P. T. Madsen, M. A. de Soto. "Acoustic behavior of beaked whales, with implications for acoustic monitoring," Oceans 2006, pp 1-6, September 2006.



Photo Credit: Bahamas Marine Mammal Research Organisation

THREE-DIMENSIONAL SINGLE-HYDROPHONE TRACKING OF A SPERM WHALE DEMONSTRATED USING WORKSHOP DATA FROM THE BAHAMAS

Christopher O. Tiemann

Applied Research Laboratories, University of Texas at Austin
P.O. Box 8029, Austin, TX 78713

ABSTRACT

A passive acoustic localization method for tracking the movement of a clicking sperm whale in three-dimensions using data from just one hydrophone is demonstrated using data made available for the 3rd International Workshop on Detection and Classification of Marine Mammals. One recording contains sperm whale clicks recorded on a bottom-mounted hydrophone on a steep slope of the Navy's AUTECH test range. When the direct-path acoustic ray arrivals from several clicks are time-aligned, persistent associated multipath arrivals of reflected ray paths can be identified for each click event and used for localization. Although the use of multipath arrival information is a standard procedure for range-depth tracking, a three-dimensional estimate of whale position can be obtained from the same multipath information with knowledge of an azimuthally-dependent environment relative to the receiver. In this case, azimuthal distinction arises from varied bathymetry. Multipath arrival patterns are matched to unique range-, depth-, and azimuth-dependent modeled arrival patterns to make an estimate of whale location. A three-dimensional whale track in range, depth, and bearing from the fixed hydrophone is presented.

SOMMAIRE

On démontre dans cet article une méthode de localisation acoustique passive pour suivre la trace d'un cachalot cliquant dans les trois dimensions utilisant les données d'un seul hydrophone, en utilisant les données disponibles à partir du Troisième Atelier Internationale sur la Découverte et la Classification des Mammifères Marins. On a enregistré les bruits secs du cachalot enregistré par un hydrophone monté au fond sur une pente raide du champ d'essai d'AUTECH de la Marine Américaine. Quand les arrivées des rayons acoustiques des plusieurs cliques avec des trajectoires directes sont alignés par intervalles, des arrivées multi-trajectoires associées et persistantes des trajectoires de rayons reflétés peuvent être identifiées pour chaque événement d'une clique et peuvent être utilisées pour la localisation. Bien que l'utilisation de l'information d'arrivée multi-trajectoire soit une procédure normale pour suivre la trace portée-profondeur, une estimation en trois dimensions de la position des cachalots peut être obtenue en utilisant les mêmes informations multi-trajectoires si l'on connaît l'environnement azimuthalement-dépendant relatif au récepteur. Dans ce cas, la distinction azimutale dérive de la bathymétrie variée. Les schémas des arrivées multi-trajectoires sont attribués aux schémas uniques qui dépendent de la portée, de la profondeur, et de l'azimut pour faire une estimation d'emplacement du cachalot. Une trajectoire en trois dimensions qui montre la portée, la profondeur, et le rapport du hydrophone fixe est donnée.

1. INTRODUCTION

Passive acoustic methods for monitoring marine mammal activity have been used for many years in censusing and behavioral studies, often in conjunction with visual surveys, because of the advantages they offer: they are unobtrusive and continue to work at times when animals are not visible (swimming underwater, nighttime, etc.) [1-3] Methods for not just detecting but also tracking the movement of animals underwater through analysis of their recorded vocalizations have advanced since early work that used geometric hyperbolic fixing techniques. [4-10] Techniques that exploit acoustic propagation models, multipath arrival information, or both can now provide alternate and possibly more accurate localization estimates. [11-15] Two-

dimensional (2D) location solutions (range and depth) can be achieved using data from as few as one hydrophone, but a full three-dimensional (3D) estimate (range, depth, and unique bearing relative to a sensor) has previously required the use of multiple receivers. [12,16-20] A technique that uses range estimates from a single hydrophone to make a hydrophone-relative 3D track does not provide an absolute measurement of azimuth.[20]

In 2006, Tiemann et al. [21] demonstrated a model-based technique for using acoustic data from just one hydrophone to make three-dimensional estimates of sperm whale locations in the Gulf of Alaska. The method exploits multipath arrival information from recorded sperm whale clicks, yet it does not require specific ray path identification (i.e., direct-path, surface-reflected). The technique can not only account for waveguide propagation physics (ray

interaction with the sea surface and sea floor in particular), but in fact relies upon reflections to estimate bearing to the source (whale). While the single-hydrophone localization technique was demonstrated successfully in its first application, a dataset made available for the 3rd International Workshop on Detection and Classification of Marine Mammals provided an opportunity to further exercise the localization algorithm in another much deeper environment. This paper describes the application and results of that localization attempt.

Among the data that the Naval Undersea Warfare Center (NUWC) provided for the workshop was a 10-minute recording (“test data #9”) containing numerous sperm whale vocalizations from what was assumed to be one animal. The data are from hydrophone #16 of the Navy’s Atlantic Undersea Test and Evaluation Center (AUTEc) in the Bahamas, positioned about 5 m off the sea floor at 1386 m depth, from March 3, 2006, at 09:48 local time with a 96 kHz sample rate. NUWC also provided bathymetry information from a multibeam survey of the range; Figure 1 shows the topography around the receiver that provided data for this demonstration.

2. METHODOLOGY

Sperm whale vocalizations appear in the data as brief (~10 ms) broadband clicks with an inter-click interval of about 1 sec. The collection of all the multipath echoes from a single click event is an arrival pattern, and one step in the localization process is to identify these patterns in the data. Additionally, an acoustic propagation model is used to predict the arrival patterns expected at the receiver from hypothesized impulsive sources at many ranges, depths, and bearings around the receiver; these modeled arrival patterns are called the replica. After comparing the measured arrival patterns with the replica, the hypothesized source position for which they agree best is the best estimate of whale location. Repeating the localization for each click event creates a track of animal motion.

In cases where the environment is unique along radials at every bearing around the receiver, a source at every range/depth/bearing bin will have a unique multipath arrival pattern “fingerprint.” It is azimuthal dependence in the environment that allows for bearing discrimination when using just one receiver, and the rough terrain around the hydrophone, as shown in Figure 1, ensures that the environment (bathymetry slice) looks different along radials in every direction.

2.1 Replica Generation

One input needed for the localization process is predictions of travel times for all ray paths between a source and the receiver for a suite of hypothesized sources on a grid of several ranges, depths, and bearings around the receiver. Specifically, the algorithm requires relative arrival times for all the multipath arrivals, not absolute travel times, so the time elapsed since the arrival of the direct ray path is calculated and saved for later use. The Gaussian beam

acoustic propagation model BELLHOP [22] provides travel time predictions for the eigenrays assuming a grid of sources spaced 10 m in range out to 5 km and 5 m in depth to 2 km. The model uses a different bathymetry profile for radials spaced every 5° in azimuth around the receiver; the bathymetry data was that provided by NUWC as shown in Figure 1. The model assumes a source frequency of 5 kHz, a range-independent downward refracting soundspeed profile taken from the Levitus historical database, and geoacoustic properties of fine-grained sediments. [23]

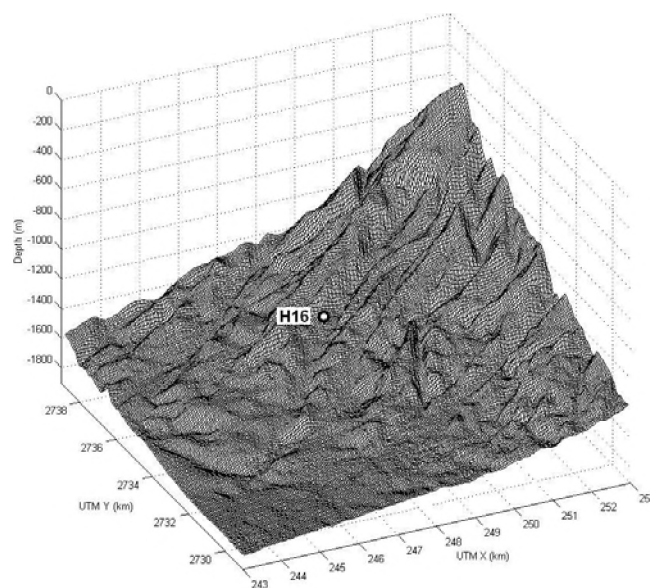


Figure 1. Bathymetry around AUTEc hydrophone #16. Coordinates are for UTM zone 18.

2.2 Arrival Pattern Extraction

A tool for automating the identification and extraction of arrival pattern information for click events in acoustic data is described in detail in Tiemann et al. [21], and it was used again here to extract information for 383 click events in the workshop data. A summary of its use follows.

Broadband sperm whale clicks are readily apparent when viewed as spectrograms like that of Figure 2a which shows four seconds of workshop data at the start of a click train; this spectrogram was made using 256-point fast Fourier transforms with 50% window overlap on data downsampled to 44 kHz (6 ms of data for each FFT). To make fainter click arrivals more apparent over background noise, spectrograms are summed over the frequency bins from 3 kHz to 22 kHz within each time bin, as shown in Figure 2b. Each peak in the spectral sum time series represents an arrival from either a direct or reflected ray path, with the direct paths typically having greater amplitudes than their associated echoes. In this example, the arrivals for five direct paths are spaced about 0.7 sec apart. Note that the inter-click interval between click events

was often less than the time separation between the two loudest arrivals of a given click event, as in this example. (The weaker reflected-path arrival at 393.9 sec is associated with the direct path arrival at 393.0 sec, not the one at 393.7 sec.)

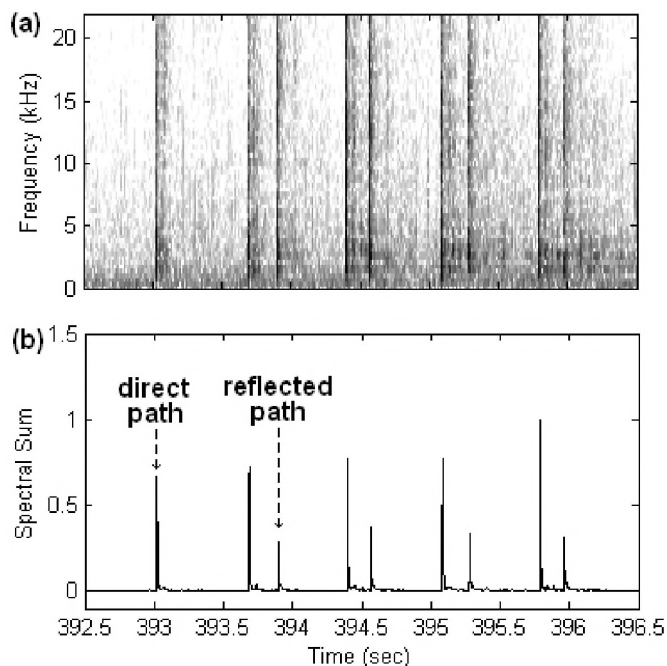


Figure 2. (a) Spectrogram of acoustic data from AUTEK; broadband sperm whale clicks appear as vertical stripes. (b) Spectrogram summed over frequency bins; broadband clicks appear as peaks.

The automated pattern matching tool assists in identifying the direct-path (earliest) arrivals for several click events in a click train. In order to assist in recognition of possibly faint multipath arrivals, a display tool time-aligns windows from the spectral sum time series which begin at the direct-path arrival. These aligned spectral sum excerpts can then be viewed as a two-dimensional color surface like that of Figure 3 where each horizontal slice conveys the relative amplitude and arrival time information of all arrivals occurring within 1.2 sec of a direct path arrival at relative time 0 sec. Note that the absolute time axis of Figure 3 indicates the time since the beginning of the data set; the relative time axis is the time elapsed since the direct path arrival of a given click event.

Persistent peaks in these surfaces that are time-aligned over every horizontal slice represent multipath arrivals that will be compared to the modeled arrival patterns, and another tool helps extract the arrival pattern information from these surfaces. Identifying the arrivals as either from surface-reflected paths, bottom-reflected paths, etc. is not required for the automated localization to follow, but such an interpretation of Figure 3 is provided as an example. The

order of the arrivals in this example is direct path, bottom-bounce path, surface-bounce path, and bottom-surface-bounce path as labeled on Figure 3. Note that the isolated peaks on these surfaces that do not align in time with other peaks can be ignored during arrival pattern extraction as they are not associated with a given click event under consideration. For example, the isolated peaks around relative time 0.7 seconds are direct-path arrivals of the *next* click event; the ~ 0.7 sec inter-click interval is shorter than the length of the spectral sum excerpts used to make this figure.

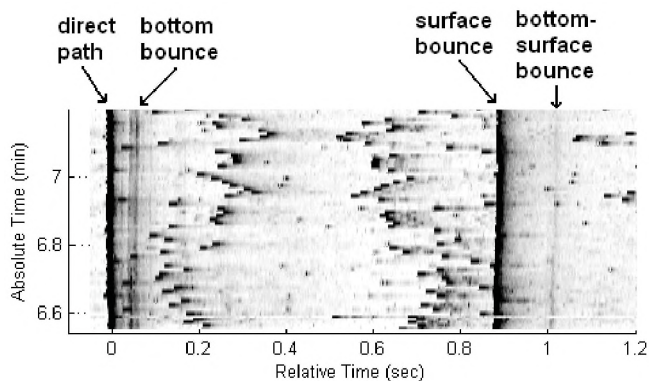
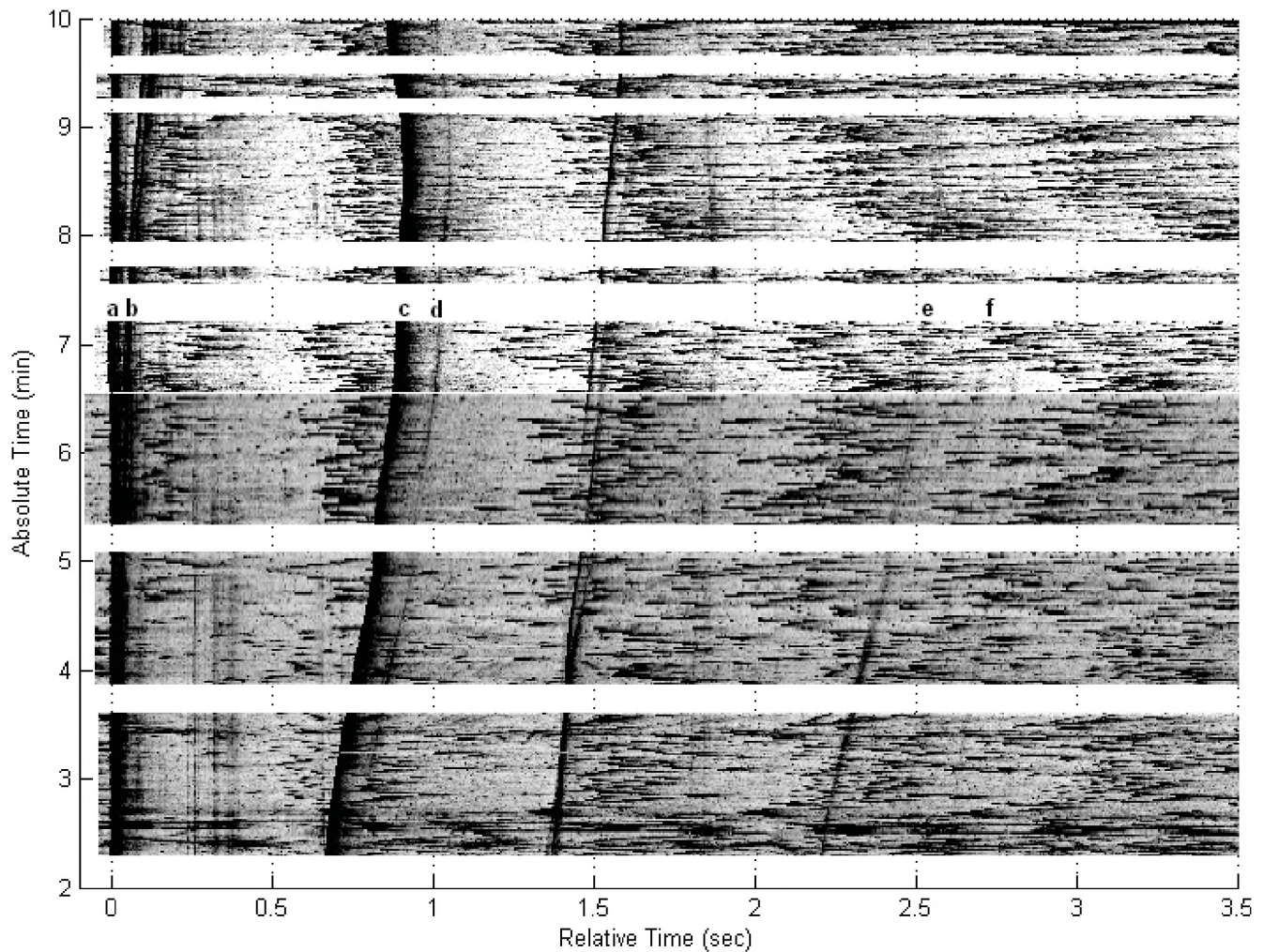


Figure 3. Time-aligned spectral sum excerpts, each starting with a direct-path arrival at relative time 0 sec, represented as a 2D surface.

A benefit of making the time-aligned spectral sum surfaces is that they allow the eye to integrate over multiple click events to recognize faint broadband arrivals that would not be obvious in any single spectral sum excerpt. For example, Figure 4 shows one of these surfaces made using data from most of the 10-minute workshop recording, 383 click events in all. Color scales were adjusted in this figure in attempts to make the faintest persistent arrivals visible, but doing so increases the visible clutter from non-aligned arrivals.

The same four persistent arrivals identified in Figure 3 are present in Figure 4 and are labeled there as well, but note how the relative spacing between the arrivals evolves with time. This is expected as the source is changing position relative to the fixed receiver, and it will have a new arrival pattern at every new location. The lengthened relative time axis of Figure 4 (now 3.5 sec) also shows two more arrivals around 2.5 sec relative time. The acoustic model predicts these should be present then and identifies them as a surface-bottom-surface reflected path and a bottom-surface-bottom-surface reflected path. To put that 2.5 sec delay into context, in the time between the direct-path arrival and its late echo, *three more* click events occurred, yet that late, faint arrival can still be associated with the correct direct-path arrival.



Ray path identification:

- | | |
|---------------------------|---|
| (a) direct path | (d) bottom-surface bounce |
| (b) bottom bounce | (e) surface-bottom-surface bounce |
| (c) surface bounce | (f) bottom-surface-bottom-surface bounce |

Figure 4. Time-aligned spectral sum excerpts, each starting with a direct-path arrival at relative time 0 sec, represented as a 2D surface. Persistent arrivals with identified ray path geometries are labeled.

Identification of the ray path geometries, though not required, can be challenging when the data indicates there are more ray arrivals than a model would predict, as was the case in this recording. For example, notice the arrivals present at about 1.5 sec relative time. These arrivals are curious not only because they split, merge, and even disappear but also because there is no match anywhere close to them in the modeled arrival patterns. One hypothesis to explain these arrivals is that they are from reflected paths outside of the vertical plane of propagation connecting the source and receiver. The acoustic propagation model used is limited to 2D problems (range/depth slices), yet the map of Figure 1 shows a rough 3D terrain with several ridges around the

receiver which may allow for reflections of ray paths outside the vertical plane of modeling.

For another example of unusual arrivals, notice several that arrive at about 0.3 sec relative time, shortly after the direct path arrival. These arrivals are persistent throughout the entire record, and they are also locked in time to the arrival of the direct path, i.e. their relative arrival time never changes. Every other arrival shifts in time as the source moves except these. One hypothesis to explain these is that they are due to sound reflecting off the receiver housing, then off some nearby fixed terrain feature and back to the receiver. That scenario would cause stripes of constant relative arrival time like those seen in the figure regardless of the source location.

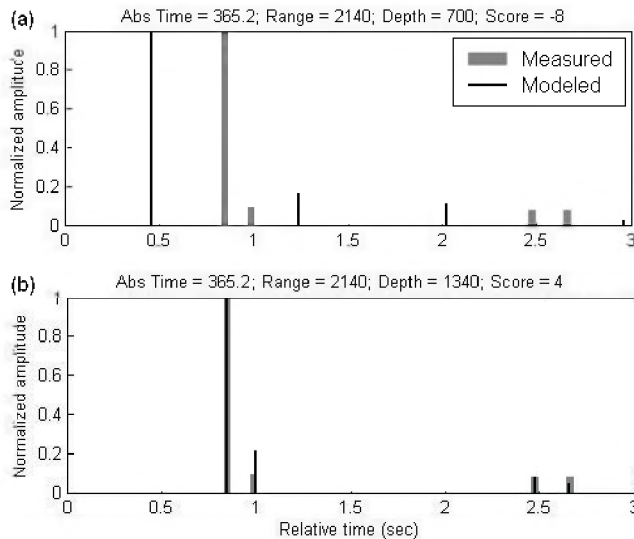


Figure 5. (a) Arrivals from a modeled source at 2140 m range, 700 m depth overlaid on measured arrivals from time 365.2 sec. No arrivals overlap, so it is given a low overlap score. (b) The same measured arrivals overlay modeled arrivals from 2140 m range, 1340 m depth. All measured and modeled arrivals overlap.

2.3 Ambiguity Surface Construction

The source location for each click event is estimated through the construction of bearing-dependent ambiguity surfaces. Each surface graphically conveys the likelihood that a whale was at a given range/depth bin when it vocalized. These surfaces have the same resolution as the replica, and a scoring mechanism assigns a score to each hypothesized source location based on how closely a measured arrival pattern matches the modeled arrival pattern for that location. The scoring technique and ambiguity surface construction are described in detail in Tiemann et al. [21]; a summary of its use here follows.

The score for every candidate source position is calculated by first counting the number of measured arrivals that have the same relative arrival times as those in the replica for that source position. From this score is subtracted the number of arrivals in both the data and replica that do not have a match in relative arrival time. A tolerance of 10 ms is used in defining a match in relative arrival times as that was a typical duration for a recorded click. To illustrate the scoring process, Figures 5a and 5b show the relative arrival time and amplitude information for a measured arrival pattern overlaid by modeled arrival patterns for two candidate sources at the same range but different depths. The first example has no overlapping arrivals among the 8 considered for a total score of -8; the second example shows all 4 arrivals from both the data and replica matching to result in a score of +4. Note that for early parts of the data set the direct-path

and bottom-bounce arrivals were difficult to distinguish. Therefore, the bottom bounce path was not considered during the scoring process.

Scores are calculated for all candidate source positions and presented on an ambiguity surface like that of Figure 6, one surface for each bearing radial in the replica. Peaks on these surfaces indicate likely source positions, and the global maximum among all ranges, depths, and bearings is declared the best estimate of source location. Repeating the scoring process for every click event results in a track of whale motion.

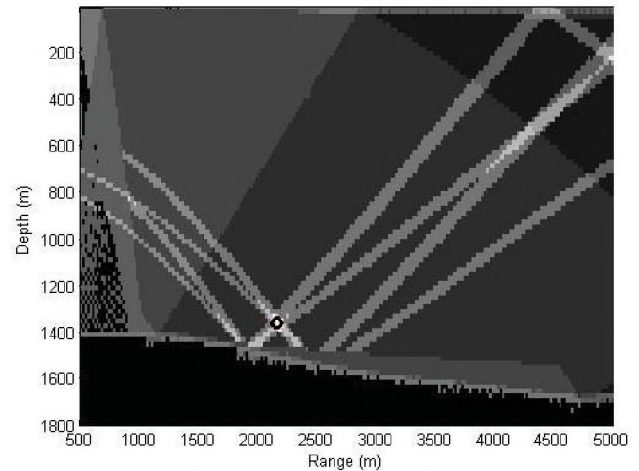


Figure 6. An ambiguity surface showing overlap scores on a vertical range/depth slice along bearing 200° from the receiver. The peak at 2140 m range, 1340 m depth is marked and indicates the best estimate of source location for the click event at time 365.2 sec.

3. RESULTS

The range, depth, and bearing estimates resulting from the localization process described above are presented in Figure 7. The whale track originates 3 km to the southwest of the receiver, ending within 2 km of the receiver and shallower. No independent ground truth data of whale motion was available for this workshop data, but the whale's average speed over this track was 2.9 m/s, comparable to other acoustically derived sperm whale swim speeds. [18,19] Figure 8 puts this track in context of the AUTECH environment, overlaying estimated positions on a plan view of the bathymetry contours. One can see that the whale's change in depth tracks the shallowing terrain. It is the 5° binning in the replica that causes a disjointed track when bearing changes; the whale did not suspend clicking.

4. DISCUSSION AND CONCLUSION

The ability to passively monitor the movement of marine mammals underwater benefits behavioral studies of such,

but the multi-sensor arrays typically required to make 3D tracks of animal motion increase the cost and complexity of those studies. An economical single-hydrophone solution for doing the same should be of benefit to the bioacoustics community. A previous effort demonstrated how to produce such a 3D estimate of sperm whale location by exploiting multipath arrival information from a single sensor, and this work used data provided by the 3rd International Marine Mammal Workshop to demonstrate that the same technique seems viable in deep water environments as well, at least in locations where there is some environmental variation in azimuth.

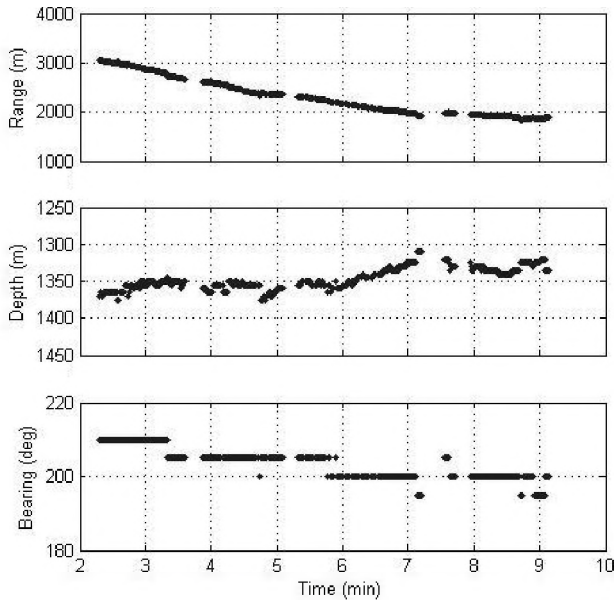


Figure 7. Range, depth, and bearing estimates relative to hydrophone #16 of the AUTEC range for a clicking sperm whale.

Azimuthal dependence in the environment is necessary in order to obtain a bearing estimate to the source. In locations where the topography is flat along all radials from the receiver, no unique bearing information can be obtained; however, the mechanics of the localization algorithm work exactly the same in providing a two-dimensional range/depth location estimate. Again, multipath arrival information can be exploited without knowledge of ray path geometry. A second recording provided for the marine mammal workshop (“test data #6”) also contained sperm whale click trains, but it was taken from a hydrophone on a relatively flat part of the AUTEC range. Although not presented here, range/depth estimates of whale location were made from that data, but bearing information could not be resolved.

Questions of accuracy are warranted when demonstrating a localization technique, and there are two types of error that can negatively affect the method shown here: mismatch between the modeled and truth environment, and errors in measurements of the relative travel times between multipath arrivals of a given click

event. Both contribute nonlinearly to an overall error. A thorough sensitivity analysis for this technique was described in Tiemann et al. [21] showing it to be reasonably robust against such errors: 14 m error due to environmental mismatch and 16 m error due to measurement inaccuracy in a simulated case.

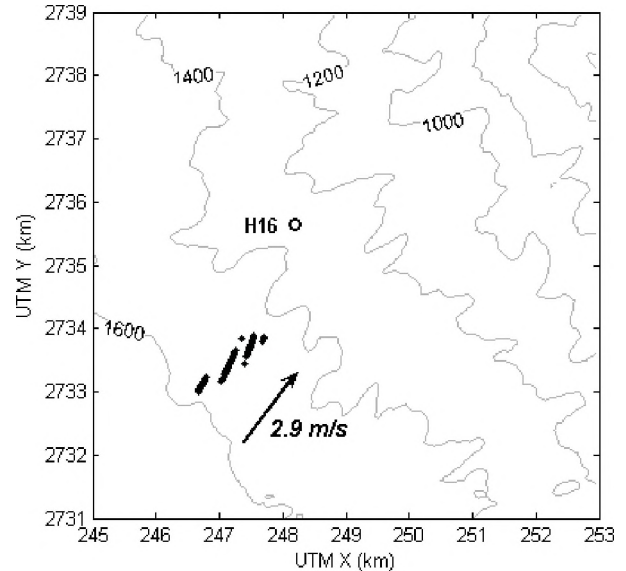


Figure 8. Estimates of sperm whale position overlaid on a plan view of bathymetry contours from the AUTEC range. Coordinates are for UTM zone 18; depths are in meters.

Lastly, although not implemented as such, the computational method for localization as presented here could be considered just a variation on the traditional hyperbolic fixing technique. In standard hyperbolic fixing, the time delay between an event recorded at two sensors defines a hyperbola of candidate source positions. The intersection of hyperbolas traced from multiple receiver pairs localizes the sound source. The same technique is basically being applied here (minus the isovelocity medium assumption), but in this single-hydrophone case, those additional receivers are *virtual*.

Recall that in ocean waveguide propagation, a ray path reflection off a boundary creates a virtual receiver vertically offset from the real receiver. Those real and virtual receivers are the foci in the definition of a hyperbola. The time of arrival difference between the direct and a reflected path, as measured during the arrival pattern extraction process, is like the time delay between an event arrival at the real and virtual receivers. Each additional multipath arrival time that can be measured contributes another virtual receiver and thus another hyperbolic path to the localization. To illustrate, consider the ambiguity surface of Figure 6. A direct ray path plus four late arrivals contributed to making this surface, so there are four paths of relatively high overlap scores. Those high-score areas even resemble hyperbolas for at least 2000 m away from the receiver, and they all

intersect at one range/depth bin to provide the localization estimate. With this understanding, a more elegant analytic solution for exploiting relative arrival times may become apparent as a substitute for the brute force computational method described here.

5. ACKNOWLEDGEMENTS

Thanks go to the mammal workshop organizers and to NUWC for providing not only the workshop dataset but also additional geometry and bathymetry information from AUTECH which made the localization possible. The Applied Research Laboratories at the University of Texas at Austin funded this analysis and algorithm development. Thanks also to Dr. Michael Revesz at ARL:UT for many helpful discussions and to Katherine Zaunbrecher and Jared LeBlanc for providing a translation of the abstract.

6. REFERENCES

1. R. Leaper, O. Chappell, and J. Gordon, "The development of practical techniques for surveying sperm whale populations acoustically," *Rep. Int. Whal. Comm.* **42**, 549–560 (1992).
2. J. Barlow and B.L. Taylor, "Estimates of sperm whale abundance in the northeastern temperate Pacific from a combined acoustic and visual survey," *Marine Mammal Sci.* **21**(3), 429–445 (2005).
3. C. Tiemann, S. Martin, J. Mobley Jr., "Aerial and acoustic marine mammal detection and localization on Navy ranges," *IEEE J. Ocean Eng.* **31**(1), 107–119 (2006).
4. S. Mitchell and J. Bower, "Localization of animal calls via hyperbolic methods," *J. Acoust. Soc. Am.* **97**, 3352–3353 (1995).
5. C.W. Clark, W.T. Ellison, and K. Beeman, "Acoustic tracking of migrating bowhead whales," *IEEE Oceans Conference Proceedings* **18**, 341–346 (1986).
6. J.L. Spiesberger and K.M. Fristrup, "Passive localization of calling animals and sensing of their acoustic environment using acoustic tomography," *Am. Nat.* **135**, 107–153 (1990).
7. A.S. Frankel, C.W. Clark, L.M. Herman, and C.M. Gabriele, "Spatial distribution, habitat utilization, and social interactions of humpback whales, *Megaptera novaeangliae*, off Hawai'i determined using acoustic and visual techniques," *Can. J. Zool.* **73**, 1134–1146 (1995).
8. K.M. Stafford, C.G. Fox, and D.S. Clark, "Long-range acoustic detection and localization of blue whale calls in the northeast Pacific Ocean," *J. Acoust. Soc. Am.* **104**(6), 3616–3625 (1998).
9. V.M. Janik, S.M. Van Parijs, and P.M. Thompson, "A two-dimensional acoustic localization system for marine mammals," *Mar. Mamm. Sci.* **16**, 437–447 (2000).
10. C.W. Clark and W.T. Ellison, "Calibration and comparison of 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 (2000).
11. C.O. Tiemann, M.B. Porter, and L.N. Frazer, "Localization of marine mammals near Hawaii using an acoustic propagation model," *J. Acoust. Soc. Am.* **115**, 2834–2843 (2004).
12. A. Thode, "Three-dimensional passive acoustic tracking of sperm whales (*Physeter macrocephalus*) in ray-refracting environments," *J. Acoust. Soc. Am.* **118** (6), 3575–3584 (2005).
13. A. Thode, D. K. Mellinger, S. Stienessen, A. Martinez, and K. Mullin, "Depth-dependent acoustic features of diving sperm whales *Physeter macrocephalus* in the Gulf of Mexico," *J. Acoust. Soc. Am.* **112**(1), 308–321 (2002).
14. W. M. X. Zimmer, M. P. Johnson, A. D'Amico, and P. L. Tyack, "Combining data from a multisensor tag and passive sonar to determine the diving behavior of a sperm whale *Physeter macrocephalus*," *IEEE J. Ocean. Eng.* **28**(1), 13–28 (2003).
15. E.-M. Nosal and L.N. Frazer, "Sperm whale three-dimensional track, swim orientation, beam pattern, and click levels observed on bottom-mounted hydrophones," *J. Acoust. Soc. Am.* **122**(4), 1969–1978 (2007).
16. R. Aubauer, M.O. Lammers, and W.W.L. Au, "One-hydrophone method of estimating distance and depth of phonating dolphins in shallow water," *J. Acoust. Soc. Am.* **107**, 2744–2749 (2000).
17. P.A. Lepper, K. Kaschner, P.R. Connelly, and A.D. Goodson, "Development of a simplified ray path model for estimating the range and depth of vocalising marine animals," in *Proc. Inst. Acoust.* (St. Albans, 1997), pp. 227–234.
18. M. Wahlberg, "The acoustic behaviour of diving sperm whales observed with a hydrophone array," *J. Exp. Mar. Biol. Ecol.* **281**, 53–62 (2002).
19. E.-M. Nosal and L.N. Frazer, "Track of a sperm whale from delays between direct and surface-reflected clicks," *Appl. Acoust.* **67**, 1187–1201 (2006).
20. C. Laplanche, O. Adam, M. Lopatka, and J. Motsch, "Male sperm whale acoustic behavior observed from multipaths at a single hydrophone," *J. Acoust. Soc. Am.* **118**, 2677–2687 (2005).
21. C. Tiemann, A. Thode, J. Straley, V. O'Connell, K. Folkert, "Three-dimensional localization of sperm whales using a single hydrophone," *J. Acoust. Soc. Am.* **120**(4), 2355–2365 (2006).
22. M.B. Porter and H.P. Bucker, "Gaussian beam tracing for computing ocean acoustic fields," *J. Acoust. Soc. Am.* **82**(4), 1349–1359 (1987).
23. F.A. Bowles, "A geoacoustic model for fine-grained, unconsolidated calcareous sediments," ARSRP Natural Laboratory, NRL/MR/7432-93-7082, March 9, 1994.

A COMPARISON OF PITCH EXTRACTION METHODOLOGIES FOR DOLPHIN VOCALIZATION

Xanadu C. Halkias and Daniel P. W. Ellis

LabRosa, Columbia University, Department of Electrical Engineering,
1300 S.W. Mudd, 500 West 120th Street, New York, NY, 10027

ABSTRACT

When collecting and analyzing marine mammal vocalizations one of the most important goals is to automatically extract the pitch/fundamental frequency of the collected calls. In dolphins we can assume that there are two main pitched sounds: whistles, which can be described as tonal AM-FM signals, and bursts, which can be described as highly harmonic signals. There are three main difficulties with pitch extraction on dolphin vocalizations that arise from the nature of the data. First, most underwater recordings are restricted to a low signal-to-noise ratio due to reflections, hardware noise and other interferences. This constitutes a big challenge for most existing pitch trackers. Second, one has to take into account the significant differences in the frequency range of bottlenose dolphin vocalizations compared to humans. Finally, dolphin whistles and bursts generally are emitted in two distinct frequency ranges, which result in different modes in the analysis data. In this work we compare our novel pitch extraction approach with two widely popular algorithms. Our approach uses hierarchy-based hidden Markov models (HMM) with cepstral coefficients as features. We quantitatively compare the performance of our algorithm with Yin, which is based on a modified autocorrelation method and `get_f0`, a popular off-the-shelf pitch tracker that utilizes linear predictive coefficients (LPC) and dynamic programming. Our approach outperforms the comparative methods by at least a factor of 10%.

SOMMAIRE

Pour la collecte et l'analyse de vocalises de mamifères marins, l'extraction de la fondamentale est une étape cruciale. Dans le cas des dauphins, nous pouvons considérer qu'il y a deux types de sons voisés : les chants qui peuvent être décrits comme des tonalités AM-FM, et les rafales ("bursts") constituées de signaux hautement harmoniques. La première des trois difficultés pour extraire le timbre est le très faible rapport signal sur bruit dû aux réflexions multiples et autres interférences. La seconde consiste à appréhender les résolutions harmoniques sur le signal de cétacés par rapport aux traitements connus en parole par exemple. Dans ce papier, nous testons notre nouvelle méthode d'extraction de timbre sur un modèle Chaîne de Markov Cachée Hiérarchique à partir de coefficient cepstraux. Nous comparons nos résultats à la méthode YIN basée sur un calcul d'autocorrélation, et à `get_f0` qui est extracteur de timbre classique par programmation dynamique utilisant des coefficients LPC. Nous montrons que notre méthode apporte un gain de 10% par rapport à ces méthodes.

1. INTRODUCTION

When analyzing dolphin vocalizations, one of the most important tasks is the extraction of the fundamental frequency/pitch of the desired calls. Several methodologies for attempting to automatically extract pitch exist as scientists are extensively studying this problem, especially with respect to human speech and musical recordings.

Most existing packages used by researchers in the analysis of dolphin vocalizations require manual interaction for the extraction of the desired calls. Moreover, they do not extract the pitch at a per-frame level; rather they provide a frequency range that is manually obtained. These packages, such as Ishmael [1] and Raven [2] are widely used in the field and have been valuable tools for onsite researchers.

In order to resolve the problem of pitch extraction on dolphin vocalizations without manual interaction we utilize methodologies that have been effectively applied in the fields of speech and music processing. One such technique is Hidden Markov Models (HMM) [3, 4]. HMM's can be used either directly on the unprocessed spectrogram of the audio or in combination with the extraction of descriptive features e.g. mel frequency cepstral coefficients. Another class of algorithms that have been used widely in speech processing is based on the auto correlation of the signal or some transformed variation of it.

These different algorithms need domain engineering in order to take into account the intricacies of dolphin recordings as compared to human speech. As discussed in the introduction, there exist two main differentiating

issues when analyzing dolphin vocalizations. The first and arguably most important difference is that dolphin recordings exhibit extremely low signal to noise ratios, which require a careful selection of robust features to be used.

Next, almost all information in human speech exists in the range below 4kHz and 20 kHz for music. These ranges could be considered the low register for dolphins that can vocalize above 90kHz. This suggests three problems with using off-the-shelf algorithms designed for speech. Parameters such as filter cutoffs and domain-specific tuning curves must be modified to accommodate the revised frequency range. Also, the wider signal bandwidth of interest necessitates a much higher dimensionality of the feature space. Most importantly, however, is that the frequency range of harmonic content present in a dolphin vocalization can be much higher than the Nyquist rate of most commonly used underwater recording devices. Unfortunately, this means much of the high-frequency content of the dolphin calls is lost during recording.

Finally, there are differences in the frequency ranges of the various types of dolphin vocalizations. It is thus possible to cluster a call based on vocalization type, and using a different classifier for each type. This suggests a hierarchical or two-stage pitch extraction system. Knowing the frequency ranges of each type of call allows us to build classifiers with a lower feature space dimensionality.

In this paper we proceed by providing an explanation of three methodologies used for pitch extraction in section 2. Section 3 summarizes the experimental results, and finally in section 4 we discuss the implications of the nature of dolphin calls on the design of pitch extraction algorithms.

Algorithm	Feature	Classifier
Cepstrum+HHMM	256 Cepstral coefficients	Hierarchy HMM
YIN	Modified autocorrelation	Local minimum
get_f0	LPC residual	Dynamic Programming

Table 1: Description of pitch extraction methodologies

2. PITCH EXTRACTION METHODS

Three algorithms are used in order to achieve a comparative result in the desired pitch extraction task. Our novel approach consists of the use of a hierarchy/decision based HMM with the use of cepstral coefficients. The second algorithm, YIN [6], is widely used in speech processing for single pitch extraction and is based on a modified autocorrelation method. Finally, indicative results from get_f0 [7, 8], a popular off-the-shelf pitch tracker are obtained. Table 1 summarizes the three algorithms and the features used.

2.1 Cepstral coefficients with hierarchically driven hidden Markov models (HMM)

Hidden Markov models (HMM) [3, 4] have been extensively used in many natural sequences such as speech, language and handwriting. They provide us with a valuable tool for the analysis and extraction of information of time dependent data.

As previously discussed, there needs to be a robust selection of features that will be able to overcome the inherent low SNR present in the recordings. In this work the use of the cepstrum is preferred given its ability to highlight the pitch of a given signal. Through the existing literature the cepstrum [5] has been successfully employed in speech to obtain the desired pitch. It assumes a source-filter model and provides a homomorphic deconvolution thus separating the detailed excitation part of the signal from the broad, filter part.

We utilize the real cepstrum using the observation that a pitch peak will appear at the high n coefficients. The real cepstrum is defined as the inverse Fourier transform of the log magnitude of the spectrum of our signal. This can be seen in Equation 1.

$$c_n = \int \log|X(e^{j\omega})| \cos n\omega d\omega \quad (1)$$

Figures 2, 3 show an example of the cepstral coefficients for one call.

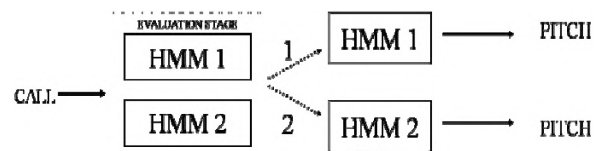


Figure 1: Description of HHMM system

We remove, as is common, the first coefficient, which captures the average energy of the original signal.

Arguably, the use of the spectrogram or even a normalized version of it could have been a suitable feature, it is clear that it will not provide a good noise suppression feature given that noise will be distributed across all frequencies, thus interfering with the task at hand.

Domain engineering suggests the existence of narrowband clusters of calls within the wide frequency range of vocalizations. In this work, analysis of the data as will be seen in section 4.1 dictated the use of a decision level. The idea behind this implementation is that there is an inherent bimodality in the data that can be taken advantage of with the use of a hierarchy. Initially, two HMM's are created with different number of hidden states that correspond to different frequency ranges of the calls. For every input vector both HMM's are evaluated using the forward algorithm and the one that gives us the highest likelihood is activated for the implementation of Viterbi decoding [3, 4], thus obtaining the most likely path across the hidden states.

Figure 1 provides a schematic description of the utilized system. It is evident from Figure 1 that two different HMM's are used each of them with different parameters representing two different frequency ranges respectively.

Each HMM is defined through parameters, λ that are extracted from the data. In this work both HMM's are continuous, implying that each state, q , can be represented with a single Gaussian probability density function:

$$\lambda_1 = (\pi_1, A_1, E_1), \text{ where } \pi_1 = (\pi_{11}, \pi_{12}, \dots, \pi_{1N}), N = 1, 2, \dots, 18 \quad (2)$$

$$A_1 = \{a_{1ij}\}_{i,j=1,2,\dots,N}, \text{ where } a_{1ij} = P(q_{1i} = j | q_{1i-1} = i)$$

$$E_{1N}(q) = N(\sigma_1 | \mu_{1N}, \sigma_{1N}), N = 1, 2, \dots, 18$$

$$\lambda_2 = (\pi_2, A_2, E_2), \text{ where } \pi_2 = (\pi_{21}, \pi_{22}, \dots, \pi_{2M}), M = 1, 2, \dots, 52 \quad (3)$$

$$A_2 = \{a_{2ij}\}_{i,j=1,2,\dots,M}, \text{ where } a_{2ij} = P(q_{2i} = j | q_{2i-1} = i)$$

$$E_{2M}(q) = N(\sigma_2 | \mu_{2M}, \sigma_{2M}), M = 1, 2, \dots, 52$$

$$\pi_2 = (\pi_{21}, \pi_{22}, \dots, \pi_{2M}), M = 1, 2, \dots, 52$$

$$A_2 = \{a_{2ij}\}_{i,j=1,2,\dots,M}, \text{ where } a_{2ij} = P(q_{2i} = j | q_{2i-1} = i)$$

$$E_{2M}(q) = N(\sigma_2 | \mu_{2M}, \sigma_{2M}), M = 1, 2, \dots, 52$$

where the states sets q_1, q_2 represent the frequency ranges of approximately 2.2kHz-11kHz and 440Hz-740Hz respectively. A noise state is also added for every HMM in order to capture the lack of pitch in a particular frame. Each state, q represents a pitch delay number that can be directly mapped to a specified frequency. Also, π_1, π_2 define the priors for state sets q_1, q_2 respectively as obtained from the statistics of the ground truth data. A_1, A_2 define the transition matrices for each HMM directly obtained from our ground truth and E_1, E_2 are the emission distributions for each state set. These are single 256 dimensional Gaussian distributions obtained from the extracted cepstral coefficients.

Once the parameters of the HMM's have been extracted we proceed to evaluate every call in order to identify its frequency range. This is shown in Equations 4, 5. The last stage of the system employs Viterbi decoding [3, 4] in order to find the most likely path across the evaluated state set, thus extracting the desired pitch at every frame, Equation 6.

$$p(Y_1) = \sum_{q_1} p(Y_1 | q_1) p(q_1) \quad (4)$$

$$p(Y_2) = \sum_{q_2} p(Y_2 | q_2) p(q_2) \quad (5)$$

$$\text{if } p(Y_2) > p(Y_1) \text{ then} \\ P = \max_{q_1, \dots, q_T} p(q_{11}, q_{12}, \dots, q_{1T-1}, q_{1T}, Y_1, Y_2, \dots, Y_T | \lambda_1) \quad (6)$$

Where $Y_t, t = 1, 2$ is a sequence of observations, $q_t, t = 1, 2$ is a sequence of the hidden states, $q_t, t = 1, 2, \dots, T$ is the maximum probability state path and λ_1 are the parameters of the HMM.

2.2 YIN: A fundamental frequency estimator

Yin, created by de Cheveigne and Kawahara [6], is a widely used algorithm for the estimation of the fundamental frequency/pitch of speech or monophonic musical sounds. It is based on a modified autocorrelation method and is extremely successful in extracting single pitches. Its popularity is also enhanced by the fact that it is a relatively simple and efficient algorithm, thus minimizing the computational cost.

Since our goal is to extract the fundamental frequency of dolphin vocalizations we can assume that our signal x_t is periodic with period T .

As mentioned previously YIN is based on the autocorrelation of the signal as defined in Equation 7.

$$r_t(\tau) = \sum_{i=-\tau+1}^{t+\tau} x_i x_{i+\tau} \quad (7)$$

where $r_t(\tau)$ is the autocorrelation at lag τ calculated at time t and W is the integration window size.

We can also see that Equation 6 holds if we take the square and average over a window, W . This implies that a difference function can be formed where an unknown period may be found while searching for those values of τ for which the function is zero. The function is seen in Equation 8.

$$d_t(\tau) = \sum_{i=-\tau+1}^W (x_i - x_{i+\tau})^2 \quad (8)$$

One of the problems that the difference function creates is that it has the value of zero at zero lag and often times a non-zero value at the lag corresponding to the period due to imperfections in the periodicity. This indicates that the method will fail since it will always choose for the zero lag. In order to alleviate the above problem, the method proposes the use of the cumulative mean normalized difference function instead of the one in Equation 8. This new function is shown in Equation 9.

$$d_t'(\tau) = \begin{cases} 1, & \text{if } \tau = 0 \\ d_t(\tau) / [(1/\tau) \sum_{j=1}^{\tau} d_t(j)], & \text{otherwise} \end{cases} \quad (9)$$

This new function is actually one at zero lag and stays large at small lags.

There are several more steps that can be employed in order to ensure a better estimate and these steps can be seen in detail in [5]. Overall, the desired pitch can be obtained by picking the smallest value of the lag/pitch delay, τ that gives the minimum d' . An example of the YIN function for a specific call as well as the cepstral coefficients for the same call is shown in Figures 2-4.

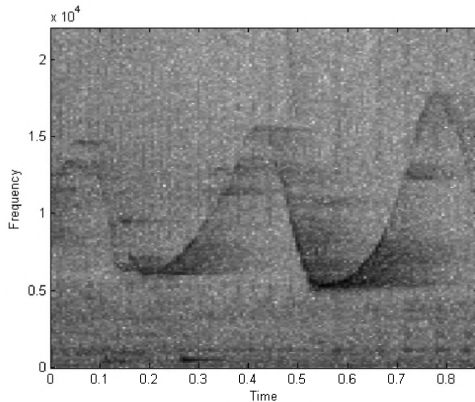


Figure 2: Whistle example

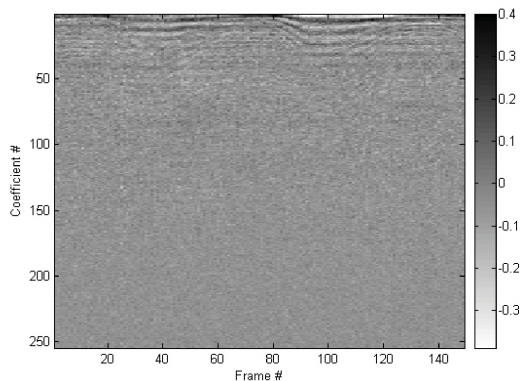


Figure 3: Cepstral coefficients

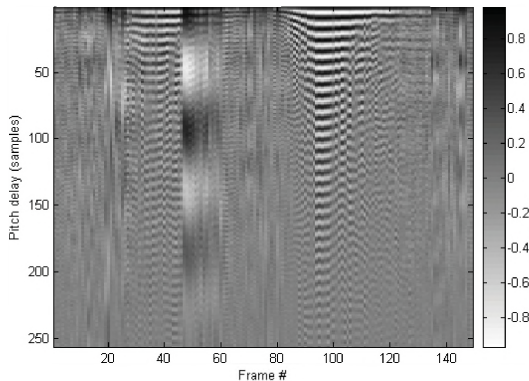


Figure 4: YIN coefficients

There are several more steps that can be employed in order to ensure a better estimate and these steps can be seen in detail in [5]. Overall, the desired pitch can be obtained by picking the smallest value of the lag/pitch delay, τ that gives the minimum d' . An example of the YIN function for a specific call as well as the cepstral coefficients for the same call is shown in Figures 2-4.

2.3 Get_f0: A software package for pitch extraction in speech

Get_f0 is one of the most popular pitch tracking algorithms. It is part of a widely used software package called Entropic Signal Processing Systems (ESPS) and Waves [8]. The majority of researchers in speech processing are familiar with this package. It is based on Doddington's and Secrest's 1983 algorithm [7] for pitch tracking in speech systems.

This method utilizes the linear prediction coding (LPC) residual error signal in order to extract the desired pitch candidate. LPC is based on the source filter model as seen for the cepstrum in section 2.1. This indicates that we theoretically expect that the residual signal will provide us with the excitation information.

To best alleviate some problems of high frequency noise, the authors devise and employ a de-emphasis filter as a pre-processing tool, whereby they low pass filter the residual signal in voiced regions of speech and high pass filter in unvoiced regions. These filters need to be redesigned for dolphin vocalizations.

To extract the candidate pitch at each instance the peaks of the normalized cross-correlation are acquired, Equation 10.

$$C(\tau) = \frac{\sum_{i=0}^{m-1} s(i) s(i-\tau)}{(\sum_{i=0}^{m-1} s(i) \sum_{i=0}^{m-1} s(i-\tau))^{1/2}} \quad (10)$$

Where τ is the lag and m is the number of samples to be correlated. As previously mentioned, the candidate pitch values are the lags at the peaks of $C(k)$ and the "goodness" measure is the corresponding value of $C(k)$ at those lags.

After having extracted the above values, dynamic programming [9] is employed in order to extract a smoother pitch contour. This requires some sort of penalty metric in order to decide what the best path amongst the candidates is. The cumulative penalty for each pitch candidate consists of a transition error in going from one frame to the next. This methodology provides a good pitch extractor specialized for speech.

3. EXPERIMENTAL RESULTS

In this section we provide the comparative experimental results as obtained from the methodologies described in section 2.

It is important to provide information on the data that was used for the experiments. Recordings from captive dolphins were obtained. From these recordings, whistles and bursts were manually extracted so that there would be no overlapping vocalizations. Overall, 110 calls were extracted of balanced type. These calls have a mean duration of 0.5sec and a mean SNR of 9.7dB. The low SNR was partly a result of analog to digital conversion given the lack of high precision hardware at the time of the recordings. The SNR was obtained by averaging the

peak SNR, Equation 10, at every frame, which was computed through the short-time autocorrelation function.

$$SNR = 10 \log_{10} \left(\frac{r(0)}{r(\tau=p)} \right) \quad (11)$$

Where $r(0)$ is the energy of the signal plus the noise and $r(\tau=p)$ is the energy of the signal with period at lag $\tau=p$.

To be better able to extract meaningful conclusions, ground truth was obtained by bootstrapping (semi-hand labeling for every frame). Initially, YIN was employed and then visually inspected in order to correct possible errors. Evidently, the extraction of ground truth allows for some errors due to resolution and rounding limitations given that we extract a pitch delay/lag for every frame. It is expected that such ground truth may incorporate some bias in the final results.

After obtaining the ground truth, the analysis of the data indicated an inherent bimodality. That led us to the choice of the hierarchy driven hidden Markov models for our approach. This is clearly shown in Figure 5. Two distinct frequency ranges are evident, thus allowing us to insert a decision level in the system. Arguably, one might explore the reasons for not choosing a single dynamic model/HMM for this task. In several experiments, a single system suffered from erroneous “doublings” and/or “halvings” at a per frame level caused by the fact that the cepstrum captures the presence of noise e.g. hardware, reflection noise.

Table 2 provides the average per frame accuracy for all three methods. It is worth noting that there are three different metrics in our results: Strict rate, which implies that the resulting pitch is an exact match with the ground truth, relaxed rate of ± 1 pitch delay (lag), and finally a relaxed rate of ± 2 pitch delays (lag). Basically, this implies a soft boundary or range of acceptable error. The relaxed rates correspond to an approximate 1.5% and 3% deviation from the ground truth, which in many applications could be acceptable. The same results are provided schematically in Figure 6. All results are generated using leave one out cross validation, otherwise known as round-robin.

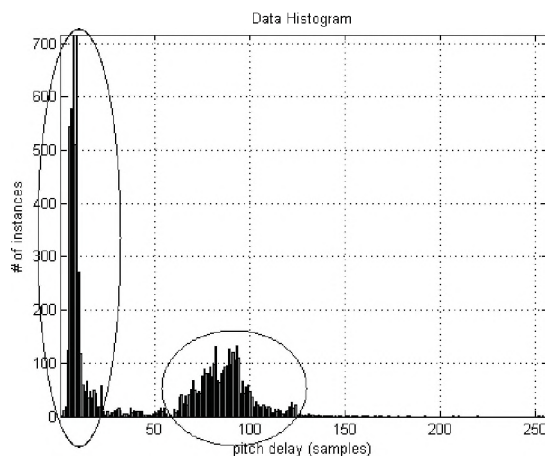


Figure 5: Data histogram. The two ellipses show the two modes of the data

Average per frame accuracy (%)		
HMM cepstrum	Yin	get f0
Strict Rate (%)		
66.12	47.09	29.3
Relaxed Rate ± 1 pitch delay (%)		
76.01	54.35	N/A
Relaxed Rate ± 2 pitch delay (%)		
77.9	55.11	N/A

Table 2: Comparative results

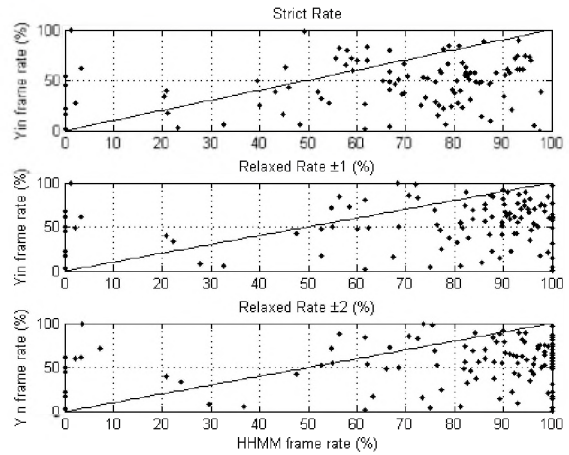


Figure 6: Comparative results of the YIN frame rate vs. the HHMM frame rate for every call

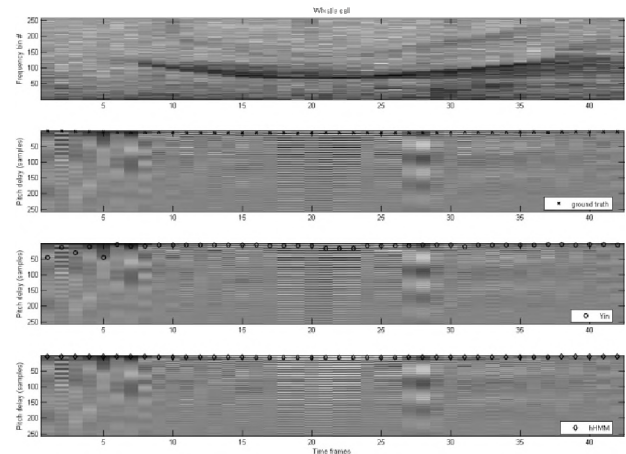


Figure 7: Example of successful pitch extraction using hierarchy HMM

In all cases it is apparent that our novel approach is superior to the baseline algorithms by over 10%. It is also interesting to note that get_f0 fails to give us comparable results for the relaxed rates due to the fact that it is highly tuned for human speech and is not able to track the desired pitch in dolphin vocalizations, which exhibit a much wider frequency range.

Furthermore, Figure 6 provides comparative results for each call for the novel approach of the hierarchy driven HMMs with the cepstral coefficients and the YIN algorithm. As it is clearly seen in the figures there is a shift of the points towards the right side of the plots. This

indicates that our methodology is superior and has a higher percentage of calls that are achieving above 80% frame accuracy.

In addition, an interesting fact arises from these plots. There appear to be a constant number of calls that are giving us a near 0% percent match. This discrepancy is caused due to the error that is introduced by the hierarchy. Basically, for these calls the decision of which frequency range they belong to is false, and thus the pitch extraction fails completely.

Lastly, Figures 7, 8 provide indicative examples of success and failure of the implemented algorithms. In both figures, the original spectrogram is shown and the comparative results are overlaid on the short time autocorrelation in order to provide a good visualization tool with results extracted from YIN.

In Figure 7, our implementation is closer to the ground truth where YIN actually exhibits a number of errors due to noise interference that leads to wrong peak picking.

In Figure 8, both YIN and the hierarchy driven HMM fail to extract the desired pitch again, due to the extremely low SNR as well as the ambiguity of the type and range of the call. In this case the evaluation stage of our system fails to classify this call in the correct range.

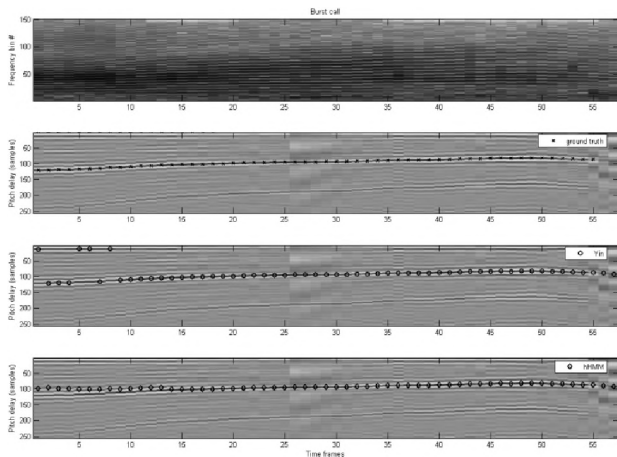


Figure 8: Example of fair pitch extraction using hierarchy HMM

4. CONCLUSIONS

As described in the previous sections, this work provides a comparative view on the success of three different pitch extraction algorithms for dolphin vocalizations. As evident from the experimental results presented in section 3, our novel approach of using hierarchy driven hidden Markov models with cepstral coefficients outperforms the other two popular methods in speech and music, YIN and `get_f0`.

The success of our approach is based on the idea of the hierarchy, which was implied from the nature of our data as seen in Figure 5. The existing bimodality allowed us to create two different HMM's with two different sets of states. This immediately reduced the state space

dimensionality of our system, thus minimizing the computational cost, while alleviating problems when training our model.

It is worth noting that the bimodality in the data needs to be explored in a larger body of data to extract meaningful conclusions in terms of the generic aspect of our method. Our data set needs to be enhanced so that we can extract possible biases from these specific recordings. Moreover, it would be interesting to compare the differences between recordings of captive dolphins versus dolphins in the wild.

Another reason for a larger labeled data set is to avoid pitfalls of over fitting when resorting to training testing methods such as leave one out cross validation.

Also, it should be noted that the hierarchy introduces an extra error term when it comes to deciding which range the call belongs to. Overall, this error accounts for only 4% of the total calls.

The use of the cepstral coefficients as a feature is considered a good match given that it showed a better descriptive feature than using the magnitude of the spectrum.

Also, it appears to be more resilient to noise. Furthermore, we can make assumptions about the location of the pitch peak, thus eliminating a number of coefficients and reducing the dimensionality of the feature space. This could also lead to a more computationally efficient algorithm.

Lastly, it is worth noting that YIN was far superior to `get_f0`. Its simplicity and efficiency make it a good candidate in simple cases. However, YIN is not so resilient to an increased noise level present in the recordings. Interestingly though, both YIN and our approach utilize dynamic programming/Viterbi, which could be an advantage to `get_f0`.

Overall, there are several steps that can be taken to improve the algorithm presented in this work. In summary, this paper clearly shows that a choice of good features and the use of a classifier which can be tuned according to a given data set can provide us with very satisfactory results for the task of pitch extraction.



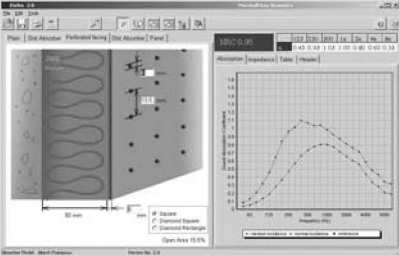

5. ACKNOWLEDGMENTS

The authors would like to thank Diana Reiss who provided us with the recordings of captive dolphins from her own collection. These recordings were hand-labeled by the authors and used for the experiments described in this work. Also, many thanks to Hervé Glotin and Olivier Adam for the French translation of the abstract.

6. REFERENCES

1. ISHMAEL, Integrated System for Holistic Multi-channel Acoustic Exploration and Localization, D. K. Mellinger, U.S. Department of Commerce, at <http://pml.noaa.gov/vents/acoustics/whales/ishmael/>

2. RAVEN, Interactive sound analysis software, Cornell Lab of Ornithology at:
<http://www.birds.cornell.edu/brp/raven/Raven.html>
3. R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification*, John Wiley & Sons, Inc., second edition 2001
4. L. R. Rabiner and B. H. Juang, "An Introduction to Hidden Markov models", IEEE ASSP Magazine, January, pp. 4-15, 1986
5. A.V. Oppenheim and R. W. Schaffer, *Digital Signal Processing*, Englewood Cliffs, NJ, Prentice-Hall, 1975
6. A. de Cheveigne and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music", *Journal of the Acoustical Society of America*, 111 (4), April 2002
7. B. Secrest and G. Doddington, "An integrated pitch tracking algorithm for speech systems", *Acoustics and Speech, and Signal Processing*, IEEE International conference on ICASSP'83, vol. 8, April 1983, pp. 1352-1355.
8. ESPS/WAVES, Entropic Signal Processing Systems, software package at:
<http://www.speech.kth.se/software/>
9. B. Gold and N. Morgan, *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*. John Wiley & Sons, Inc., New York, 1999.

	 <h1 style="margin: 0;">SoundPLAN</h1>
 <p>ZORBA is an easy to use software tool for predicting the sound absorption coefficients of porous materials such as fiberglass, mineral wool and polyester. It predicts both normal and random incidence absorption using simple input of physical parameters. ZORBA predicts the performance of perforated, slatted and panel absorbers. It estimates the absorption coefficients as well as acoustic impedance</p> <p>Trial Version: www.navcon.com/zorba.htm Navcon Engineering Network Phone: 714-441-3488 Email: forschner@navcon.com</p>	 <p>SoundPLAN is a graphics oriented noise prediction program used for noise planning, noise assessment & the development of noise mitigation measures. The database and management structure allows for a quick & easy generation of variants for small & complex noise models (i.e., Road & Railroad Projects, Industrial Plants, Quarry & Mines Operation, Power Plants, Amusement Parks, Wind Farms, Manufacturing Buildings/Rooms & Enclosures).</p> <p>SoundPLAN is based upon 30+ standards such as ISO 9613, Concawe, Nord2000, FHWA RD 77-108, TNM™2.5, FRA, VDI 3760. It generates traceable result tables and professional looking maps visualizing the input & output data. Noise Control & Optimization Tools include Noise Barrier Design and Industrial Noise Control Planning.</p> <p>Please visit us www.navcon.com/soundplan.htm for more information. Occasional users please check out SoundPLAN essential www.navcon.com/soundplan_essential.htm</p>

ATTRACTIVE TIME-VARIANT ORTHOGONAL SCHUR-LIKE REPRESENTATION FOR CLICK-TYPE SIGNAL RECOGNITION

Maciej Lopatka^{‡,†}, Olivier Adam[‡], Jean-François Motsch[‡] and Jan Zarzycki[†]

[†]Wrocław University of Technology I-28 Signal Theory Section
Build. C5 W. Wyspińskiego 27 Poland Wrocław 50370

[‡]Université Paris 12 Laboratoire Images Signaux et Systèmes Intelligents (Lissi)
Ingénierie des Signaux Neurosensoriels, 61 Av. de Gaulle 94000 Créteil, France

ABSTRACT

Analysis of click-type signals in the presence of noise with time-varying statistics is a challenging task, especially in low signal-to-noise ratio conditions. This well-known problem is commonly present in underwater passive acoustics applications. In this paper we present a novel solution for this dilemma as applied to marine mammal acoustics - a well-established basis for marine mammal study and protection. The adaptive orthogonal Schur-like algorithm is proposed to classify medium-frequency odontocete clicks. This technique is characterized by excellent convergence behaviour, very fast parametric tracking capability and robustness. The difficulty of recognition (classification) resides in the extraction of the signal's intrinsic information; i.e. extraction of an efficient signal signature. It is expected that the distances between the signatures within the class are minimal (small intra-class variance) and between the classes are maximal (high inter-class variance). This condition ensures a good recognition performance (separability of classes). The 2D signature proposed in this work and based on a selected set of the time-varying Schur coefficients assures this requirement. When compared to the classical Fourier approach, the proposed recognition method is four times as efficient for inter-class distances and twice as efficient for intra-class distances. The results of the recognition are given for sperm whale (*Physeter macrocephalus*) regular clicks and striped dolphin (*Stenella coeruleoalba*) *nacchere* clicks. They are very satisfactory and promising for other applications. The proposed technique can be easily implemented in real-time applications such as underwater acoustic monitoring systems.

RESUME

Les analyses des signaux du type cliquetis noyés dans un bruit dont les statistiques sont temps-variant est un challenge, surtout dans des conditions de rapports signal-sur-bruit défavorables. Cette problématique largement connue est couramment présente dans des applications de l'acoustique passive sous-marine. Dans cet article, nous présentons une solution novatrice appliquée dans le domaine de l'acoustique des cétacés qui actuellement constitue une base bien établie de l'étude et la protection des mammifères marins. L'algorithme orthogonal adaptatif de Schur est proposé pour classifier des clics de 2 espèces d'odontocètes. La technique introduite est caractérisée par une excellente convergence, un très bon suivi des paramètres et est robuste au bruit. Les difficultés de reconnaissance (classification) résident dans l'extraction de l'information intrinsèque du signal i.e. la mise en forme d'une signature efficace du signal. Il est attendu que les distances entre les signatures de la même classe soient minimales (petite variance intra-classe) et pour les différentes classes soient maximales (grande variance inter-classe). Cette condition assure de bonnes performances de reconnaissance (séparation des classes). La signature bidimensionnelle proposée dans ce travail et basée sur un ensemble sélectionné des coefficients temps-variant de Schur assure cette exigence. En comparant cette méthode avec l'approche classique de Fourier le gain d'efficacité est multiplié par 4 pour les distances inter-classe et par deux pour les distances intra-classe. Les résultats de la reconnaissance sont donnés pour les clics usuels de cachalots (*Physeter macrocephalus*) et les clics du type *nacchere* de dauphins bleus et blancs (*Stenella coeruleoalba*). Ils sont très satisfaisants et promettant pour d'autres applications. La technique proposée peut être facilement implémentée dans des applications temps-réel telles que des systèmes acoustiques de surveillance sous-marine.

1. INTRODUCTION

The click-type signal is characterized by short duration (microseconds to milliseconds), wide bandwidth (quasi flat spectrum), and is generally far from stationary. The

processing of such a signal is a complex and challenging task, especially in low signal-to-noise ratio conditions. This becomes more difficult when the statistics of the background noise are time-varying. Click-type signal analysis requires signal processing techniques that fulfil the

following principal conditions: robustness (to non-stationary noise), good time resolution (a click can last from tens to several hundred samples), efficient extraction of the signal's intrinsic characteristics (for detection, recognition, etc. purposes). A variety of methods are used for automatic recognition of transient signals. Some of them employ time and/or frequency representations. Other methods transform the signal to another space. These representations can be used for classification (e.g. template matching). Different parameters can be extracted for statistical classification. Many different classical (Fourier transform and its derivatives, parametric filters, time domain statistics) and advanced techniques (wavelets, Hilbert Huang Transform, High Order Statistics) are commonly used in processing of non-stationary brief signals, although not all are well suited for such processing. Classical temporal techniques use parameters such as duration, mean, variance, energy, amplitude, instantaneous phase, zero crossing rate or moments [16]. The bio-acoustics community widely uses Fourier based techniques and parameters such as principal frequency, bandwidth, cepstral coefficients or variations of the frequency or of the phase [17, 18]. Time-frequency representations are also used for signal description and classification [13]. The comparison of AutoRegressive (AR) modelling and the wavelet transforms as feature extraction tools is given in [19]. The use of neural networks for underwater signal processing is proposed in [20]. The chosen technique depends on the application and other factors such as implementation or budget issues. For example, for acoustic monitoring systems, real-time processing is paramount. Therefore it is expected that complex and time consuming methods would not be used, though there may be deterioration in performance.

In this paper we introduce the adaptive orthogonal Schur-like parameterization, a novel technique for analysis of brief acoustic signals. The adaptive Schur algorithm as shown in this paper is a powerful, low complexity technique that is also very easy to implement. A first study of this technique as applied to underwater passive acoustics is presented in [1]. This technique has already been applied to detection and analysis of sperm whale clicks [2]. This paper is an endeavour to classify mid-frequency marine mammal clicks.

The adaptive Schur algorithm is composed of two steps. First, recordings are analyzed to extract all non-stationary transients (detection of clicks) [2,7]. Secondly, the extracted clicks are assigned to different classes (recognition of clicks) [7]. We introduce a click-type signature that is based on a selected set of the time-varying Schur coefficients. The objective of this study is to recognize (classify) broadband acoustic transients emitted by two odontocete species, the sperm whale (*Physeter macrocephalus*) and the striped dolphin (*Stenella coeruleoalba*). The sperm whale regular clicks [3-5] and striped dolphin *nacchere* clicks [6] have very similar time and frequency characteristics; i.e. duration of a few milliseconds and a wide bandwidth [7]. These two

odontocete species were chosen because they seem to represent the most difficult scenario for marine mammal click-type calls: similar duration and frequency bandwidths that overlap by more than 90% [7].

The clicks considered here are emitted in sequences. The principal parameter characterizing the sequence of clicks is the ICI (inter-click interval). This slow time-varying parameter defines the time distances between consecutive clicks within the sequence of clicks. Therefore, the recognition of such clicks can be carried out in two ways: by a global and a local approach. In this paper we consider the latter approach, which means that the classification is performed on every single click. This method is much more challenging than the global approach. In the global approach, the distinction between sperm whale regular clicks and *nacchere* striped dolphin clicks can be performed by exploiting the ICI, which is about 0.5-2 s for the sperm whale and about 0.1 s for striped dolphin. The problem appears when clicks are missing or when different click sequences overlap, making estimation of the ICI very complicated. The local classification approach applied to a sequence of clicks can be considered as a pre-processing step to the global classification approach (support for the ICI estimation).

In this paper, we give results of the recognition obtained on sperm whale regular clicks (called *Pma* clicks) and striped dolphin *nacchere* clicks (called *Sco* clicks). We discuss the performance and present perspectives.

2. MATERIAL AND METHODS

Sperm whale clicks were recorded in the canyon of Toulon (Mediterranean Sea, France) (42°58'N, 5°51'E, 42°39'N, 5°43'E, 42°39'N, 6°30'E, 42°58'N, 6°27'E) in August 2004 [8]. The recordings of striped dolphin clicks were made in the Ligurian Sea in 2002 [6]. Both recordings were performed with the omnidirectional hydrophone (0-30 kHz) towed at a depth of about 50-100 m. The acoustic signals were recorded with a 44.1 kHz sample rate and 16 bit resolution via a commercial audio PC card.

2.1 Adaptive Schur Algorithm

The adaptive Schur-like parameterization [9,10] was proposed for the recognition of the bio-acoustic clicks emitted by sperm whales and striped dolphins. More detailed discussion of this technique as applied to short-term stochastic signal processing is given in [7].

The adaptive Schur algorithm, also called the innovations filter or whitening filter, is in fact an optimal orthogonal linear prediction filter. At every time-instant the filter calculates an optimal value of the signal at instant t taking into account all its past values. The solution of the prediction is calculated from the orthogonal projection of the current signal on its past samples. The forgetting factor is introduced to weigh the past samples [7]. The filter is

always stable numerically because all the signals within the filter are normalized to unity.

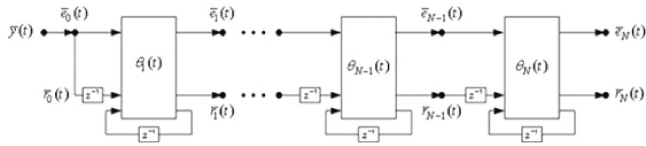


Figure 1. Adaptive orthogonal Schur filter

The ladder-form orthogonal filter is adaptive (in time) and recursive (in order) (fig.1, eq. (1)-(2)). The filter is composed of N identical sections $\theta_n(t)$ (the number defines the filter order) which are completely defined with second-order statistics by the time-varying Schur coefficients $\rho_n(t)$ (also called the reflection coefficients). The adaptive Schur algorithm is defined by the three following equations:

$$\begin{cases} \rho(n+1,t) = \rho(n+1,t-1)AB - e(n,t)r(n,t-1) \\ e(n+1,t) = CB^{-1} [e(n,t) + \rho(n+1,t)r(n,t-1)] \\ r(n+1,t) = CA^{-1} [\rho(n+1,t)e(n,t) + r(n,t-1)] \end{cases} \quad (1)$$

where A, B and C are as follows:

$$\begin{cases} A = (1 - e^2(n,t))^{1/2} \\ B = (1 - r^2(n,t-1))^{1/2} \\ C = (1 - \rho^2(n+1,t))^{-1/2} \end{cases} \quad (2)$$

The variables $\rho(n,t)$, $e(n,t)$ and $r(n,t)$ denote respectively the time-varying Schur coefficient, the normalized forward prediction error and the normalized backward prediction error on the n th section at the time t . The requisite number of sections depends on the signal type. This is closely linked to the signal energy distribution on the filter sections. As it was demonstrated in [7], the energy on the filter sections globally decreases as the number of sections increases. In practice the order of the filter is chosen between 10 and 20.

The signal $y(t)_{t \in \{1, \dots, T\}}$ input to the filter is transformed into the 2D matrix $\Theta_{NxT} = [\theta_1, \dots, \theta_N]$ (see fig.1):

$$\Theta_{NxT} = \begin{bmatrix} \rho(1,1) & \rho(1,2) & \dots & \rho(1,T) \\ \rho(2,1) & \rho(2,2) & \dots & \rho(2,T) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(N,1) & \rho(N,2) & \dots & \rho(N,T) \end{bmatrix} \quad (3)$$

The matrix columns represent time and the rows represent order. In our work real-time processing is essential. Therefore we present the mathematical complexity of the algorithm in table 1. We also estimated the algorithm processing time for a 1 second signal sampled at 44.1 kHz as a function of the filter order (commercial

PC: processor 1.6 GHz, RAM memory 1 GB). This relation is given in fig.2.

Table 1 – Mathematical complexity of the algorithm (number and weight of mathematical operations for time and order loops)

Operation	Number of cycles According to IEEE Standard 754	Number of operations Order loop n	Number of operations Time loop t
+ or -	1	6	3
*	2	12	5
$\sqrt{\quad}$	5	3	1
/	5	3	1

The adaptive Schur algorithm has two loops: the major loop in time t and the minor loop in order n . The mathematical complexity of the algorithm for the minor loop is $O(N)$ and for the major loop is $O(T*N)$. Due to recursive and adaptive processing the complexity is linear, which is very attractive for practical implementation.

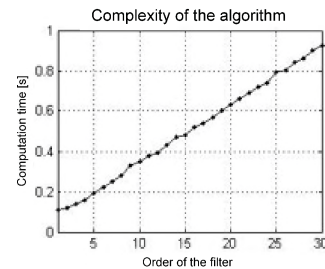


Figure 2. Computational complexity of the technique

This signal analysis is based on the matrix of time-varying Schur coefficients:

$$\Theta(n,t) = \{\rho(n,t) : n \in \{1, \dots, N\}, t \in \{1, \dots, T\}\} \quad (4)$$

which reflect the second-order statistics of the filtered signal. They gravitate towards their optimal values when the signal is (quasi) stationary. When there is an important variation in the signal covariance, the time-varying Schur coefficients reflect these changes.

2.2 Recognition of click-type signals

Processing of the click-type signal is especially challenging due to its short duration and wide bandwidth. Classical methods for click-type signals analysis have difficulty capturing the signal's intrinsic information. There is a need for new techniques that are better suited for this task. Here, our method is appropriate for transient signal recognition. The analysis is based on the 2D Schur-like representation i.e. the set of time-varying Schur coefficients.

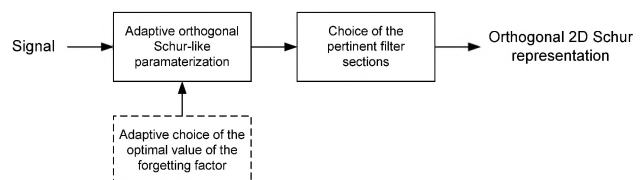


Figure 3. Generation of 2D orthogonal (Schur-like) signature

The recognition problem posed in this work is supervised, i.e. we use a specific set of click-type signal representatives for each class (sperm whale regular clicks (“clicks Pma”) and striped dolphin *nacchere* clicks (“clicks Sco”). The proposed signature (pattern Sc_{TV} – Schur time-variant) of the click-type signal based on time-varying Schur coefficients is successful for discrimination (high inter-class variance) and invariance (low intra-class variance).

The signatures Sc_{TV} for both classes are calculated according to the method presented in fig.3 and are shown in fig.4. A random set of clicks for each class (sperm whale and striped dolphin) input to the adaptive Schur filter results in a set of the time-varying Schur coefficients (2D representation, see eq. (3) and fig.4). The signature Sc_{TV} is calculated as the mean of a set of patterns for that class. In our study we used 50 clicks chosen randomly from datasets. These sets are used for determining the most discriminating K Schur coefficients. First, for the two classes Pma and Sco , we calculate the vector of discrimination H_{1xN} :

$$\forall_{i=1\dots N} \mathbf{H}_{1xN}(i) = \left\| \rho_i^{Pma} - \rho_i^{Sco} \right\| \quad (5)$$

and with:

$$H(1) > \dots > H(j) > \dots > H(N) \xrightarrow{\text{choice}} \max(H)_{1xK} \quad (6)$$

Finally, we conserve K of the most unlike (between classes) Schur coefficients, which guarantee very good class separability. The signatures Sc_{TV} are calculated for three different frequency bands: low (LF, 1-4kHz), medium (MF, 7-10 kHz) and high frequency (HF, 12-16 kHz) bands. The signatures Sc_{TV} for sperm whale regular clicks (Pma) and striped dolphin *nacchere* clicks (Sco) for each of the three bandwidths are presented in fig.4.

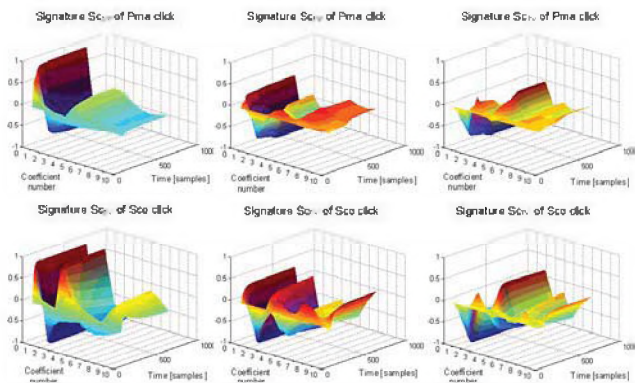


Figure 4. Signatures Sc_{TV} (2D) of Pma and Sco clicks
First line is for Pma clicks and second line is for Sco clicks
The signature Sc_{TV} is given for three different frequency bands (in columns): LF (1-4 kHz), MF (7-10 kHz) and HF (12-16 kHz)

For comparison purposes we decided to also evaluate the performance of a widely used classical recognition technique based on the Fourier technique. The Fourier signature is given as a set of 32 Fourier coefficients. The number of Fourier coefficients was chosen to capture the global spectral structure of the signal, and not local changes, which can be influenced by noise or propagation effects.

The signal description (recognition) aims to obtain the signature (pattern) that most effectively represents the signal. Ipso facto, it is expected to reach a high discrimination between classes and a high invariance of the signature within each class. In this work we proposed two supervised classification approaches:

- template matching,
- statistical.

For the first approach we use four different dissimilarity metrics: correlation coefficients, and Euclidian (d_E), Chebyshev (d_{Ch}) and Minkowski (d_M) distances, which are defined as follows (for two signals x and y):

$$d_E = \left\{ \sum_{i=1}^p (x_i - y_i)^2 \right\}^{1/2} \quad (7)$$

$$d_{Ch} = \max_i |x_i - y_i| \quad (8)$$

$$d_M = \left\{ \sum_{i=1}^p (x_i - y_i)^m \right\}^{1/m} \quad (9)$$

In the statistical approach, we proposed three parameters (variables v_1 , v_2 and v_3) calculated from the 2D Schur representations (fig.4):

$$v_1 = \rho_2(T) - \rho_3(T) \quad (10)$$

$$v_2 = \sum_{j \in \{7,8,9\}} \left(\frac{1}{T} \sum_{i=1}^T i \rho_j^*(i) \right), \quad \rho_j^* = \rho_j - \min(\rho_j(i)) \quad (11)$$

$$v_3 = \frac{1}{N} \sum_{j=1}^N \sum_{i=1}^T \rho_j^2(i) \quad (12)$$

These variables allow an almost perfect discrimination between sperm whale regular clicks and striped dolphin *nacchere* clicks. They were chosen *a posteriori* based on our two datasets.

3. RESULTS AND COMMENTS

The performance of the click-type signal recognition is obtained from two odontocete calls: sperm whale regular clicks and striped dolphin *nacchere* clicks. We present the similarities between these two categories of clicks in time and frequency domains for LF, MF and HF bandwidths (see table 2).

Table 2 – Correlation results between sperm whale (Pma) clicks and striped dolphin (Sco) clicks in time and frequency domains for four LF (1-4 kHz), MF (7-10 kHz), HF (12-16 kHz) bandwidths and wideband (whole frequency range).

	LF band	MF band	HF band	Wideband
Correlation : click Pma – click Pma				
Time	0.421±0.12	0.411±0.10	0.391±0.12	0.264±0.09
Frequency	0.709±0.13	0.744±0.11	0.722±0.11	0.642±0.16
Correlation : click Sco – click Sco				
Time	0.459±0.14	0.316±0.7	0.29±0.08	0.368±0.12
Frequency	0.740±0.15	0.671±0.14	0.684±0.10	0.741±0.15
Correlation : click Pma – click Sco				
Time	0.311±0.09	0.320±0.07	0.318±0.07	0.201±0.06
Frequency	0.573±0.14	0.660±0.13	0.694±0.11	0.501±0.10

We note that neither in time nor frequency domains is it possible to propose a threshold for distinguishing these classes. First, this is due to the fact that both clicks classes have high variance within-class (we obtain low values of the correlation between clicks of the same species). This diversity results from the natural intrinsic richness of the clicks and from propagation effects. Secondly, the *Pma* and *Sco* clicks are very similar in duration and frequency band, and thus the values of the correlation between clicks of the two species are significant.

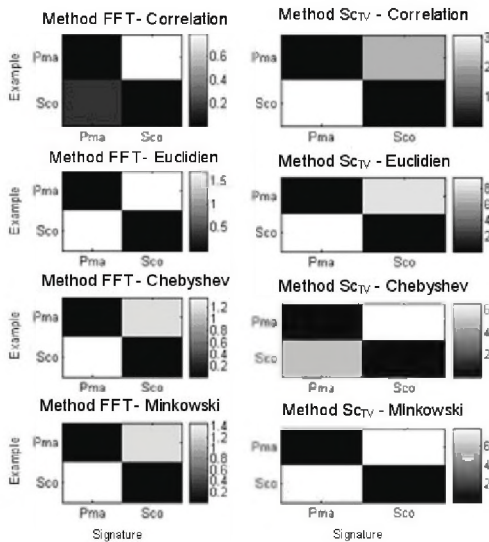


Figure 5. Classification performance for Pma and Sco clicks (mean values of intra- and inter- class distances in dB) The results are given for LF band (Template Matching)

The performance of classification by the template matching approach is given in fig.5-6. The patterns of each class are compared to the signatures *Pma* and *Sco*. The mean values of these intra-class and inter-class distances are given in fig.5. These values are normalized for each class to 0 dB for the intra-class distances. This means also that the inter-class and intra-class distances should be minimal. The distribution of values of the inter-class and intra-class distances is given in fig.6 (for the Minkowski metric). We note that the proposed signature *ScTV* ensures lower intra-class distances and higher inter-class distances, which results in a much improved discrimination performance. When compared to the performance of the

Fourier based recognition technique, the proposed method is four times more efficient for inter-class distances, and twice as efficient for intra-class distances (see fig.6). The separability of clicks, which was impossible in the time and frequency domain, becomes attainable in the space of the time-varying Schur coefficients.

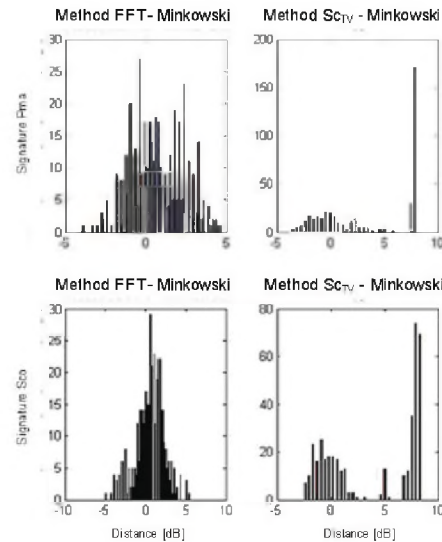


Figure 6. Histogram of intra- and inter- class distances for the Minkowski metric (Template Matching)

The parameters proposed in eq. (10)–(12) allow the accurate discrimination between sperm whale regular clicks and striped dolphin *nacchere* clicks (see fig.7). However, we note that these variables for sperm whale clicks (black in fig.7) are somehow correlated. This can be attributed to different diving phases of sperm whales. The classification results depend also from the performance of the data acquisition. This requires further research and analysis.

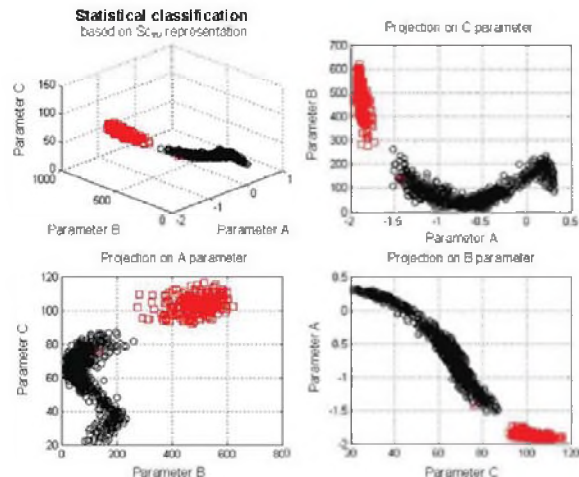


Figure 7. Statistical classification of Pma (black) and Sco clicks (red) in 3D representation space (parameters A, B and C correspond to variables v_1 , v_2 and v_3)

The correlation results shown in table 2 and the classification results based on the Fourier signatures (fig. 5) compared to the performance of the recognition method

proposed in this paper let conclude that the two proposed recognition approaches, i.e. the template matching and the statistical classification based on the 2D orthogonal Schur-like representation, are very efficient and robust for underwater click-type signal analysis.

4. CONCLUSION

In this paper we presented a novel click-type signal recognition method based on the time-variant Schur algorithm. This orthogonal technique appears well suited for underwater signal processing. The adaptive and recursive nature of the proposed algorithm is very attractive for real-time processing. We proposed an efficient 2D signature for click-type signals. We evaluated our recognition method on sperm whale (*Physeter macrocephalus*) regular clicks and striped dolphin (*Stenella coeruleoalba*) nacchere clicks. These two species clicks present some common characteristics that make classification quite challenging, especially for the classifier based on the Fourier transform. The recognition results showed that concerning classification performance and resistance to noise, the 2D Schur signature is considerably more efficient than the classical Fourier descriptor. Moreover, this signature is more compact and is characterized by a lower variability. Motivated by very promising results obtained from this study, we would like to investigate the proposed recognition approach on other marine mammal click-type and chirp-type calls. We are currently working on the issue of independence of the recognition algorithm from acquisition system set-up.


5. ACKNOWLEDGMENTS

We would like to thank the Association DIRAC (France) for the financial support of this work. We also thank Dr. Gianni Pavan for providing striped dolphin acoustic recordings used in this work.


6. REFERENCES

- [1] Lopatka M., Adam O., Laplanche C., Motsch JF., Zarzycki J. New effective analytic representation based on the time-varying Schur coefficients for underwater signals analysis. Proc. of the IEEE/OES Oceans'05 Europe Conference, 2005
- [2] Lopatka M., Adam O., Laplanche C., Zarzycki J., Motsch J.F. Sperm whale clicks analysis using recursive time-variant lattice filters. Applied Acoustics 67 (11-12): 1118-1133, Special Issue on Detection and localization of marine mammals using passive acoustics, Ed. Elsevier, 2006
- [3] Mohl B., Wahlberg M., Madsen P.T., Miller L.A. Sperm whale clicks: directionality and source level revisited. J. Acoust. Soc. Am. Vol. 107: 638-648, 2000
- [4] Goold J.C., Jones S.E. Time and frequency domain characteristics of sperm whale clicks. J. Acoust. Soc. Am. Vol. 98 (3): 1279-1291, 1995
- [5] Zimmer W. M. X., Tyack P. L., Johnson M. P., Madsen P. T. Three-dimensional pattern of regular sperm whale clicks confirms bent-horn hypothesis. J. Acoust. Soc. Am., 117 (3): 1473-1485, 2005
- [6] Pavan G., Fossati C., Manghi M., Priano M. Nacchere: an acoustic behavior of striped dolphins. European Research on Cetaceans: 17, 2005
- [7] Lopatka M. Détection et reconnaissance des signaux stochastiques transitoires. Application à l'identification des mammifères marins. Ph.D. Thesis. Université Paris 12 (France) and Wrocław University of Technology (Poland), 2007
- [8] Adam O., Laplanche C., Lopatka M., Gandilhon N., Girou E., Arquier R., Motsch JF. Detection and localization of a sperm whale observed using passive acoustics off the coast of Toulon. 19th annual conference of the European Cetacean Society. Poster A2. La Rochelle, France, 2005
- [9] Kailath T. A theorem of I. Schur and its impact on modern signal processing. In I. Schur methods in operator theory and signal processing, Ed. I. Gohberg, Operator theory: advances and applications, Vol. 18, Birkhäuser-Verlag: 9-30, 1986
- [10] Lee D.T.L., Morf M., Friedlander B. Recursive Least Squares Ladder Estimation Algorithms. IEEE Trans. on Circuits and Systems, Vol. CAS-28, No. 6: 467-481, 1981
- [11] Marple L., Brotherton T. Detection and classification of short duration underwater acoustic signals by Prony's method. Proc. of the ICASSP'91 : 1309-1312, 1991
- [12] Huynh Q.Q., Greene W., Impagliazzo J. Feature extraction and classification of underwater acoustic signals. In Full Field Inversions Methods in Ocean and Seismo-Acoustics (Diachok O., Caiti A., Gerstoft P., Schimdt H., eds) : 183-188, Kluwer. Dordrecht, The Netherlands, 1995
- [13] Leprette B., Martin N. Extraction of pertinent subsets from time-frequency representations for detection and recognition purposes. Signal Processing, Vol. 82 (2): 229-238, 2002
- [14] Loe S., Anderson K., Jung K. Comparative analysis results for underwater transient classification. Proc of the SPIE 2242 (Wavelet Applications): 815-823, 1994
- [15] Ghosh J., Deuser L., Beck S. A neural network based hybrid system for detection, characterization and classification of short-duration oceanic signals. IEEE Journal of Oceanic Engineering, Vol. 17: 351-363, 1992
- [16] Jin J., Shi J. Automatic feature extraction of waveform signals for in-process diagnostic performance improvement. J. Intelligent Manufacturing 12: 257-268, 2001

- [17] Houser D.S., Helweg D.A. and Moore P.W. Classification of dolphin echolocation clicks by energy and frequency distributions. *J Acoust. Soc. Am.* 106 (3): 1579-1585, 1999
- [18] Bozkurt B., Couvreur L. On the use of phase information for speech recognition. *Proc. of the 13th European Signal Proc. Conf. (EUSIPCO 2005)*, Antalya, Turkey, 2005
- [19] Fargues M.P., Bennett R. Comparing wavelet transform and AR modelling as feature extraction tools for underwater signal classification. *29th Asilomar Conference on Signals, Systems and Computers*, Vol. 2: 915-919, 1995
- [20] Ghosh J., Deuser L., Beck S. A neural network based hybrid system for detection, characterization and classification of short-duration oceanic signals. *IEEE Journal of Oceanic Engineering*, Vol. 17: 351-363, 1992



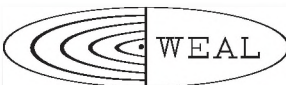
**WESTERN
ELECTRO-ACOUSTIC
LABORATORY**

A division of Veneklasen Associates, Inc. 

ACOUSTICAL TESTING & MEASUREMENTS

Laboratory Testing	
Sound Transmission Loss, STC	ASTM E-90* (ISO 140*)
Sound Absorption, NRC	ASTM C-423* (ISO 354*)
Calibration of Microphones	ANSI SI-10*
Acoustic Power	ANSI S12-32
<i>Full Anechoic Chamber Measurements also available</i>	
Field Testing	
Noise Reduction, NIC, FSTC	ASTM E-336*
Impact Sound Transmission, FIIC	ASTM E-1007*
Building Facades	ASTM E-996*

*NVLAP Accredited



TEL: 661.775.3741 FAX: 661.775.3742
25132 Rye Canyon Loop Santa Clarita, CA 91355
www.weal.com

ACOUSTIC BEHAVIOR OF DOLPHINS IN THE PACIFIC OCEAN: IMPLICATIONS FOR USING PASSIVE ACOUSTIC METHODS FOR POPULATION STUDIES

Shannon Rankin¹, Julie N. Oswald², and Jay Barlow¹

¹Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, California 92037

²Scripps Institution of Oceanography, University of California San Diego, La Jolla, California, 92038

ABSTRACT

The Southwest Fisheries Science Center has been conducting shipboard visual line-transect cetacean surveys for over 30 years, and combined visual and acoustic surveys for seven years. Full incorporation of passive acoustics as a tool for population assessment requires an understanding of the acoustic behavior of cetaceans as well as the limitations of the methods used in these surveys. Our research summarizes data collected during seven years of combined visual and acoustic surveys throughout the central and eastern North Pacific Ocean, ranging from the Aleutian Island chain in the north, to Peru in the south. Phonations from 2,034 dolphin schools were examined to better understand the acoustic behavior of cetaceans. Equally important are the cetacean schools that were seen but not heard, and this analysis includes an examination of these groups by species, group size, geographic location, and time of day. The results of this analysis allow us to take the first steps to incorporate passive acoustics into line-transect cetacean surveys.

RÉSUMÉ

Le Southwest Fisheries Science Center a étudié les cétacés à bord de navires en utilisant des transects linéaires pour des données visuelles depuis plus de 30 ans, et une combinaison des méthodes visuelles et acoustiques depuis seulement sept ans. L'incorporation complète de l'acoustique passif comme outil d'évaluation de la population exige une bonne compréhension du comportement vocal des cétacés, ainsi que de connaître les limites des méthodes utilisées dans ces études. La présente recherche résume les données provenant de sept années d'études visuelles et acoustiques tout au long de la partie centrale et orientale du Pacifique Nord, depuis les îles septentrionales d'Aleutian, jusqu'au Pérou, au sud. Les vocalisations des dauphins, à partir de 2034 groupes suivis, ont été examinées afin de mieux comprendre le comportement vocal des cétacés. Les groupes de cétacés qui ont été vus mais non entendus, sont également importants ; cette analyse examine ces groupes par espèce, taille du groupe, position géographique et heure du jour. Les résultats nous permettent de prendre en compte les premières mesures pour incorporer l'acoustique passif dans un transect linéaire dans l'étude des cétacés.

1. INTRODUCTION

Population studies of cetaceans in offshore waters have typically relied on shipboard visual observations, which are limited to daylight hours and must be suspended when poor weather conditions prohibit reasonable visual detection of animals. In recent years, passive acoustic detection of cetacean phonations using towed hydrophone arrays has been used to complement visual shipboard surveys (Thomas *et al.* 1986, Gordon *et al.* 2000, Oswald *et al.* 2007a). Acoustic detection of cetacean phonations is not limited by time of day, nor is it affected by most weather conditions. The primary limitation of acoustic methods is that the animals must be producing sounds within the frequency range of the equipment.

Dolphin phonations have been grouped into three categories: whistles, burst pulses, and echolocation clicks. Whistles are tonal, frequency-modulated signals used for communication (Janik and Slater 1998, Herzing 2000, Lammers *et al.* 2003). Most dolphin species produce

whistles, which typically have fundamental frequencies between 2 and 30 kHz (Lammers *et al.* 2003, Oswald *et al.* 2004). Burst pulses are broadband click trains that have very short inter-pulse intervals. These sounds are also thought to be used for communication, although they may also be for echolocation (Herzing 2000). Echolocation clicks are short, broadband, pulsed sounds used for navigation and object detection. Echolocation clicks have peak frequencies ranging from tens of kilohertz to well over 100 kHz (Au 1980, Au 1993). Basic descriptions of acoustic repertoire exist for many species; however, little is known of the acoustic behavior of most species in their natural habitat.

The Southwest Fisheries Science Center (SWFSC) has been conducting visual observations of cetaceans during shipboard line-transect surveys for over thirty years. In 2000-2006, a towed hydrophone array was added to examine the potential for the use of passive acoustics during these surveys. In this paper we present a preliminary examination of the acoustic behavior of dolphins in the

Pacific Ocean and our ability to detect their phonations using a towed hydrophone array with a limited bandwidth (2 - 24 kHz).

2. METHODS

We conducted cetacean surveys in the Pacific Ocean from 2000 to 2006 using simultaneous visual and acoustic line-transect methods. The acoustic effort during these surveys is shown in Figure 1. The dates, study area, and effort for each survey varied, and a summary of this information is given in Table 1.

Visual observation methods followed standard SWFSC protocol that has been used since the 1980s (Kinzey *et al.* 2000). A team of three experienced visual observers rotated

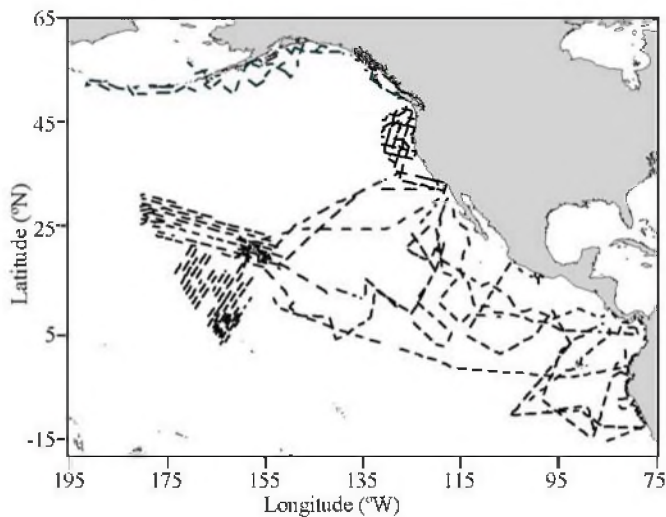


Figure 1. Map of survey area and tracklines with passive acoustic effort using a towed hydrophone array shown as dashed lines.

between two ‘big-eye’ 25x150 binoculars and one data-recording position. Visual observation occurred during daylight hours in Beaufort sea states 0-5. When animals were sighted by the visual observation team, they were approached for species identification and group size estimation.

A towed hydrophone array was used for acoustic detection of cetacean phonations. The array was typically towed 200-300 m behind the ship during daylight hours and in sea states less than Beaufort 7. Several array configurations were used, each with its own specifications. The five-element ‘Sonatech’ array (Sonatech, Inc., Santa Barbara) had a flat frequency response from 2 kHz to 45 kHz (± 4 dB at -132 dB re 1 V/ μ Pa), the three-element high-frequency ‘HF’ array (Sonatech, Inc., Santa Barbara) had a flat frequency response from 2 kHz to 120 kHz (± 3 dB at -164 dB re 1V/ μ Pa), and the ‘SWFSC’ array had a flat frequency response from 500 Hz to 30 kHz (± 5 dB at -155 dB re 1V/ μ Pa). The specific arrays used during each survey are shown in Table 1.

Signals from the array were equalized using a Mackie CR1604-VLZ mixer and recorded using a Tascam DA-38 eight-channel digital recorder (sample rate 48 kHz). Sounds were monitored by an acoustic technician both aurally, using headphones, and visually, using real-time scrolling spectrographic software (ISHMAEL, Mellinger 2001). Acoustic localization of dolphin schools was performed based on the convergence of bearing angles plotted on Whaltrak, a custom-written plotting program. Bearing angles to phonating dolphin schools were calculated using the phone-pair bearing algorithm in ISHMAEL (Mellinger 2001). All data presented here are based on monitoring within the limitations of the hydrophones and recording equipment; only sounds detected between 2 kHz and 24 kHz were included in the analyses.

Table 1. Summary information for seven cetacean surveys conducted by the Southwest Fisheries Science Center, including the cruise name, dates, region surveyed, survey vessel, hydrophone arrays used, and the number of acoustic detections. Three surveys were conducted in the eastern tropical Pacific Ocean (ETP).

Cruise Name	Dates	Region Surveyed	Survey Vessel	Array	# Detections
STAR	28 July - 9 Dec, 2000	ETP	<i>McArthur</i>	Sonatech, HF	374
ORCAWALE	30 July - 9 Nov, 2001	US West Coast	<i>Jordan</i>	Sonatech, HF	132
HICEAS	27 July - 8 Dec, 2002	Hawai'i	<i>Jordan</i>	SWFSC	273
STAR	6 Oct - 9 Dec, 2003	ETP	<i>McArthur II</i>	SWFSC	260
SPLASH	29 June - 20 Oct, 2004	Alaska	<i>McArthur II</i>	SWFSC	35
PICEAS	29 July - 14 Nov, 2005	Pacific Islands	<i>McArthur II</i>	SWFSC	229
STAR	30 July - 6 Dec, 2006	ETP	<i>McArthur II</i>	SWFSC	731

Acoustic activity (presence/absence of phonations) within the limits of our monitoring bandwidth was compared among species. The acoustic detection distance, or the greatest distance at which phonations could be confidently matched to a known dolphin sighting, was compared for each species. Variation in acoustic activity

was examined using Classification and Regression Tree analysis (CART) to determine which factors influenced the detection of dolphin schools (latitude, longitude, group size, sea state).

3. RESULTS

This analysis includes 2,034 acoustic detections of dolphin schools made during seven years of combined visual-acoustic line-transect surveys of cetaceans in the Pacific Ocean. A total of 971 single species schools were identified to species by experienced visual observers and included: *Stenella attenuata*, *S. coeruleoalba*, *S. longirostris*, *Delphinus* spp., *Tursiops truncatus*, *Steno bredanensis*, *Pseudorca crassidens*, *Globicephala* spp., *Lagenorhynchus obliquidens*, *L. obscurus*, *Lissodelphis borealis*, *Grampus griseus*, *Orcinus orca*, *Berardius bairdii*, and *Feresa attenuata*. Phonations produced by *Delphinus delphis* and *Delphinus capensis* were grouped together as *Delphinus* spp., as were detections produced by *Globicephala macrorhynchus* and *Globicephala melas* (*Globicephala* spp.). In addition, mixed-species schools of *S. attenuata* and *S. longirostris* were included in some analyses.

Overall, 73% of sighted dolphin schools were also detected acoustically. The percentage of sighted schools that were detected both visually and acoustically ranged from 28% for *Berardius bairdii* to 100% for *Pseudorca crassidens* (Table 2). Dolphin species that had a high acoustic detection rate (> 80% of schools) were found in significantly larger schools than species with a low acoustic detection rate (Mann-Whitney *U*, $p < 0.001$). The mean group size of schools detected acoustically was significantly (Mann-Whitney *U*, $\alpha = 0.05$) greater than the mean group size of schools not detected acoustically for most species (Table 2). The CART analysis showed that group size was the most important factor associated with the acoustic detection of dolphin schools, both overall and for each species individually.

Most dolphin species found in our study areas are known to produce whistles within the acoustic detection bandwidth of the equipment used during these surveys (Table 3). Whistles were evident in 93% of the 2,034 acoustic detections; however, not all species produced whistles. No whistles were detected from schools of *Lissodelphis borealis*, *Lagenorhynchus obliquidens*, *L. obscurus*, or *Berardius bairdii*. Maximum acoustic detection distance varied from 1.5 nmi for *Lissodelphis borealis* to 10 nmi for *Stenella coeruleoalba* (Table 3). Dolphin species in which most schools were found to produce whistles were generally detected at greater distances (Table 3).

Many dolphin groups were detected and localized using acoustic methods but were not seen by visual observers. Species was not known with certainty for groups that were not seen. These data were not examined for this study.

4. DISCUSSION

This study provides the largest dataset of simultaneous visual and acoustic observations of cetaceans during shipboard line-transect surveys published to date. The limited frequency bandwidth of our acoustic system did not allow for an examination of the full frequency range of dolphin phonations, however, for the purposes of population

surveys, detection of the school is of greater importance than detection of the full acoustic repertoire.

Nearly two-thirds of sighted dolphin schools were detected acoustically; however, acoustic detection of dolphin schools was not equal among species. Of the variables included in the analysis, group size was found to be the single most important factor influencing the acoustic detection of dolphin schools, both among and within species. Most dolphin schools that were not detected by the acoustic team contained fewer than 20 animals. Species that were consistently detected acoustically had large mean group sizes. For example, 85% of *S. attenuata* schools were detected acoustically and this species had an average school size of 93.1. There are exceptions to this trend, however. All *P. crassidens* schools and 96.8% of *Steno bredanensis* schools were detected acoustically, but these species had small mean group sizes (10.7 and 15.3, respectively). In the case of *P. crassidens*, individual group sizes were small, but encounters included a large number of these small groups spread out over large areas. *Steno bredanensis*, on the other hand, are found in small isolated groups, and there is no clear explanation for their high level of acoustic activity.

For some species, fewer than 70% of sighted schools were detected using acoustic methods, including: *G. griseus*, *Lagenorhynchus* spp., *O. orca*, *Lissodelphis borealis*, *F. attenuata*, and *B. bairdii*. With the exception of three sightings of *Lagenorhynchus obscurus*, all of these were relatively small schools. Also, with the exception of *F. attenuata*, whistles were detected from fewer than half of the schools of these species. It is possible that these species mainly produce high frequency clicks and that the limited bandwidth of our equipment prevented the detection of these sounds.

Given that 93% of the groups that were detected acoustically produced whistles, the use of whistle sounds for detection would allow most schools to be picked up. Whistles tend to be lower in frequency than most click sounds, and can therefore be detected using less expensive, lower bandwidth systems than would be necessary for click detection and identification. In addition, lower frequencies propagate further than higher frequencies, suggesting that whistles can be detected over greater distances than clicks. It is possible that whistles play an important role in communication over the large areas occupied by these groups.

From our analysis of the acoustic detection of dolphin schools during these surveys, we define two detection categories: dolphin species with a high rate of acoustic detection (>80%) and dolphin species with a low rate of acoustic detection (<80%). Dolphin species with a high rate of acoustic detection were typically found in large schools and frequently produced whistles. Most of these species were found in the tropical study areas (Hawai'i, Pacific Islands, eastern Tropical Pacific Ocean). The species with a low rate of acoustic detection were typically found in smaller schools and produced few, if any, whistles. These species were more common in the temperate study areas off the west coast of the United States, Canada, and Alaska.

Table 2. Mean group size for dolphin schools detected (1) both visually and acoustically, and (2) only visually. For all detections, the percent vocal indicates the percentage of sighted schools that were detected using acoustic methods. For acoustic detections, the percentage of detections that included whistles is given. Species are arranged according to the percentage of schools detected acoustically (percent vocal). A statistical comparison was made of the group sizes for acoustic/visual detections and for visual-only detections (Mann-Whitney U test).

Species	Acoustic/Visual Detections		Visual-Only Detections		Mann-Whitney U Test	All Acoustic Detections	All Detections
	Sample Size	Group Size	Sample Size	Group Size	Significance	% with whistles	% vocal
<i>P. crassidens</i>	19	10.7	-	-	-	100.0%	100.0%
<i>S. bredanensis</i>	30	15.3	1	7.3	0.434	90.3%	96.8%
<i>S. attenuata, S. longirostris</i>	71	351.5	4	131.5	0.122	100.0%	94.7%
<i>Delphinus</i> spp.	134	192	23	62	0.001	98.5%	85.4%
<i>S. attenuata</i>	81	93.1	14	41.9	0.012	97.4%	85.3%
<i>T. truncatus</i>	62	78.1	13	10.1	0.020	96.7%	82.7%
<i>S. longirostris</i>	37	116.4	9	38.1	0.008	100.0%	80.4%
<i>S. coeruleoalba</i>	149	60.4	37	48.3	0.047	100.0%	80.1%
<i>Globicephala</i> spp.	55	21.2	21	14.4	0.064	92.6%	72.4%
<i>L. obscurus</i>	3	280	2	9.5	0.083	0.0%	60.0%
<i>G. griseus</i>	28	21.2	30	9.8	0.021	44.8%	48.3%
<i>L. obliquidens</i>	4	19.5	5	11.5	0.712	0.0%	44.4%
<i>O. orca</i>	21	11.9	28	5.6	0.011	50.0%	42.9%
<i>L. borealis</i>	7	27.3	13	7.8	0.021	0.0%	35.0%
<i>F. attenuata</i>	2	23.9	4	7.9	0.064	100.0%	33.3%
<i>B. bairdii</i>	2	16	5	7.6	0.245	0.0%	28.6%

In general, the limited bandwidth of the acoustic equipment used during these surveys was sufficient for the detection of dolphin schools encountered in tropical and sub-tropical study areas (*P. crassidens*, *Steno bredanensis*, *Delphinus*

spp., *Stenella* spp., *T. truncatus*). Further examination of the data may provide a better understanding of why some tropical dolphin schools were not detected using these acoustic methods.

Table 3. Acoustic detection distance and whistle frequency range for each species. The maximum acoustic detection distance (nmi) provides the range at which our equipment detected sounds from each species. Frequency ranges (kHz) of whistles were obtained from the literature, and all fall within the 2-24 kHz detection range of our equipment (note: the authors have detected whistles in the presence of *F. attenuata*, but there are no published descriptions of whistles for this species). Species are labeled from highest acoustic detection rate to the least (Table 2).

Species	Detection Distance		Whistle Range		Reference
	Mean (St. Dev)	Maximum	Low Frequency	High Frequency	
<i>P. crassidens</i>	2.87 (1.64)	6	1.8	18	Oswald <i>et al.</i> (2007b)
<i>S. bredanensis</i>	1.53 (1.19)	4.5	4	9.5	Oswald <i>et al.</i> (2007b)
<i>Delphinus</i> spp.	2.22 (1.6)	6	3.5	23.5	Oswald <i>et al.</i> (2007b)
<i>S. attenuata</i>	1.88 (1.54)	6	3	21	Oswald <i>et al.</i> (2007b)
<i>T. truncatus</i>	1.75 (1.31)	6	1.9	21.6	Ding, <i>et al.</i> (1995)
<i>S. longirostris</i>	2.57 (1.56)	6	4	25	Oswald <i>et al.</i> (2007b)
<i>S. coeruleoalba</i>	2.61 (1.83)	10	1	23	Oswald <i>et al.</i> (2007b)
<i>Globicephala</i> spp.	2.58 (1.79)	8.5	0.3	23.6	Oswald <i>et al.</i> (2007b)
<i>L. obscurus</i>	0.98 (1.33)	2.5	1	27	Ding, <i>et al.</i> (1995)
<i>G. griseus</i>	0.93 (0.7)	2.3	2	24	Rendell <i>et al.</i> (1999)
<i>L. obliquidens</i>	0.71 (0.87)	2	2	20	Caldwell and Caldwell (1971)
<i>L. hosei</i>	2	2	4.3	24	Oswald <i>et al.</i> (2007a)
<i>O. orca</i>	0.73 (0.71)	2.3	1.5	18	Thomsen <i>et al.</i> (2001)
<i>L. borealis</i>	0.58 (0.67)	1.5	-	-	Rankin, <i>et al.</i> (2007)
<i>F. attenuata</i>	1 (1.05)	1.75	-	-	*
<i>B. bairdii</i>	1.1 (0.84)	1.7	4	8	Dawson, <i>et al.</i> (1998)

Many of the species encountered in the temperate study areas (*Lagenorhynchus* spp., *G. griseus*, *O. orca*,

Lissodelphis borealis, and *B. bairdii*) had low rates of acoustic detection. It is possible that the limited bandwidth

of our acoustic equipment prevented the detection of many of these dolphin schools. Acoustic studies conducted in these areas should be carried out using broadband equipment to guarantee the detection of higher-frequency click sounds produced by these species. An increased bandwidth (over 100 kHz) for cetacean studies in the temperate regions would also allow for detection of porpoise species, which could not be included in this study. Despite our bandwidth limitations, we detected both clicks and click bursts from many groups and were able to describe the sounds produced by *L. borealis* (Rankin *et al.* 2007).

5. CONCLUSION

The Southwest Fisheries Science Center has been using a standard protocol for combined visual and acoustic shipboard line-transect cetacean surveys for seven years. Using this standard protocol, we have been able to detect and localize odontocete groups in situations in which the visual team was unable to work due to weather or darkness. Our ability to detect dolphin schools varies by species, group size, and acoustic behavior. These results highlight the variation in acoustic behavior within and among species, and the need for a more rigorous examination of the acoustic behavior of each species. Nonetheless, the high rate of acoustic detection of dolphin schools in the tropical and sub-tropical Pacific Ocean justifies the use of acoustic methods for the detection of most dolphin schools within these areas.

REFERENCES

Au, W. W. L. 1993. The sonar of dolphins (Springer-Verlag, New York).

Au, W. W. L. 1980. Echolocation signals of the Atlantic bottlenose dolphin (*Tursiops truncatus*) in open waters. In Animal sonar systems (R. G. Busnel, and J. F. Fish, eds). Plenum, New York, pp. 251-282.

Caldwell, M. C., and D. K. Caldwell. 1971. Statistical evidence for individual signature whistles in Pacific white-sided dolphins, *Lagenorhynchus obliquidens*. *Cetology* 3:1-19.

Dawson, S., J. Barlow, and D. Ljungblad. 1998. Sounds recorded from Baird's beaked whale, *Berardius bairdii*. *Mar. Mamm. Sci.* 14:335-343.

Ding, W., B. Würsig, and W. Evans. 1995. Comparison of whistles among seven odontocete species. In Sensory systems of marine mammals. (R.A. Kastelein, J. A. Thomas, and P. E. Nachtigall, eds.), De Spil Publishers, Woerden, pp. 299-323.

Gordon, J. C. D., J. N. Matthews, S. Panigada, A. Gannier, J. F. Borsani, and G. Notarbartolo di Sclara. 2000. Distribution and relative abundance of striped dolphins in the Ligurian Sea Cetacean Sanctuary: results from an acoustic collaboration. *J. Cet. Res.* 2:27-36.

Herzing, D. L. 2000. Acoustics and social behavior of wild dolphins: implications for a sound society. In Hearing by whales and dolphins. (W. W. L. Au, A. N. Popper, and R. R. Fay, eds.). Springer-Verlag, New York, pp. 225-272.

Janik, V. M., and P. J. B. Slater. 1998. Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls. *Anim. Behav.* 56:829-838.

Kinzey, D., P. Olson, and T. Gerrodette. 2000. Marine mammal data collection procedures on research ship line-transect surveys by the Southwestern Fisheries Science Center. SWFSC Admin Repot. LJ-00-08. 32p.

Lammers, M. O., W. W. L. Au, and D. L. Herzing. 2003. The broadband social acoustic signaling behavior of spinner and spotted dolphins. *J. Acoust. Soc. Am.* 114:1629-1639.

Mellinger, D. K. 2001. ISHMAEL 1.0 User's Guide. NOAA Tech. Mem. OAR-PMEL-120, available from NOAA/PMEL, 7600 Sand Point Way, NE, Seattle, WA 98115-6349.

Oswald, J. N., S. Rankin, and J. Barlow. 2004. The effect of recording and analysis bandwidth on acoustic identification of delphinid species. *J. Acoust. Soc. Am.* 116:3178-3185.

Oswald, J. N., S. Rankin, and J. Barlow. 2007a. A tool for real-time acoustic species identification of delphinid whistles. *J. Acoust. Soc. Am.* 122:587-595.

Oswald, J. N., S. Rankin, and J. Barlow. 2007b. First description of whistles of Pacific Fraser's dolphins, *Lagenodelphis hosei*. *Bioacoustics*, *in press*.

Rankin, S., J. N. Oswald, J. Barlow, and M. O. Lammers. 2007. Patterned burst-pulse vocalizations of the northern right whale dolphin, *Lissodelphis borealis*. *J. Acoust. Soc. Am.* 121(2):1213-1218.

Rendell, L. E., J. N. Matthews, A. Gill, J. C. D. Gordon, and D. W. Macdonald. 1999. Quantitative analysis of tonal calls from five odontocete species, examining interspecific and intraspecific variation. *J. Zool. Lond.* 249:403-410.

Thomas, J. A., S. A. Fisher, and L. M. Ferm. 1986. Acoustic detection of cetaceans using a towed array of hydrophones. *Rept. Int. Whal. Comm. Spec. Iss.* 8:139-148.

Thomsen, F., D. Franck, and J. K. B. Ford. 2001. Characteristics of whistles from the acoustic repertoire of resident killer whales (*Orcinus orca*) off Vancouver Island, British Columbia. *J. Acoust. Soc. Am.* 109:1240-1246.

ACKNOWLEDGEMENTS

This work could not have been accomplished without the assistance and cooperation of the scientists, officers and crew of the *McArthur*, *McArthur II*, and the *David Starr Jordan*. Special thanks to Liz Zele and the many acousticians who helped in the field. Funding for this research was provided by the Southwest Fisheries Science Center and the U. S. Navy.

DEFINITION OF THE ANTARCTIC AND PYGMY BLUE WHALE CALL TEMPLATES. APPLICATION TO FAST AUTOMATIC DETECTION

Samaran Flore¹, Olivier Adam², Jean-François Motsch², Christophe Guinet¹

¹Centre d'Études Biologiques de Chizé, CNRS, 79360 Villiers en bois, France. samaran@cebc.cnrs.fr

²Laboratoire Images, Signaux et Systèmes Intelligents, groupe Ingénierie des Signaux NeuroSensoriels, Université Paris-Est, France

ABSTRACT

This paper deals with the automatic detection of low-frequency Antarctic (*Balaenoptera musculus intermedia*) and Pygmy (*B. m. brevicauda*) blue whale sounds produced in the Southwestern Indian Ocean. A new detection method based on a matched filter is introduced. Four original match templates are presented and tested against original blue whale subspecies calls. The mathematical formulas of these templates, defined by Gaussian curve models, are provided. The detection threshold is based on the correlation coefficients. The threshold was set to reduce false detections obtained on simulated signals at various signal-to-noise ratios. We focus our work on the true detections of whale calls. Moreover, to obtain a real-time system, we decrease the computational time by decimating the recorded signal ($F_s=250\text{Hz}$). We show that this new method enables us to effectively detect both subspecies in various ambient noises, in the Southern Ocean.

RESUME

Dans ce papier, les sons de basses fréquences émis par les baleines bleues Antarctique (*Balaenoptera musculus intermedia*) et pygmées (*B. m. brevicauda*) dans le secteur sud – ouest de l'Océan Indien ont été détecté automatiquement à partir d'une technique de filtrage adapté. Pour ce faire, des signaux synthétiques ont été créés à partir de signaux originaux en modélisant leurs équations mathématiques à partir de courbes gaussiennes. La détection se fait alors par la corrélation entre le signal entrant et le modèle calculé (template). Le seuil de détection a été choisi au préalable en simulant une série de signaux dans des rapports signal sur bruit différents. Au final, un seuil de détection élevé a été choisi pour minimiser les fausses alarmes au risque d'augmenter les détections manquées. Pour diminuer le temps de calcul, le signal original ($F_s=250\text{Hz}$) a été décimé. Cette méthode originale c'est révélée très efficace pour détecter les sons émis par ces deux sous espèces de baleines bleues dans des niveaux de bruit ambiant très variés comme c'est le cas dans cette partie de l'Océan Indien.

1. INTRODUCTION

Knowledge of marine mammal sounds, and in particular baleen whale sounds, has been largely enhanced thanks to new acoustic data available from a wide variety of instruments that were originally designed to monitor the seismicity of the earth or for defence purposes. Instruments designed to monitor low frequency earthquakes (Watkins, 1981; Nishimura & Colon, 1994; Nieuwkerk et al., 2004), record seismic-acoustic signals and underwater seismicity (Stafford et al., 2004; Rebull et al., 2006), and listen to Soviet submarines during the cold war via the Navy SOSUS arrays (Costa, 1993; Gagnon & Clark, 1993; Clark & Mellinger, 1994; Mellinger & Clark, 2003) recorded a great variety of calls in the lower frequency range. These calls included Blue (*Balaenoptera musculus*), Fin (*B. physalus*), and Humpback (*Megaptera novaeanglia*) whales over long periods. Recordings of baleen whale calls document the seasonal distributions, the relative abundance, and the acoustic behaviour of particular species. Moreover, they have also been useful in tracking animals in their natural habitat (Mellinger & Clark, 2003; Stafford et al., 2001,

2003, 2004; Širović et al., 2004; McDonald et al., 2006).

The hydroacoustic stations of the International Monitoring System (IMS) were primarily designed to continuously record natural and artificial sounds in the oceans, particularly sounds generated by man-made explosions in support of the Comprehensive Nuclear-Test-Ban Treaty (CTBT) (Roueff et al., 2004). Between May 2003 and April 2004 six IMS stations, provided by the Commissariat à l'Énergie Atomique (CEA) were deployed off the coast of Possession Island (Crozet archipelagos in the French Indian Ocean Territory). The low frequency hydrophones (1-100 Hz) have enabled recordings of a large variety of signals: time-variant ambient underwater noise, biological signals including large baleen whales calls, and anthropogenic sounds.

Our aim is to detect the Antarctic blue whale calls (*B. m. intermedia*) and the Pygmy blue whale calls "Madagascar-type" (*B. m. brevicauda*) in the CEA dataset. Using spectrograms, our first analysis identifies the presence of 2 subspecies calls in the approximately 40,000 hour-long dataset. These calls contain some uniform patterns with one or more units, high acoustic intensity (above 180 dB re 1 μPa at

1m), very low frequency range ([28-35 Hz]) and are repetitive (Clark, 1990; Ljungblad et al., 1998; Mellinger & Clark, 2003; Širović et al., 2004, 2007; Stafford et al., 2004; Rankin et al., 2005; McDonald et al., 2006).

Manually detecting each specific blue whale call among a large amount of data would require many hours of effort to scan the spectrograms visually and to listen to the recordings (Stafford et al., 2007). In this context, automatic processes to identify the location, the characteristics, and the abundance of the calls in the dataset are necessary. Moreover, automated detection methods provide objective criteria to detect and count a known sound in a year-long dataset within hours or days. An automated, fast, real-time detection method for blue whale calls was used to analyze the dataset obtained from permanent acoustic stations.

Recently, a variety of methods have been developed and used for automatic recognition of marine mammal sounds. The classical technique is based on the spectrogram matched filter, i.e. the cross-correlation between the spectrograms of the signal of reference (template) and the recorded signal. This cost minimisation technique constitutes the basis of the dynamic time-warping (DTW) developed in human speech recognition (Silverman & Morgan, 1990). In this case, the recorded signal could be compressed or dilated before being compared to the template. A variant of this approach, called crosswords reference templates (CWRTs), consists of comparing the recorded signal with a great variety of templates (Abdulla, 2003). Another approach is based on the use of templates defined in the frequency domain. The cross-correlation templates are obtained from the shapes of the known recorded signal spectrograms (Mellinger & Clark, 2000); this method has been used successfully to classifying right whale calls (*Eubalaena japonica*) (Munger et al., 2005). An edge detector has also been tried directly on the spectrogram (Gillepsie, 2004). The choice of referent spectrograms from real recordings determines the performance of the detector. Moreover, the performance may depend on the dataset, in which case it is difficult to generalize the results to other datasets. The referent call contains features of a single individual. If these features are not close to those of other individuals (of the same subspecies) referent spectrograms become non exhaustive. Recently, new methods were proposed including Hidden Markov Modeling (HMM) associated with Artificial Neural Networks (ANN) techniques (Trentin & Gori, 2003), and methods based on time-frequency or time-scale representation such as wavelets. The main disadvantage of these methods is their computational complexity as compared to the matched filter.

The signal conditioner and template definition are key to the successful implementation of the matched filter. To optimize performance detection, we have not chosen to extract one call randomly from the dataset and to use it as the referent signal. It is also important to spend time on the signal pre-treatment, especially the filtering process. This step contributes to improving performance of the detector. In our application, the signal-to-noise ratio varies with each hydrophone for the duration of the dataset. The use of multiple

templates improves detector performance due to variation and distortion among blue whale calls (Munger et al., 2005; Mellinger, 2004).

In this paper, we provide details on the pre-processing of the signal and we describe the mathematical formulas of the different synthetic waveforms used for the detection of both blue whale subspecies calls. Our work is based on the analysis of the CEA dataset and on the knowledge of the blue whale calls (Stafford et al., 2004, 2005; Rankin et al., 2005; Širović et al., 2004, 2007; McDonald et al., 2006). The results obtained for different cross-correlation thresholds and different signal-to-noise ratios (SNR) are presented. It should be noted that our work was aimed at minimizing computational complexity. Before concluding, we present results of true and false detections on real signals.

2. MATERIALS AND METHODS

2.1 Dataset and blue whale calls

This IMS dataset has been made available for the analysis of South Indian Ocean biological signals. In May 2003, six autonomous stations were moored on the northern (H04N1, H04N2, and H04N3) and southern coasts (H04S1, H04S2, H04S3) of Possession Island (Crozet archipelagos in the French Indian Ocean Territory) in the Indian Ocean between 46°09'S-46°51'S and 51°48'E-51°53'E. Each station consisted of an anchor, a buoy and a hydrophone, called an Underwater Monitoring Unit (UMU).

The optical fiber cable and the converter transmitter constitute the digital communication link, once the analog-to-digital conversion (performed in the UMU) has been carried out. The digital acquisition and storage system perform data format changes without affecting the sampling rate and sample values. The data are dated by a 1ms precision absolute clock synchronized by GPS. Data are transmitted in real time via satellite link to the International Data Center to be analyzed at the CEA / DIF / DASE - Bruyères-le-Châtel FRANCE. These instruments are moored to the seafloor between 1100 and 1500 meter depths. Sensors are suspended near the sound channel axis (SOFAR) at a depth of approximately 300m. They were deployed in a triangular configuration (triad) far from the northern and southern coasts of the island with approximately 2 km spacing between moorings and 60 km between two triads. Acoustic data for H04N2, H04N3, H04S1 and H04S3 were available for the entire recording period; data for H04S2 were available from May 2003 to December 2003; and no data were available for H04N1 due to instrument failures.

The UMU contains the sensor, the analog signal conditioning circuits, and the analog-to-digital converter. These instruments monitored sound continuously, at a sampling rate of 250 Hz, coded by 24 bits (S/N: 126.5 dB), and a flat (± 3 dB) frequency response of 1.2-102.5 Hz. Note that the ambient underwater noise is time-variant for the duration of the dataset. For example, during a given month (March 2004), the mean acoustic pressures recording by the hydrophones

are different between the northern network (92.6 ± 2.5 dBrms) and the southern network (109.5 ± 6.4 dBrms) (Table 1). In our preliminary study, we focus on two types of easily recognizable calls: the Antarctic blue whale call (BMi) and the Pygmy blue whale call (BMb) “Madagascar type” (bandwidth [15-35 Hz]).

Antarctic blue whale calls

Spectrograms of the first category of detected calls are similar to those of typical Antarctic blue whale calls (Ljungblad et al., 1998; Širović et al., 2004; Stafford et al., 2004; Rankin et al., 2005). These calls consist of three tonal units repeated in patterned sequences every 40-50 seconds over a period of a few minutes or hours (Figure 1).

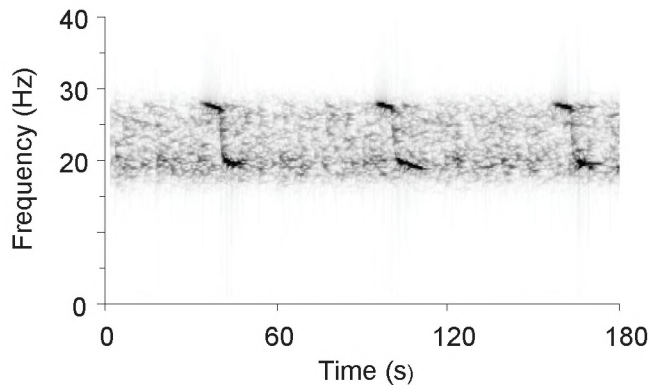


Figure 1: Spectrogram of Antarctic blue whale calls recorded off Crozet Island (Spectrogram parameters: 1024 points FFT length, 90% overlap, 250 Hz sample rate, Hanning, for a filter bandpass between 18 and 28 Hz)

The first component is a constant frequency tone centered at 28 Hz followed by a short frequency-modulated (FM) down-sweep from 28 Hz to 20 Hz ending with the third component, a slightly modulated tone (20-18 Hz). This call lasts approximately 26 seconds but sometimes only the first one or two components are present. This degree of variability in the presence of the three individual components was previously reported (Stafford et al., 2004; Rankin et al., 2005). In the dataset, the calls have variable amplitudes (from 84.3 to 117.8 dB re $1 \mu\text{Pa}$ at 1m) depending on the distance of the whales to the hydrophones and the original amplitude of their sounds (Table 1).

Pygmy blue whale calls

Since the first pygmy blue whale call description established by Ljungblad (1998), information regarding the content of these calls has been scarce. These low frequency calls were often present in the dataset. Like Antarctic blue whale calls, these signals occur in patterned sequences of long tonal calls every 90-100 seconds over the course of a few minutes or hours (Figure 2).

Each sequence is composed of two long units that repeat themselves. The first component is primarily a constant frequency tone at 35 Hz lasting 15-20 seconds. A silence

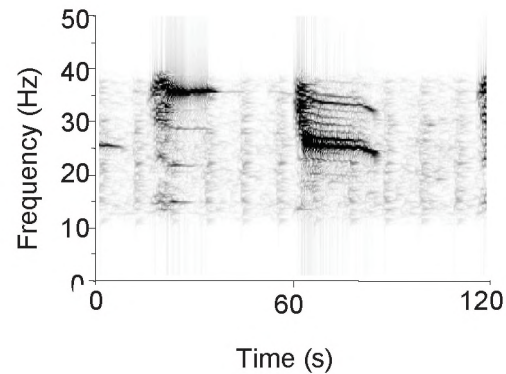


Figure 2: Spectrogram of pygmy blue whale calls “Madagascar type” recorded off Crozet Island (Spectrogram parameters: 1024 points FFT length, 90% overlap, 250 Hz sample rate, Hanning, for a filter bandpass between 12 and 40 Hz).

(approximately 20 sec) separates the two-part phrase. The second component starts with a 1-2 second 15-28 Hz FM down-sweep that ends with a long (20 sec) slightly modulated tone. Each component has strong associated harmonics. In the dataset, the signal-to-noise ratio is time-variant and could have a negative value.

2.2 Automatic detection methods

We present the specific synthetic waveforms, the process for the matched filter and our approach for choosing the detection threshold.

Definition of the templates

In both cases (BMi and BMb), we follow the approach described in Figure 3. The first step is to condition the original signal. As previously mentioned, the sample frequency is 250 Hz. We applied first a high-pass filter then a low-pass filter on the dataset. We used Butterworth filters which present a frequency maximally flat response. Since the frequency bandwidths vary for the 2 subspecies whale calls, different filters for the BMi and the BMb whales are necessary. For the BMi (resp. BMb), the order of the filter is 10 (resp. 12) and the cut-off frequencies are 13 Hz (resp. 17 Hz) and 30 Hz (resp. 50 Hz). The signal is decimated by 2.

The second step allows the extraction of the common features of the parts of the signal with high energy in this bandwidth. This first detection method is based on the energy with non-overlapping sliding windows of 24.6 sec. The noise is reduced when using the average method. Our first choice is to use the recordings with the higher signal-to-noise ratio but we obtain similar results with the complete dataset (Table 2).

The objective of the third step is to synchronize each part of the signal that contains the call. To that effect, we calculate the cross-correlation between the dataset and the averaged signal obtained at the end of Step 2. The averaged signal is used to define the model of the template. Finally both subspecies calls are modelled using Gaussian curves to obtain the equations of the templates. Step 4 will be described in the following section.

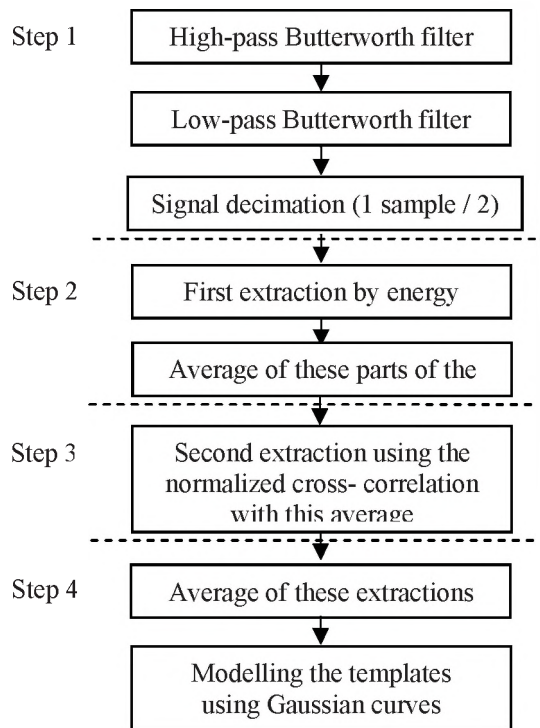


Figure 3: Algorithm for the definition of the call templates. This algorithm is applied first to the BMi calls and after to the BMb calls.

Step 1: Conditioning the original recorded signal.

Step 2: Search for common features

Step 3: Time-synchronization of each part of the original signal

Step 4: Template obtained with the Gaussian model

Definition of the Antarctic blue whale calls template

For the Antarctic blue whale (BMi) the equation of the synchronized averaged signals is modelled in 2 different parts. For part 1 of the signal, the main frequency is 28 Hz and the spectrum amplitude is modelled using a single Gaussian curve. For the second part, the main frequency is 19 Hz and the spectrum amplitude is modelled using 4 Gaussian curves (Figure 4).

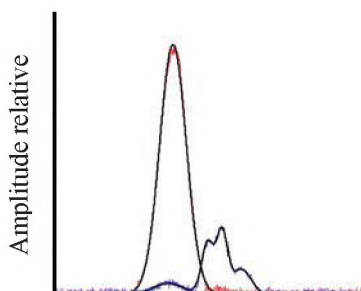


Figure 4: Modelling the spectrum with Gaussian curves

The mathematical formula for the BMi call template is:

$$\begin{aligned}
 tpl_{BMi}[k] = & \underbrace{a_0^{BMi1} \times e^{-\left(\frac{k-b_0^{BMi1}}{c_0^{BMi1}}\right)^2} \times \sin(2\pi f_0^{BMi1} k)}_{PART 1} \\
 & + \underbrace{\left(\sum_{j=0}^3 a_j^{BMi2} \times e^{-\left(\frac{k-b_j^{BMi2}}{c_j^{BMi2}}\right)^2} \right) \times \sin(2\pi f_0^{BMi2} k)}_{PART 2} \quad (1)
 \end{aligned}$$

with $k=1 \dots 3000$ and

$$f_0^{BMi1} = 27.53\text{Hz}, f_0^{BMi2} = 19.36\text{Hz}.$$

The parameters are:

Part 1:

$$a_0^{BMi1} = 10.9, b_0^{BMi1} = 2378, c_0^{BMi1} = 372.4$$

Part 2:

$$a_0^{BMi2} = 0.4648, b_0^{BMi2} = 2288, c_0^{BMi2} = 420$$

$$a_1^{BMi2} = 2.659, b_1^{BMi2} = 3353, c_1^{BMi2} = 148.5$$

$$a_2^{BMi2} = 2.265, b_2^{BMi2} = 3070, c_2^{BMi2} = 169.4$$

$$a_3^{BMi2} = 1.081, b_3^{BMi2} = 3730, c_3^{BMi2} = 261.2$$

Equation 1 allows for the reconstructing of the template sample by sample. The time representation of the BMi call template is shown in Figure 5. The duration of this template is 24 seconds. To validate our template, we apply the modelling process on the 1-hour length signal having the highest signal-to-noise ratio and on the complete dataset. We obtain two similar templates. The frequencies of the 2 parts of these templates are presented in Table 2. The correlation coefficients obtained between an unknown signal and these two templates are similar because of the similitude of the two templates.

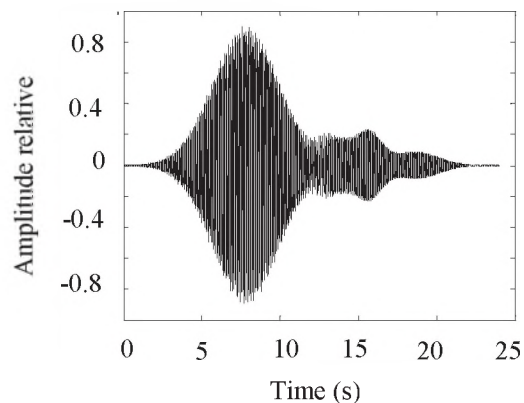


Figure 5: The template of BMi calls

Definition of the Pygmy blue whale calls template

For the Pygmy blue whale (BMb), the equation of the synchronized signals is more complex. We distinguish three

parts. The durations of part 1, part 2 and part 3 are 22.3 seconds, 20 seconds and 26.7 seconds respectively. Note that we consider the second part as a silence between part 1 and part 3. Employing the same approach as described before, we obtain the model of the spectrum using Gaussian curves (Figure 6).

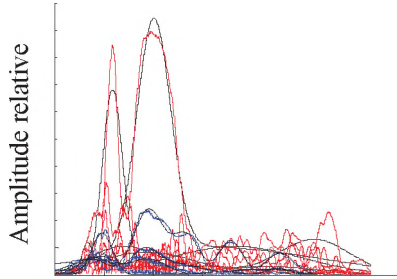


Figure 6: Modelling the spectrum using Gaussian curves

The equations of the 3 parts of the template are:

Part 1:

$$tpl_{BMb1}[k] = \sum_{i=0}^3 \left(\sum_{j=0}^4 a_{ij}^{BMb1} \times e^{-\left(\frac{k_{ij}^{BMb1} - b_{ij}^{BMb1}}{c_{ij}^{BMb1}} \right)^2} \right) \sin(2\pi f_i^{BMb1} k) \quad (2)$$

With $k=1 \dots 2790$ and the frequencies are

$$f_0^{BMb1} = 35.03 \text{ Hz}, f_1^{BMb1} = 14.13 \text{ Hz}, \\ f_2^{BMb1} = 21.11 \text{ Hz}, f_3^{BMb1} = 22.72 \text{ Hz}.$$

The duration and the parameters of each Gaussian curve are respectively reported in Table 3 and Table 4.

Part 2:

$$tpl_{BMb2}[k] = 0 \quad (3)$$

With $k = 2791 \dots 5295$;

Part 3:

$$tpl_{BMb3}[k] = \sum_{i=0}^4 \left(\sum_{j=0}^1 a_{ij}^{BMb3} \times e^{-\left(\frac{k_{ij}^{BMb3} - b_{ij}^{BMb3}}{c_{ij}^{BMb3}} \right)^2} \right) \sin(2\pi f_i^{BMb3} k) \quad (4)$$

With $k = 5296 \dots 8639$ and the frequencies are

$$f_0^{BMb3} = 24.96 \text{ Hz}, f_1^{BMb3} = 26.05 \text{ Hz}, \\ f_2^{BMb3} = 23.96 \text{ Hz}, f_3^{BMb3} = 27.15 \text{ Hz}, f_4^{BMb3} = 33.0 \text{ Hz}$$

The parameters of the Gaussian curves are reported in Table 5 and Table 6. The time-representation of the complete template is given in Figure 7.

Presentation of our detector based on the matched filter

Our preliminary analysis of the CEA dataset regularly shows incomplete calls for the 2 subspecies of blue whales. For the Antarctic blue whale, this observation is well documented *Canadian Acoustics / Acoustique canadienne*

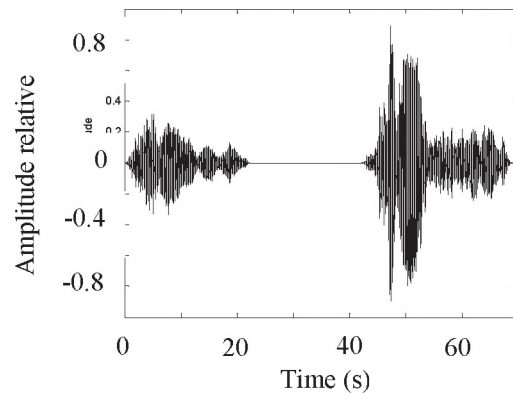


Figure 7: The template of BMb calls

(Stafford et al., 2004; Rankin et al., 2005). However, for the sounds emitted by the Pygmy blue whale, there are few references available. To our knowledge, no study has reported such incomplete calls. We define 2 new templates corresponding to these incomplete calls. These templates are deduced from the previous templates: each new template is composed of the first part of the calls only. We respectively note that BMie and BMbe are the incomplete calls for BMi and BMb.

$$tpl_{BMie}[k] = tpl_{BMi1}[k] \quad (5)$$

$$tpl_{BMbe}[k] = tpl_{BMb1}[k] \quad (6)$$

The algorithm is based on the cross-correlation of the dataset and these 4 templates (Figure 8):

$$R_{xy} = \frac{TF^{-1}(X \times Y^*)}{\sqrt{\sum x^2 \times \sum y^2}} \quad (7)$$

where X is the dataset spectrum and Y the template spectrum. Note that spectrums for the 4 templates are calculated before starting the detection process to reduce the computation time. The results list the occurrence of the calls for the 2 blue whale subspecies and some features are saved, like the name of the station, the time of the beginning and the end of the call (year, month, day, hour, minute and second), the signal intensity (Peak and RMS), and the value of the correlation coefficient.

3 RESULTS AND COMMENTS

3.1 Selected threshold for the cross-correlation

The objective of the signal detection method is to validate one of these 2 hypotheses (Harvey, 1992):

$$\begin{cases} H_0 : x = n \\ H_1 : x = s + n \end{cases} \quad (8)$$

with x, s, n respectively the observation, the signal that we have to detect, and the noise.

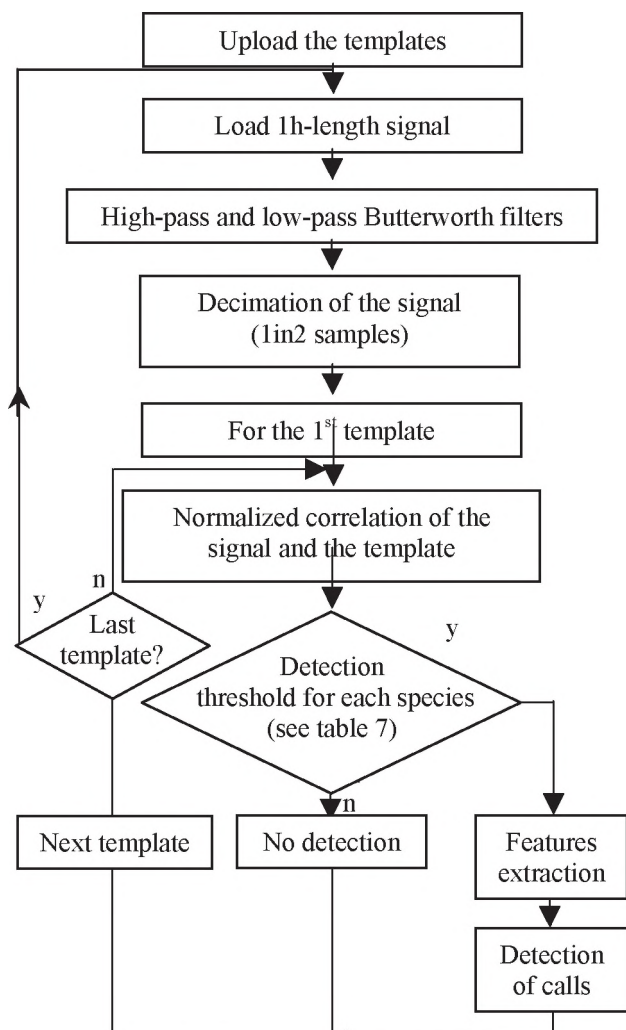


Figure 8: The automatic detection method

Whatever the detector, a threshold can be used to distinguish both hypotheses. The performance of the detector is based on the choice of the threshold. We have selected the value of the threshold from the analysis of simulated signals. These synthetic signals are composed of 100 BMi templates corrupted with a white Gaussian noise. We can assume that the white Gaussian noise properties are close to those of the underwater noise on the specific narrow bandwidth of our application ([20-40Hz]). The distribution of the templates is coherent with a real recorded blue whale signal (same rhythm). We change the signal-to-noise ratio from -30 to 25 dB (range 5 dB). The goal is to assess the value of the threshold for the efficiency of the detection method. Results of the total number of detections, correct detections and false alarms are shown in Table 7. Note that we consider the detection correct when the call is localized at ± 1 sec.

First, the number of the total detections decreases with the SNR and increases when the threshold value decreases. For Gaussian white noise only, no call was detected. Second, the rate of the correct detections is 100% for SNR higher than -15 dB, showing the resistance of this approach to noise. This rate decreases dramatically when the SNR is less than -20 dB.

The number of false alarms increases significantly when the threshold decreases. For example, when the SNR is 25 dB, the false alarm rate varies from 0 to 66 when the threshold value varies from 0.19 to 0.1. On the other hand, the false alarm rate increases proportionally as the SNR decreases and reaches a maximum value for SNR=-30 dB. Table 7 shows that when the threshold is superior to 0.17, the number of total detections and the correct detection are 100%, except when the SNR<-15 dB.

The same method was applied for BMie, BMb, and BMbe in choosing the threshold. The selected thresholds were respectively 0.17 and 0.15 for the complete and incomplete Antarctic blue whale calls. For both pygmy blue whale calls, the threshold is 0.14. Nevertheless, our margin for error allowed for an occasional missed call because the calls are produced very regularly and the main objective in this process is to decipher whether calls are present or not, and not to determine the exact number of calls. Figure 9 is plotted from data of table 7. The ROC curves are calculated with 6 different SNRs from -25 to 25 (range 10 dB). We deduce the threshold for BMi (0.17 in bold in Fig.9). Note that the signal-to-noise ratios are different in the northern and southern acoustic stations involving more false alarms in the south.

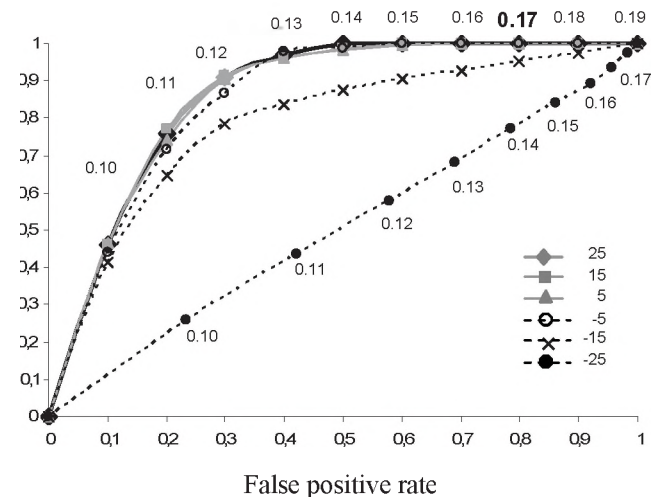


Figure 9: ROC curves from Table 7 value for BMi threshold.

3.2 Performance comparison for matched filter between templates and real referent calls

To validate the templates, we compare the results obtained from the matched filter using our templates and the matched filter using real blue whale calls. We used 4 different calls, 2 by subspecies. For each subspecies, we first extract from the dataset the call with the best shape and with a high signal-to-noise ratio. Second, we choose another call with a low signal-to-noise ratio. The correlation coefficient between the real calls with high (low) SNR and our templates gives 0.58 (0.17) for BMi and 0.35 (0.14) for BMb. For matched filters we use these 4 calls as templates and we apply the same detection thresholds. The results are given in Table 8.

The matched filter using high SNR correctly detects 47% (15%) of BMi (BMb) calls for the northern network dataset. These results decrease dramatically (to reach just less than 1%) for matched filters based on low SNR real calls. The performance is comparable for the southern network. These results show the benefit of using the templates in the matched filter.

The synthetic signals are not representative of a single individual but rather contain features common to calls of other individuals. We note too, a better resistance to the presence of the non-stationary underwater noise.

3.3 Detection of real blue whale calls

The number of calls detected for each blue whale subspecies reflects the appearance of the species, i.e. the vocal activity or the migration pattern of both subspecies (Širović et al., 2004; Stafford et al., 2004; McDonald et al. 2006). The number of blue whale calls detected in a given month (March 2004) is shown in Table 9. We choose March 2004 because the 2 subspecies are present during this month of the year and, secondly, the dataset is almost complete (98% on the 744 hours of the month).

The variation in the number of calls depends on the localization of the calling whales relative to each network and to the signal-to-noise ratio (recording conditions, ambient noise in the recording area). The correlation coefficients are proportional to the quality of the blue whale signals received at the hydrophones. Correlation coefficients vary between 0.17 and 0.72 for BMi calls and between 0.14 and 0.55 for BMb calls (Table 9). Note that the minimum values correspond to our thresholds (see §3.1).

For both whale subspecies, the number of detected calls is higher for the northern network dataset compared to the southern network dataset. This could be justified if the whales were constantly present in the north. This is true for BMb whales. But catches of Antarctic whales show that BMi whales were localized south of Crozet Island (Branch et al., 2007). The reason is that the noise level is higher on the hydrophones in the southern stations (Table 1). As seen in the previous section, our method detects fewer calls when the signal-to-noise ratio is less than -15 dB.

This presupposes that the detected calls are reliable blue whale calls. This result is reinforced by the correlation coefficient means superior to 0.2 for each whale and each network (Table 7).

3.4 Computation time

One of our objectives is to develop a method for real-time application. We attach great importance to computation time. Taking this constraint into account, we do not consider methods based on time-frequency representation. Moreover, we reduce the computation time by decimating the original signal by 2 and implementing the matched filter in the time domain. The algorithm code (Figure 8) is developed with Mathworks Matlab 7.04 and processed on Dell Pentium 4 CPU 2.4 GHz

(1 Go RAM). Note that we take into account the load of the recorded signal and the saved results. For the 2 subspecies and the 4 types of calls, the computation time is 3382.04 seconds for analysis of the entire March 2004 dataset. The ratio is approximately 1/1000.

This result allows us to consider real-time application. Computation time could be decreased by using machine code in place of Matlab. Moreover, most of the time is dedicated to the uploading and the conversion of the dataset. This is a drawback of post-processing analysis. This step is avoided for real-time application.

4 CONCLUSION

In this paper, we investigated the performance of matched filters dedicated to the automatic detection of the calls of 2 blue whale subspecies in long-term acoustic recordings in the Southwestern Indian Ocean. We presented the definition of 4 templates corresponding to the complete and incomplete calls of these whales. We provided the mathematical formulas for Antarctic blue whale and pygmy blue whale call templates. Our automatic detection is based on the cross-correlation method; we optimized the process to be time-efficient in analysing such long recordings.

This automated detection method was useful in detecting blue whale calls in the whole dataset. The limited range of variability in the Antarctic blue whale and pygmy blue whale calls allowed us to create the synthetic waveforms for both calls. Four templates were used for the matched filter. The choice of the detection threshold was based on minimizing the false alarms.

Compared to alternative automated detection methods where the performance can be modified by the length of the dataset, the acoustic characteristics of the call, the behaviour of the calling whale, the properties of the water, and the physical environment of the recording location, the pattern chosen here is efficient in detecting all Antarctic and pygmy blue whale calls present in a given recording. It is also human and dataset independent. Moreover, this automated detection method could evolve by completing the current library with other baleen whale calls.

We intend to test the method on another training set of blue whale calls recorded in the northern and eastern parts of the Indian Ocean. Our first perspective is to use certain specific features, in particular, rhythm of the whale calls, for increasing detection reliability. As an end goal, we will use this detector for extracting the time of arrival of calls on each hydrophone to localize the whales.

ACKNOWLEDGEMENTS

This work was supported by the French Ministry of Ecology, the Association DIRAC (France) and the Commissariat à l'Énergie Atomique. The IMS hydrophone data have been available through the CEA under contract (CEA n°169-C-BEFI & CNRS n°781513). The information released in this paper conforms to the conditions set forth in this contract. The

authors wish to thank Y. Cansi and G. Ruzié who provided access to the data used in this manuscript and L. Lieblein, A. Graff and P. Gaillard for the improvements in English usage. The authors also appreciate the valuable comments of anonymous reviewer which helped to improve the manuscript. This work represents a portion of F.S.'s dissertation.

REFERENCES

- Abdulla, W.H., D. Chow and G. Sin. 2003. Cross-words reference template for DTW-based speech recognition systems. Conference on Convergent Technologies for Asia-Pacific Region (TENCON). 4: 1576 – 1579.
- Branch T. A., Stafford K. M., Palacios D. M., Allison C., Bannister J. L., Burton C. L. K., Cabrera E., Carlson C. A., Vernazzani B. G., Gill P. C., Huckle Gaete R., Jenner K. C. S., Jenner M. N. M., Matsuoka K., Mikhalev Y. A., Miyashita T., Morrice M. G., Nishiwaki S., Sturrock V. J., Tormosov D., Anderson R. C., Baker A. N., Best P. B., Borsa P., Brownell R. L., Childerhouse S., Findlay K. P., Gerrodette T., Ilangakoon A. D., Joergensen M., Kahn B., Ljungblad D. K., Maughan B., Mccauley R. D., Mckay S., Norris T. F., Whale O., Rankin S., Samaran F., Thiele D., Van Waerebeek K., Warneke R. M. 2007. Past and present distribution of blue whale in the Southern Hemisphere and Northern Indian Ocean. *Mam. Rev.*37: 116-175.
- Clark, C.W. 1990. Acoustic behavior of mysticete whales. In: Thomas, J., Kastelein, R. (Eds.), *Sensory Abilities of Cetaceans*. Plenum Press, New York, 571-583.
- Clark, C.W., and D. K. Mellinger, 1994. Application of Navy IUSS for whale research. *J. Acoust. Soc. Am.* 96: 3315 (A).
- Costa, D.P. 1993. The secret life of marine mammals. Novel tools for studying their behaviour and biology at sea. *Oceanography*. 6: 120-128.
- Gagnon, G.J. and C.W. Clark. 1993. The use of U.S. Navy IUSS passive sonar to monitor the movements of blue whales. *Proc. Tenth Biennial Conf. Biol. Mar Mamm.* P.50.
- Gillespie, D. 2004. Detection and classification of right whale calls using an “edge” detector operating on a smoothed spectrogram. *Canad. Acoust.* 32: 39-47.
- Harvey, L.O.Jr 1992. The critical operating characteristic and the evaluation of expert judgment. *Organizational Behavior & Human Decision Processes*. 53(2): 229-251.
- Ljungblad, D., C.W. Clark and H. Shimada. 1998. A comparison of sounds attributed to pygmy blue whales (*Balaenoptera musculus brevicauda*) recorded south of the Madagascar Plateau and those attributed to ‘true’ blue whales (*Balaenoptera musculus*) recorded off Antarctica. *Rep. Int. Whal. Commn.* 49: 439–442.
- McDonald, M.A., S.L. Mesnick, and J.A. Hildebrand,. 2006. Biogeographic characterisation of blue whale song worldwide: using song to identify populations. *J. Cetacean Res. Manage.* 8(1): 55-65.
- Mellinger, D., and C.W. Clark. 2000. Recognizing transient *Canadian Acoustics / Acoustique canadienne*
- lowfrequency whale sounds by spectrogram correlation. *J. Acoust. Soc. Am.*107: 3518–3529.
- Mellinger, D., and C.W. Clark. 2003. Blue Whale (*Balaenoptera musculus*) sounds from the North Atlantic. *J. Acoust. Soc. Am.* 114(2): 1108-1119.
- Mellinger, D.K. 2004. A comparison of methods for detecting right whale calls. *Canad. Acoust.* 32: 55-65.
- Munger, L., D. Mellinger, S. Wiggins, S. Moore and J. Hildebrand. 2005. Performance of spectrogram cross-correlation in detecting right whale calls in long-term recordings from the bering sea. *Canad. Acoust.* 33(2): 25-34.
- Nieukirk S.L., K.M. Stafford, D.K. Mellinger, R.P. Dziak and C.G. Fox. 2004. Low-frequency whale and seismic airgun sounds recorded in the mid-Atlantic Ocean. *J. Acoust. Soc. Am.* 115(4):1832-1843.
- Nishimura C.E. and D.M. Conlon. 1994. IUSS Dual Use : Monitoring Whales and Earthquakes Using SOSUS. *Mar. Tech. Sci. J.* 27(4): 13-21.
- Rankin S., D. Ljungblad, C. Clark and H. Kato. 2005. Vocalizations of Antarctic blue whales, *Balaenoptera musculus intermedia*, recorded during the 2001/2002 and 2002/2003 IWC/SOWER circumpolar cruises, Area V, Antarctica. *J. Cetacean Res. Manage.* 7(1):13-20.
- Rebull, O.G., J.D. Cusi, M.R. Fernández and J.G. Muset. 2006. Tracking fin whale calls offshore the Galicia Margin, North East Atlantic Ocean. *J. Acoust. Soc. Am.* 120: 2077-2085.
- Roueff A., P-F. Piserchia, J.-L. Plantet, Y. Cansi and G. Ruzié. 2004. Noise characterization at the international station of Crozet Islands (HO4). *J. Acoust. Soc. Am.* 116: 2649.
- Silverman, H.F. and Morgan, D.P. 1990. The application of dynamic programming to connected speech recognition. *IEEE ASSP Magazine*, 7-25
- Širović, A., J.A. Hildebrand, S.M. Wiggins, M.A. McDonald, S.E. Moore and D. Thiele. 2004. Seasonality of blue and fin whale calls and the influence of sea ice in the Western Antarctic Peninsula. *Deep Sea Res. II.* 51 : 2327-2344.
- Širović, A., J.A. Hildebrand and S.M. Wiggins. 2007. Blue and fin whale call source levels and propagation range in the Southern Ocean *J. Acoust. Soc. Am.* 122: 1208-1215.
- Stafford K.M., L.N. Sharon and C.G. Fox. 2001. Geographic and seasonal variation of blue whale calls in the North Pacific. *J. Cetacean Res. Manage.* 3(1): 65-76.
- Stafford K.M. 2003. Two types of blue whale calls recorded in the Gulf of Alaska. *Mar. Mamm. Sci.* 19 (4): 682-693.
- Stafford K.M., D.R. Bohnenstiehl, M. Tolstoy, E. Chapp, D.K. Mellinger and S.E. Moore. 2004. Antarctic-type blue whale calls recorded at low latitudes in the Indian and eastern Pacific Oceans. *Deep Sea Res. I.* 51: 1337-1346.
- Stafford, K.M., D.W.R. Bohnenstiehl, E. Chapp. And M.

Tolstoy, 2005. Location, location, location: acoustic evidence suggests three geographic stocks of “Pygmy” Blue Whales in the Indian Ocean. Proceeding of the 16th Conference on the Biology of Marine Mammals. San Diego, December 12-16, 2005.

Stafford K.M., D.K. Mellinger, S.E. Moore and C.G. Fox. 2007. Seasonal variability and detection range modelling of baleen whale calls in the Gulf of Alaska, 1999-2002. J.

Acoust. Soc. Am. 122(6):3378-3390.

Trentin, E. and M. Gori. 2003. Robust combination of neural networks and hidden Markov models for speech recognition. IEEE Transactions on Neural Networks, 14(6): 1519 -1531.

Watkins, W.A. 1981. Activities and underwater sounds of fin whales. Scientific reports of the Whales Research Institute, 33: 83-117.

	Northern network			Southern network		
	Ambient noise level	BMi	BMb	Ambient noise level	BMi	BMb
Min	85.9(97.7)	84.3(94.6)	89.3(101.5)	100.1(113.8)	99.8(111.6)	104.4(116.6)
Max	106.2(121.0)	107.1(132.0)	119.4(133.2)	133.4(145.3)	117.8(137.6)	120.0(139.9)
Mean	92.6(104.5)	91.1(103.9)	97.2(111.3)	109.5(120.7)	103.9(115.4)	109.3(122.9)
SD	2.5(3.2)	3.2(4.2)	3.4(4.0)	6.4(6.6)	1.3(1.9)	3.2(3.9)

Table 1: Acoustic intensities (rms (peak) re 1µPa at 1m) (calculated on 1 month)

	One hour recording	All dataset
1 st frequency (Hz)	27.57	27.53
2 nd frequency (Hz)	19.35	19.36

Table 2: Difference between the templates modelling from 1-hour recordings and from the whole dataset.

	k_{ij}^{BMB1}			
	$i=0$	$i=1$	$i=2$	$i=3$
$j=0$	1...679	1...691	1...679	1...2790
$j=1$	68...1572	692...1686	680...1464	0
$j=2$	1573...2035	692...1686	1465...1989	0
$J=3$	1573...2035	1687...2790	1990...2790	0
$j=4$	2036...2790	0	0	0

Table 3: Duration of each Gaussian curve (0 for $k \neq k_{ij}^{BMB1}$) for the part 1

	a_{ij}^{BMB1}				b_{ij}^{BMB1}				c_{ij}^{BMB1}			
	$i=0$	$i=1$	$i=2$	$i=3$	$i=0$	$i=1$	$i=2$	$i=3$	$i=0$	$i=1$	$i=2$	$i=3$
$j=0$	2.944	3.153	3.309	7.63	385.2	398.8	464.2	484.8	211.6	270.7	198.9	147.8
$j=1$	12.02	4.215	3.001	0	956.8	920.9	946	0	210.5	189.7	268.8	0
$j=2$	7.359	2.08	0.7191	0	1335	1251	1706	0	189.3	324	218.3	0
$j=3$	6.306	1.148	1.262	0	1790	2135	2272	0	203.1	404.6	169	0
$j=4$	3.97	0	0	0	2295	0	0	0	203.1	0	0	0

Table 4: Parameters of the Gaussian curves for the BMB call part 1

	k_{ij}^{BMB3}				
	$i=0$	$i=1$	$i=2$	$i=3$	$i=4$
$j=0$	5296...8639	5296...6173	5296...7090	5296...6692	5296...5745
$j=1$	5296...8639	6174...8639	7091...8639	6693...8639	5746...8639

Table 5: Duration of each Gaussian curve (0 for $k \neq k_{ij}^{BMB3}$) for the part 3

	a_{ij}^{BMb3}					b_{ij}^{BMb3}					c_{ij}^{BMb3}				
	$i=0$	$i=1$	$i=2$	$i=3$	$i=4$	$i=0$	$i=1$	$i=2$	$i=3$	$i=4$	$i=0$	$i=1$	$i=2$	$i=3$	$i=4$
$j=0$	5.15	33.1	4.2	5.21	5.24	1537	590	940.9	714	338.3	1591	144.7	694.7	527.3	78.5
$j=1$	42.5	5.25	6.54	1.53	1.76	1010	1840	2640	2189	1289	242.8	949	516.2	757.9	917.1

Table 6: Parameters of the Gaussians for the BMb call part 3

	0.10	0.11	0.12	0.13	0.14
25	166(100,66)	143(100,43)	121(100,21)	110(100,10)	103(100,3)
20	167(100,67)	135(100,35)	121(100,21)	108(100,8)	101(100,1)
15	171(100,71)	148(100,48)	122(100,22)	107(100,7)	104(100,4)
10	169(100,69)	144(100,44)	121(100,21)	110(100,10)	104(100,4)
5	179(100,79)	146(100,46)	128(100,28)	109(100,9)	104(100,4)
0	167(100,67)	137(100,37)	122(100,22)	110(100,10)	105(100,5)
-5	156(100,56)	135(100,35)	119(100,19)	109(100,9)	105(100,5)
-10	175(100,75)	145(100,45)	127(100,27)	109(100,9)	105(100,5)
-15	165(96,69)	135(96,39)	119(96,23)	105(96,9)	102(96,6)
-20	170(87,83)	146(87,59)	128(87,41)	117(87,30)	112(87,25)
-25	157(72,85)	131(72,59)	117(70,47)	102(68,34)	96(65,31)
-30	155(27,133)	119(22,97)	81(18,63)	50(13,37)	30(11,19)

(a)

	0.15	0.16	0.17	0.18	0.19
25	101(100,1)	101(100,1)	100(100,0)	100(100,0)	100(100,0)
20	101(100,1)	101(100,1)	100(100,0)	100(100,0)	100(100,0)
15	102(100,2)	100(100,0)	100(100,0)	100(100,0)	100(100,0)
10	103(100,3)	101(100,1)	101(100,1)	100(100,0)	100(100,0)
5	102(100,2)	101(100,1)	100(100,0)	100(100,0)	100(100,0)
0	102(100,2)	100(100,0)	100(100,0)	100(100,0)	100(100,0)
-5	103(100,3)	100(100,0)	100(100,0)	100(100,0)	100(100,0)
-10	101(100,1)	101(100,1)	100(100,0)	100(100,0)	100(100,0)
-15	101(96,5)	100(96,4)	100(96,4)	100(96,4)	100(96,4)
-20	105(87,18)	101(87,14)	101(87,14)	100(87,13)	99(86,13)
-25	84(62,22)	67(50,17)	50(35,15)	38(25,13)	25(17,8)
-30	15(9,6)	11(7,4)	5(4,1)	3(3,0)	2(2,0)

(b)

Table 7: Evaluation of the detection threshold value (lines show the threshold values and columns show SNR (dB)). Number of total detections (correct detections, false alarms)

Matched filter used	Northern network		Southern network	
	BMi	BMb	BMi	BMb
Template	6313	7 495	2856	717
Real call with high SNR	2971	1116	2082	93
Real call with low SNR	148	130	461	28

Table 8: Number of calls detected by using template and real call in various SNR for matched filter (calculated on 1 month)

	Northern network		Southern network	
	BMi	BMb	BMi	BMb
Number of calls detected	6313	7 495	2856	717
Min	0.17	0.14	0.17	0.14
Max	0.72	0.55	0.49	0.45
Mean	0.23	0.23	0.19	0.20
SD	0.07	0.07	0.02	0.06

Table 9: Number of calls detected and correlation coefficient (calculated on 1 month)

Why Purchase from a Single Manufacturer... ...When You Can Have the Best in the Industry From a Single Supplier?

Scantek is the company to call when you want the most comprehensive assortment of acoustical and vibration equipment. As a major distributor of the industry's finest instrumentation, we have the right equipment at the right price, saving you time and money. We are also your source for instrument rental, loaner equipment, product service, technical support, consulting, and precision calibration services.

Scantek delivers more than just equipment. Since 1985, we have been providing solutions to today's complex noise and vibration problems with unlimited technical support by acoustical engineers that understand the complex measurement industry.

Suppliers of Instruments and Software:

- Norsonic
- RION
- CESVA
- DataKustik (Cadna & Bastian)
- KCF Technologies
- BSWA
- Castle Group
- Metra
- RTA Technologies
- G.R.A.S.

Scantek
Sound and Vibration
Instrumentation and Engineering

Applications:

- Building Acoustics & Vibration
- Occupational Noise and Vibration
- Environmental and Community Noise Measurement
- Sound Power Testing
- Calibration
- Acoustical Laboratory Testing
- Loudspeaker Characterization
- Transportation Noise
- Mechanical Systems (HVAC) Acoustics

Scantek, Inc. • 7060 Oakland Mills Road • Suite L • Columbia, MD 21046 • 800•224•3813 • www.scantekinc.com

DETECTION AND LOCALIZATION OF BLUE AND FIN WHALES FROM LARGE-APERTURE AUTONOMOUS HYDROPHONE ARRAYS: A CASE STUDY FROM THE ST. LAWRENCE ESTUARY

Yvan Simard^{1,2} and Nathalie Roy²

¹Marine Sciences Institute, University of Québec at Rimouski, P.O. Box 3300, Rimouski, Québec G5L-3A1, Canada.

²Maurice Lamontagne Institute, Fisheries and Oceans Canada, P.O. Box 1000, Mont-Joli, Québec G5H-3Z4, Canada

ABSTRACT

The feasibility of using passive acoustic methods (PAM) to monitor time-space distribution of fin and blue whales in the Saguenay–St. Lawrence Marine Park was explored using large-aperture sparse hydrophone arrays. The arrays were deployed during summers 2003 to 2005 at the head of the 300-m deep Laurentian Channel. They were composed of 5 AURAL autonomous hydrophones moored at mid-water depths, near the summer sound channel. A small coastal array complemented the deployment in 2003. The apertures were from 20 to 40 km and the configurations were changed from year to year. The most frequent calls recorded were blue and fin whale signature infrasounds. Noise from transiting ships on the busy St. Lawrence Seaway often masked the calls on the nearest hydrophones. Sometimes this resulted in an insufficient number of receivers for localizing the whales using time difference of arrival (TDoA) methods. The technical characteristics of the arrays and data processing are presented, with an example of call detection and localization. Despite the difficulties inherent to this environment, PAM can be effectively implemented there, eventually for real-time operations.

RÉSUMÉ

La faisabilité d'utiliser la technologie de monitoring acoustique passif (PAM) pour suivre la distribution spatio-temporelle des rorquals bleus et communs dans le Parc Marin Saguenay–Saint-Laurent a été explorée à l'aide de réseaux d'hydrophones à maille lâche couvrant de grandes distances. Les réseaux ont été déployés pendant les étés 2003 à 2005 à la tête du chenal Laurentien, profond de 300 m. Ils étaient composés de 5 hydrophones autonomes AURAL mouillés à mi-profondeur, près du couloir de son estival. Un petit réseau côtier de faible ouverture complétait le déploiement en 2003. Les ouvertures des réseaux étaient de 20 à 40 km et leurs configurations étaient changées à chaque année. Les vocalisations les plus fréquentes étaient les infrasons identitaires des rorquals bleus et communs. Le bruit de navires transitant dans la Voie Maritime achalandée du Saint-Laurent masquait souvent les vocalisations sur les hydrophones les plus proches, ce qui parfois résultait en un nombre insuffisant de récepteurs pour localiser les baleines à l'aide de méthodes utilisant les différences de temps d'arrivée (TDoA). Les caractéristiques techniques des réseaux et du traitement des données sont présentées avec un exemple de détection et de localisation. Malgré les difficultés inhérentes à cet environnement, la technologie PAM peut y être efficacement implémentée, éventuellement pour des opérations en temps réel.

1. INTRODUCTION

The development of the methodology for localising whales from their sounds in their habitats was initiated by Watkins and Schevill (1972) in the 1970s. It was then rapidly applied to tracking whales over large distances (e.g. Cummings and Holliday 1985, Clark et al. 1986). Advances in electronics, computers and numerical analysis now make this PAM technology more accessible and affordable to small research budgets. Various systems have been used, including shore-cabled and radio-linked systems, drifting buoys, and arrays of autonomous recorders for versatile and long-term deployments (e.g. Janik et al. 2000, Hayes et al. 2000,

Watkins et al. 2000, Tiemann and Porter 2004, Simard et al. 2004, Sirovic et al. 2007, Stafford et al. 2007). The goal of such PAM systems, is the continuous mapping of presence and distribution of whales over ocean basins (e.g. Greene et al. 2004, Simard et al. 2004, Sirovic et al. 2007, Stafford et al. 2007) and assessing their densities, (e.g. Ko et al. 1986, McDonald and Fox 1999, Clark and Ellison 2000), sometimes in quasi real-time (e.g. Thiemann and Porter 2004). Their performance in effectively accomplishing these tasks, depends on the characteristics of the targeted whale calls, the environment, the type of equipment used, its deployment and configuration. This performance may significantly vary from case to case.

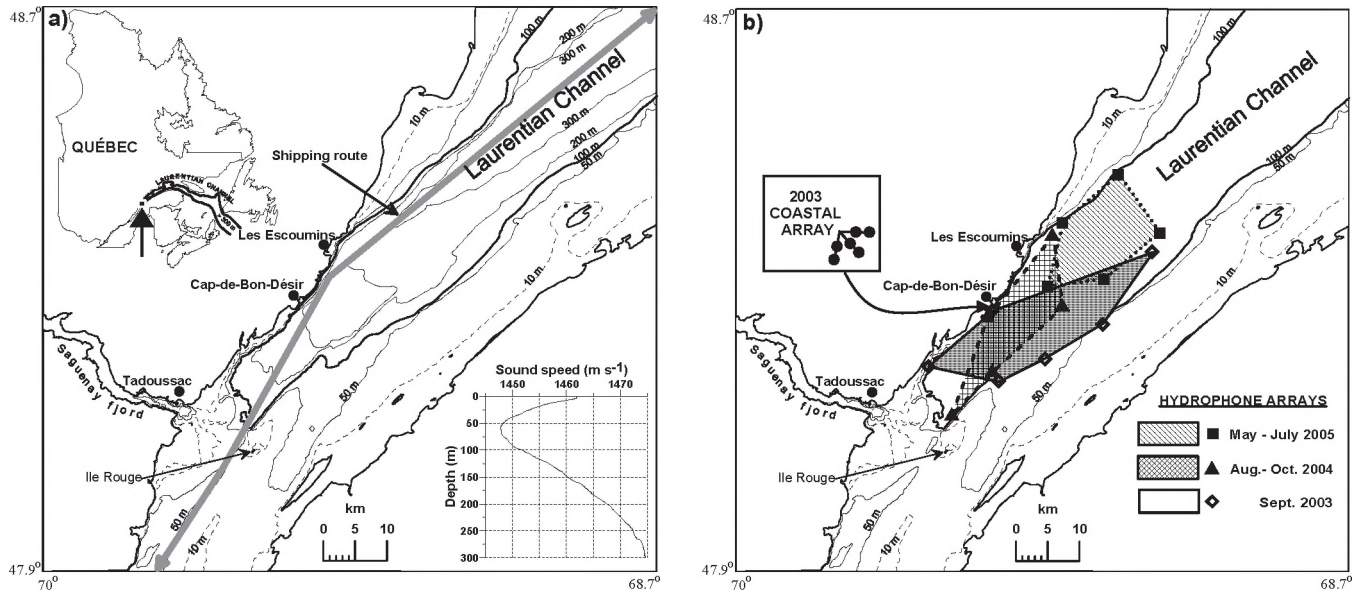


Figure 1. a) Study area at the head of the Laurentian channel in the Saguenay–St. Lawrence Marine Park, with bathymetry and typical summer sound speed profile. b) Configurations of the hydrophone arrays deployed in 2003–2005.

PAM's success first depends on the capacity to isolate the targeted calls from the rest of the acoustic signal in which they are imbedded, especially for distant sources and low signal to noise ratios (SNR). Call source level (SL), propagation loss, and the local "ocean noise" level determine detection ranges (c.f. Sirovic et al. 2007, Stafford et al. 2007). Whale calls' SLs vary considerably among species and within a species' vocal repertoire (e.g. Kuperman and Roux 2007, p. 199.). Ocean noise level also exhibits considerable variability in space and time, in response to fluctuating natural sources, such as wind, ice, rain, sounds produced by various organisms, and anthropogenic sources such as shipping (c.f. review NRC 2003). When a series of hydrophones are available at each node of the larger PAM array, beamforming and matched-field processing (c.f. Jensen et al. chap. 10) can improve signal detection by SNR enhancement. Signal processing can improve detection of some calls by exploiting their distinctiveness in time-frequency space compared to noise (e.g. Mellinger and Clark 2000). Sound speed structures over the water column can focus sounds from distant sources into sound channels, thereby reducing propagation loss from multiple interactions with absorptive and scattering surface and bottom interfaces. This is true for both the signal and the noise sources. The signal with the lowest transmission loss depends on the 3D spatial arrangements of the sources and the local propagation characteristics. The spatial arrangement, horizontal distance between the hydrophones, and their depth relative to the sound channel are relevant to the PAM problem. The optimal configuration could be explored from simulation models.

SNR not only affects the detection of calls, but also the capacity to precisely estimate their TDoAs on the hydrophone array (Clark and Ellison 2000, Buaka Muanke and Niezrecki 2007). High precision is essential for precise localisation (Spiesberger and Wahlberg 2002, Spiesberger 2004, 2005). Precise estimation of the TDoAs is hindered by low SNR and multipath propagation conditions where reflected and refracted signals overlap. TDoA accuracy also depends on proper synchronisation of the array, which is often problematic with the multiple independent clocks of autonomous hydrophone arrays (e.g. Thode et al. 2006, Sirovic et al. 2007).

Additional constraints for operational PAM setups include minimizing interfering noise from the hydrophone deployment accessories such as strumming from the mooring. Low-frequency vibration and flow noise (Haddle and Skudrzyk 1969) can arise due to strong currents often encountered on continental shelf habitats where whales vocalizing at low frequencies forage on aggregated preys (e.g. Simard and Lavoie 1999).

Examples of PAM applications used to non-intrusively study whales in their large-scale habitat from a sparse array of distant omnidirectional hydrophones are expanding around the world. Details of experiments from several case studies in different environments should help improve the development, efficient use, and robustness of this new methodology. The present paper contributes to this effort by presenting an example for blue and fin whale localization in the Saguenay–St. Lawrence Marine Park (SSLMP) located at the head of the Laurentian Channel in the Lower St. Lawrence Estuary (Fig. 1).

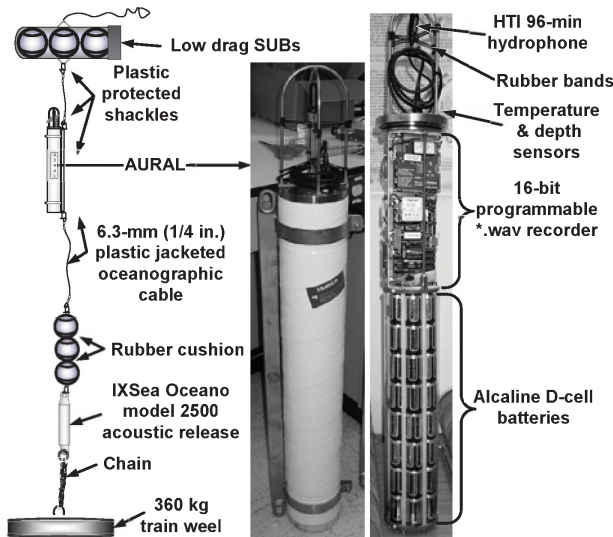


Figure 2. Typical mooring of an AURAL autonomous hydrophone.

2. MATERIAL AND METHODS

Study site

For centuries, North-West Atlantic baleen whales have migrated to the head of the 300-m deep Laurentian Channel during summer for feeding on prey concentrated along its bordering steep slopes by strong tidal upwelling processes (Simard and Lavoie 1999, Lavoie et al. 2000, Simard et al. 2002, Cotté and Simard 2005). The summer water column in this part of the North-West Atlantic is characterized by a prominent Cold Intermediate Layer (CIL) centered around 60 m, creating a well defined sound channel at these depths (Fig. 1a). The bottom is composed of more than 200 m of silt overlying the bedrock in the trough with sand and gravel on the surrounding shallow areas. The high tidal energy of this environment generates fronts, semidiurnal upwelling at the channel head, and propagating internal tide and high-frequency internal waves, moving the CIL depth by up to 100 m (c.f. Saucier and Chassé 2000). These processes modify the propagation conditions in time and space, notably by swinging the sound channel up and down. Shipping noise from St. Lawrence Seaway traffic is high. Levels in the 18-22.6 Hz and 35.6-89.8 Hz targeted call bands can reach 130 dB re 1 $\mu\text{Pa}_{\text{rms}}$ and exceed 102 dB re 1 $\mu\text{Pa}_{\text{rms}}$ more than 50% of the time (Simard et al., unpublished results from 15960 h of recordings).

Equipment

The PAM arrays were deployed in the study area during summers of 2003 to 2005 (Fig. 1b). They were made up of 5 AURAL autonomous hydrophones (Multi-Electronique Inc, Rimouski, Qc, Canada) programmed for 16-bit continuous sampling (M1-mode) after 17 or 23 dB amplification. The AURALS also recorded the ambient temperature and depth. The temperature compensated crystal oscillators of their clocks minimized temperature effects on clock drifts. The instruments were anchored with typical oceanographic

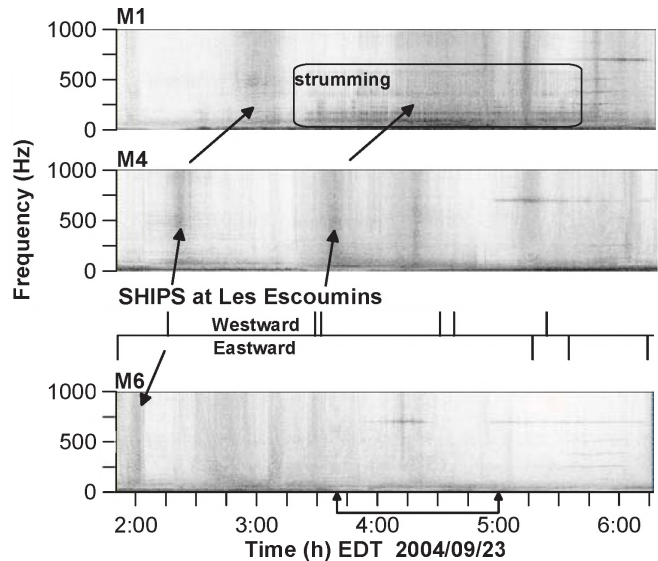


Figure 3. Spectrograms showing typical non-synchronized noise patterns while ship transits throughout the study area from upstream (M1) to downstream (M6) with their reporting time at the pilot station, 10 km upstream of M4. Period corresponding to Fig. 4 is pointed on X axis.

moorings, taking special care to minimize the noise from the mooring components (Fig. 2). All hydrophones were HTI 96-min (High Tech Inc., Gulport, Ms, USA) with a nominal receiving sensitivity (RS) in the low frequency band (< 2 kHz) of -164 dB re $1 \text{ V}/\mu\text{Pa}$, confirmed by calibration at the Defense Research Development Canada (Dartmouth, NS, Canada) facility. The hydrophones were placed at intermediate depths in the water column near the summer sound channel axis (Fig. 1a). After the first deployment in 2003, the hydrophones were deployed in deeper water farther from the channel slopes in order to avoid local maximum tidal currents (c.f. Lavoie et al. 2000, Saucier and Chassé 2000) that were generating vibrations of the mooring and flow noise. In 2003, a coastal array of 6 HTI 96-min hydrophones with an aperture of ~ 650 m was also deployed along a cape in the middle of the study area (Fig. 1b). The acquisition system consisted of a 16-bit ChicoPlus Servo-16 data acquisition board (Innovative Integration, Simi Valley, CA, U.S.A.) connected to a PC. The exact locations of these hydrophones on the bottom were determined from hyperbolic fixing (receiver and sources inverted) by sending series of 8-kHz pulses from the IXSea Oceano acoustic release transmitter (Marly-le-Roy, France) from a network of surrounding stations surveyed by the R/V Coriolis II. CTD profiles (SBE 19, Seabird Electronics, Bellevue, Wa., USA) from the study area were used to compute sound speed profiles.

Synchronization

The synchronization of the autonomous hydrophones exploited a combination of means: starting and stopping the AURALS with a PPS (pulse per second) impulse from a GPS receiver, simultaneous recording of same acoustic

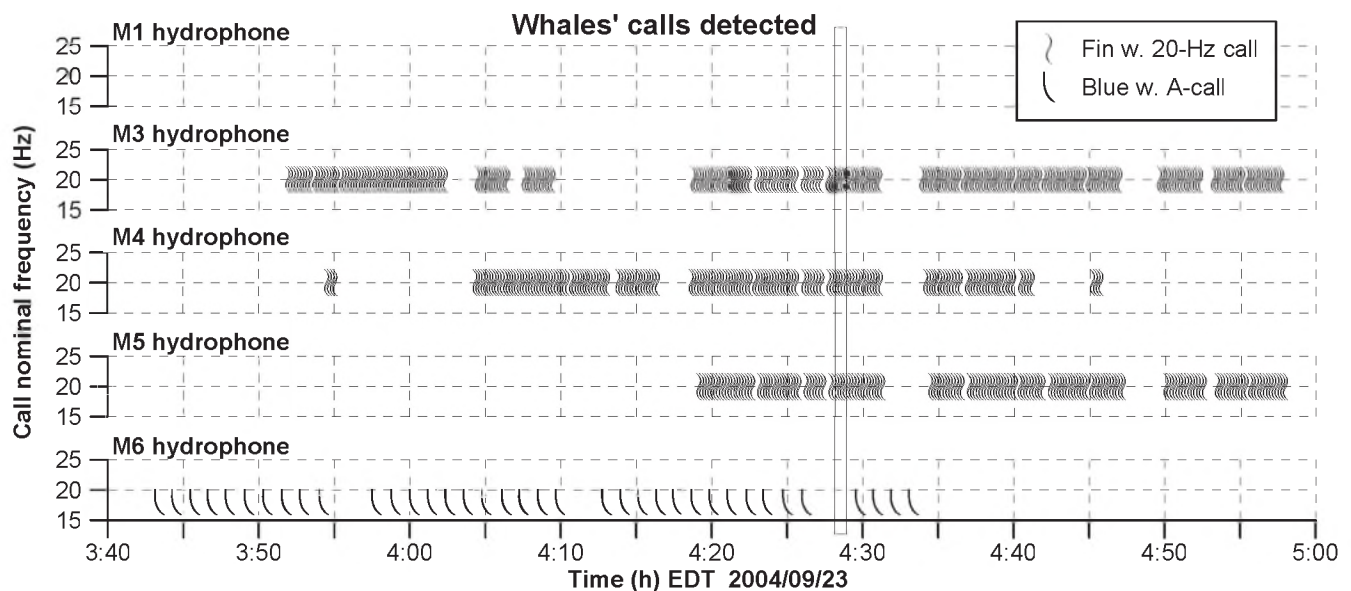


Figure 4. Blue and fin whale calls series detected on the hydrophone array during the period pointed on Fig. 3. The bar marks the 1-min sequence used for the localization shown in Figs. 5, the hyperbole traces and isodiachron clouds of Fig. 6.

signals on all units, cross-checking with the coastal array clock, and linear interpolations assuming constant drift of the internal M1-mode clocks over the deployment periods. This drift was quite stable from one instrument to the other and estimated to $4.1 \pm 0.1 \text{ s d}^{-1}$. The relative drift for estimating the TDoAs at the hydrophones was $< 0.2 \text{ s d}^{-1}$. It was also consistent over years, and therefore seems to be a characteristic of the particular instrument's clock. Other synchronization approaches could be tested. (see Discussion).

Data analysis

Call detection requires initial SNR enhancement by detrending the spectrogram as the first noise filtration step (details in Mellinger 2004, Mouy 2007). Fixed-template time-frequency call detection algorithms (e.g. Mellinger and Clark 2000) then generally performed well for the stereotyped infrasound calls of blue (A and B calls) and fin (20-Hz pulse) whales (Mouy 2007). A time-frequency contour detection algorithm combined with DTW (dynamic time warping) classification algorithm (*ibid.*) was used for the variable blue whale D call (Berchok et al. 2006). TDoA estimation was generally easier using spectrogram cross-coincidence (i.e. computing the time lag required to best match the call blueprint on the binary images of the spectrograms at hydrophone pairs from a logical AND on the pixel values of 0 or 1, e.g. Simard et al. 2004, Fig. 5) than by cross-correlating the filtered signal in time domain because of noise interference. Localization was performed by hyperbolic fixing (Spiesberger and Fristup 1990), isodiachrons and Monte-Carlo simulations (Spiesberger and Whalberg 2002, Spiesberger 2004), and by an acoustic propagation model (Tiemann and Porter 2004) (details in Roy et al. 2008).

3. RESULTS

Ships transiting in both directions along the study area increased noise over the whole spectrum for 0.5-1 h around the ships' closest point of approach to the hydrophones (Fig. 3). During ships' ~ 3 -h transits, their intense noise successively polluted the hydrophones of the array along their route. At times, strong currents induced strumming and flow noise that polluted the calls' band at tidal peaks, thus negatively impacting the hydrophones located in the maximum flow.

Call time series for the 80-min period marked on Fig. 3 show a 26-min sequence where 3 hydrophones detected 20-Hz fin whale calls (Fig. 4, 4h19 to 4h45). This call series comprised two ~ 15 -min bouts separated by 3 min. The calls are repeated at ~ 11 -s intervals, but occasional calls, named backbeats, lag their preceding call by ~ 17 -s (Fig. 5; c.f. Samaran 2004), a characteristic that can help confirm adequate time alignments. TDoAs were estimated from spectrogram cross-coincidence for 21 1-min sequences (e.g. Fig. 5). The whale was found to be close to M4 hydrophone on the northern slope of the Channel; the locations on land allowed easy removal of ambiguous localizations (Fig. 6). The whale showed slight displacements ($< 0.5 \text{ km}$ from hyperbolic fixing) during this 26-min period from both localization methods. The localization uncertainty was $\sim 1.2 \text{ km}$ from the radius of the isodiachron Monte-Carlo localization cloud (4000 simulations taking into account an error of 20-m in hydrophone position, 0.5 s in TDoAs, and 5 m s^{-1} in effective sound speed). The mean distance between the locations of the peak density of the isodiachron Monte-Carlo simulations for each of the 21 sequences of 1 min and the hyperbolic fixing solutions was 100 m (SD = 32 m).

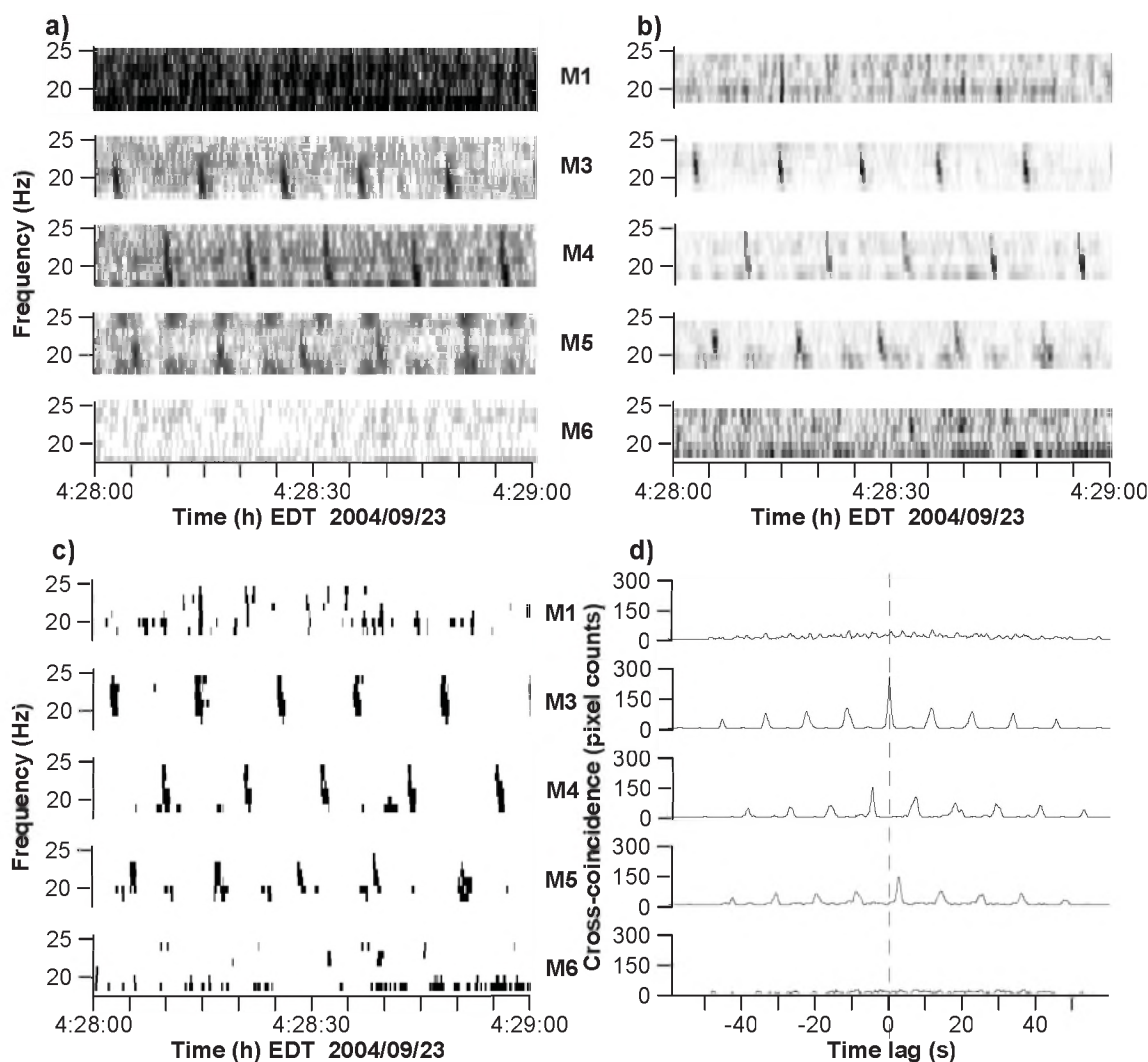


Figure 5. Example of signal processing for TDoA estimation for the 1-min segment pointed on Fig 4.: a) Spectrograms of the calls where same relative dB palette is applied to all hydrophones; analysis window 0.1 s, frequency resolution 1 Hz. b) Images of the spectrograms in a) where the palette is specific to the hydrophone and frequency view, and the extreme frequencies were trimmed out. c) Binary images of b) retaining only the strongest 5% intensities. d) Cross-coincidence of images in c) resulting from a logical AND operation using M3 hydrophone as reference to estimate TDoAs from peak values.

4. DISCUSSION

Despite the difficulty of applying PAM to track whales in a highly-fluctuating and noisy environment such as the head of the Laurentian Channel, the results show that it is feasible for a significant proportion of the time, using a sparse array of hydrophones. Even though detection and localization was not always possible because of masking noise, the frequent vocalisations and the low displacement rate of the whales allowed their mapping with a reasonably good resolution in time and space. With a good knowledge of the oceanographic, propagation and noise characteristics of the study area, it is possible to effectively implement PAM technologies to monitor whales in this meso-scale basin over long periods from a sparse array of autonomous hydrophones.

Further attention should be given to determining the optimal hydrophone density and 3D spatial arrangement, regular clock synchronisation, and minimization of masking from mooring strumming, vibrations and flow noise. Covering the hydrophone with open-cell foam or membranes might reduce flow noise for short-term deployments, but bio-fouling negates their long-term use. Choosing the hydrophone location after considering currents' 3D spatial structure helped to reduce the problem. Hydrophones less sensitive to vibrations and flow noise, but still affordable to limited research budgets, would be desirable. Directional sensors (e.g. Greene et al. 2004) able to simply and accurately determine the source direction under noise conditions should also help improving PAM efficiency.

Our simple synchronization approach was successful but other more elaborated methods could be explored. Spiesberger (2005) proposed a Monte-Carlo method to

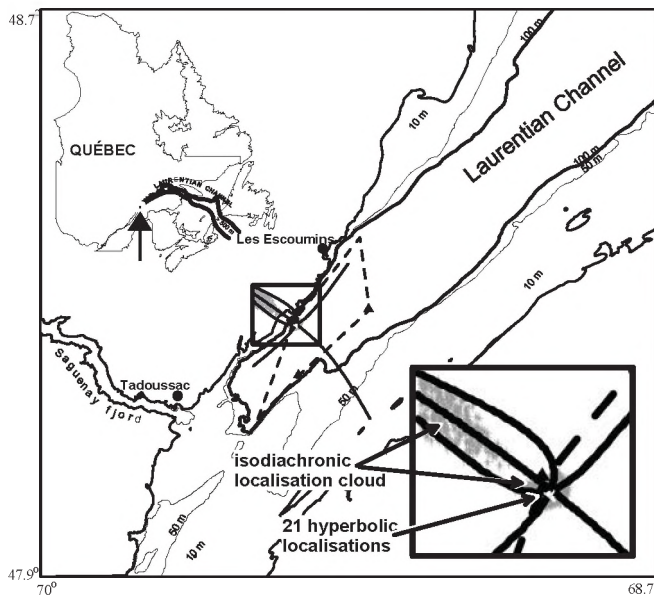


Figure 6. Localization of the 20-Hz call series shown in Fig. 4. Hyperbolic localizations every 1-min intervals between 4h19 and 4h45 (white line). The solution at 4h28 is shown with hyperboles and the isodiachron clouds of possible localizations (in gray), whose peak density is 54 m away from the hyperbolic solution.

assess probability distributions of all localization variables, including TDoAs. This approach could be adapted to track relative clock drifts at different times from the differences between the observed and estimated TDoAs of a series of independent sources recorded by the array. Thodes et al. (2006) proposed a matched-field modeling approach for synchronizing a small line array of autonomous hydrophones by simultaneous geoaoustic inversions for both whale localizations and clock offsets, combined with cross-correlation of diffuse background noise.

To choose the optimal array configuration for the study area, propagation modeling could be used to provide detection and localization probability maps under the observed local noise probability density function in the call bands and published SLs for the targeted signature calls. Augmenting the hydrophone density of the array to minimize masking by shipping noise, which depends on the relative distance between the ship and whale and their SLs difference (c.f. Simard et al. 2006a), appears the simplest way of enhancing the detection and localisation probability over the whole study area.

Efficient signal processing algorithms are required to minimise noise and multipath interferences in detecting and identifying the calls. Time-frequency domain algorithms with adequate resolution proved to be effective at this task as well as for reasonably estimating TDoAs. Localization error will never be eliminated due to imprecision associated with the input variables. Isodiachronic Monte-Carlo localization proved helpful to assess the extent of this localization uncertainty, but robust error estimation methods need further research (Roy et al. 2008). Tracking of a fixed sound source emitting at regular intervals appears highly

suitable to accurately monitor the localization error, especially under such variable environments.

PAM information was not available in real-time but only after the recovery of the array at the end of the observation period. Real-time PAM (e.g. Tiemann and Porter 2004) is often required for management and protection purposes, mitigation of anthropogenic activities, and implementation of early warning systems. Such low-cost telecommunicating real-time detection, classification and localization systems are presently in development and experimentation (Simard et al. 2006b) and could eventually become versatile alternatives to cabled real-time PAM systems.

5. ACKNOWLEDGEMENTS

This work was funded by the authors' supporting institutions. We thank all technicians, students, assistants and partners who contributed to its realization.

6. REFERENCES

- Buaka Muanke, P., and Niezrecki, C. 2007. Manatee position estimation by passive acoustic localization. *J. Acoust. Soc. Am.* 121: 2049-2059.
- Berchok, C.L., Bradley, D.L., and Gabrielson, T.B. 2006. St. Lawrence blue whale vocalizations revisited: Characterization of calls detected from 1998 to 2001. *J. Acoust. Soc. Am.*, 120, 2340-2354.
- Clark, C.W., Ellison, W.T., and Beeman, K. 1986. Acoustic tracking of migrating bowhead whales. *Oceans* 86: 341-346.
- Clark, C.W., and Ellison, W.T. 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.* 197: 3509-3517.
- Cotté, C., and Simard, Y. 2005. The formation of rich krill patches under tidal forcing at whale feeding ground hot spots in the St. Lawrence Estuary. *Mar. Ecol. Progr. Ser.* 288: 199-210.
- Cummings, W.C., and 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.
- Greene, C.R., McLennan, M.W., Norman, R.G., et al. 2004. Directional frequency and recording (DIFAR) sensors in seafloor recorders to locate calling bowhead whales during their fall migration. *J. Acoust. Soc. Am.*, 116, 799-813.
- Haddle, G.P., and Skudrzyk, E.J. 1969. The physics of flow noise. *J. Acoust. Soc. Am.* 46: 130-157.
- Hayes, V.M., Mellinger, D.A., Croll, D.A., et al. 2000. An inexpensive passive acoustic system for recording and localizing wild animal sounds. *J. Acoust. Soc. Am.* 107: 3552-3555.
- Janik, V.M., Van Parijs, S.M., and Thompson, P.M. 2000. A two-dimensional acoustic localization system for marine mammals. *Mar. Mammal Sci.* 16: 437-447.
- Ko, D., Zeh, J.E., Clark, C.W., et al. 1986. Utilization of acoustic location data in determining a minimum number of spring-

- migrating bowhead whales unaccounted for by the ice-based visual census. *Rep. Int. Whaling Comm.* 36: 325-338.
- Kuperman, W.A., and Roux, P. 2007. Underwater acoustics. In Rossing, T.D. (ed.) *Handbook of acoustics*. Springer, N.Y. pp. 149-201.
- Lavoie, D., Simard, Y., and Saucier, F.J. 2000. Aggregation and dispersion of krill at channel heads and shelf edges: the dynamics in the Saguenay–St. Lawrence Marine Park. *Can. J. Fish. Aquat. Sci.* 57: 1853-1869.
- Matthews, J.N., Rendell, L.E., Gordon, J.C.D., MacDonald D.A. 1999. A review of frequency and time parameters of cetacean tonal calls. *Bioacoustics* 10: 47-71.
- McDonald, M.A., and C.G. Fox. 1999. Passive acoustic methods applied to fin whale population density estimation. *J. Acoust. Soc. Am.* 15: 2643-2651.
- Mellinger, D. K. 2004. A comparison of methods for detecting right whale calls. *Canadian Acoustics*, 32(2):55–65.
- Mellinger, D.K., and Clark, C. W. 2000. Recognizing transient low-frequency whale sounds by spectrogram correlation. *J. Acoust. Soc. Am.* 107: 3518-3529.
- Mouy, X. 2007. Détection et identification automatique en temps-réel des vocalises de rorqual bleu (*Balaenoptera musculus*) et de rorqual commun (*Balaenoptera physalus*) dans le Saint-Laurent. M.Sc. Thesis, Univ. du Québec à Rimouski, Rimouski, Qc, Canada.
- NRC, 2003. Ocean noise and marine mammals. The National Academies Press. Washington, D.C.
- Roy, N., Simard, Y., and Rouat, J. 2008. Performance of three acoustical methods for localizing whales in the Saguenay–St. Lawrence Marine Park. *Canadian Acoustics* 00: 000-000 (this issue)
- Samaran, F. 2004. Déteabilité des vocalisations de rorquals communs (*Balaenoptera physalus*) à partir d'une station côtière dans la voie maritime de l'estuaire du Saint-Laurent. M.Sc. Thesis Univ. du Québec à Rimouski, Rimouski, Qc, Canada.
- Saucier, F.J., and Chassé, J. 2000. Tidal circulation and buoyancy effects in the St. Lawrence estuary. *Atmosphere-Ocean* 38: 505-556.
- Simard, Y., Bahoura, M., and Roy, N. 2004. Acoustic detection and localization of baleen whales in Bay of Fundy and St. Lawrence Estuary critical habitats. *Canadian Acoustics* 32(2), 107-116.
- Simard, Y., and Lavoie, D. 1999. The rich krill aggregation of the Saguenay—St. Lawrence Marine Park: hydroacoustic and geostatistical biomass estimates, structure, variability and significance for whales. *Can. J. Fish. Aquat. Sci.* 56: 1182-1197.
- Simard, Y., Lavoie, D., and Saucier, F.J. 2002. Channel head dynamics: Capelin (*Mallotus villosus*) aggregation in the tidally-driven upwelling system of the Saguenay–St. Lawrence Marine Park's whale feeding ground. *Can. J. Fish. Aquat. Sci.* 59: 197-210.
- Simard, Y., Roy, N., and Gervaise, C. 2006a. Shipping noise and whales: World tallest ocean liner vs largest animal on earth. *in OCEANS'06 MTS/IEEE – Boston, IEEE Cat. No. 06CH37757C*, Piscataway, NJ, USA. 6 p.
- Simard, Y., Bahoura, M., Park, C.W., et al. 2006b. Development and experimentation of a satellite buoy network for real-time acoustic localization of whales in the St. Lawrence. *in OCEANS'06 MTS/IEEE – Boston, IEEE Cat. No. 06CH37757C* Piscataway, NJ, USA. 6 p.
- Sirovic, A., Hildebrand, J.A., and Wiggins, S.M. 2007. Blue and fin whale call source levels and propagation range in the Southern Ocean. *J. Acoust. Soc. Am.* 122: 1208-1215.
- Spiesberger, J.L. 2004. Geometry of locating sounds from differences in travel time: Isodiachrons. *J. Acoust. Soc. Am.* 112: 3046-3052.
- Spiesberger, J.L. 2005. Probability distributions for locations of calling animals, receivers, sound speeds, winds, and data from travel time differences. *J. Acoust. Soc. Am.* 118: 1790-1800.
- 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.
- Stafford, K.M., Mellinger, D.K., Moore, S.E. and Fox, C.G. 2007. Seasonal variability and detection range modeling of baleen whale calls in the Gulf of Alaska, 1999–2002. *J. Acoust. Soc. Am.* 122: 3378-3390.
- Thode, A., Gerstoft, P., Burgess, W., et al. 2006. A portable matched-field processing system using Passive acoustic time synchronization. *IEEE J. Ocean. Eng.* 31: 696-710
- Tiemann, C.O. and Porter, M.B. 2004. Localization of marine mammals near Hawaii using an acoustic propagation model. *J. Acoust. Soc. Am.* 115: 2834-2843.
- Watkins, W.A., and Schevill, W.E. 1972. Sound 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, et al. 2000. Seasonality and distribution of whale calls in the North Pacific. *Oceanography* 13: 62-67.

DETECTION AND RECOGNITION OF NORTH ATLANTIC RIGHT WHALE CONTACT CALLS IN THE PRESENCE OF AMBIENT NOISE

Ildar R. Urazghildiiev, Christopher W. Clark and Timothy P. Krein
Cornell Lab of Ornithology, 159 Sapsucker Woods Road, Ithaca, NY 14850

ABSTRACT

The problem of detection and recognition of contact calls produced by North Atlantic right whales, *Eubalaena glacialis*, is considered. A proposed solution is based on a multiple-stage hypothesis-testing technique involving a spectrogram-based detector, spectrogram testing, and feature vector testing algorithms. Results show that the proposed technique is able to detect over 80% of the contact calls detected by a human operator and to produce about 26 false alarms per 24 h of observation.

SOMMAIRE

Un problème de détection et reconnaissance des baleines noires, *Eubalaena glacialis*, en présence de bruit ambiant est étudié. Une solution proposée est basée sur une technique de test d'hypothèses en plusieurs étapes, impliquant le détecteur, des tests de spectrogramme et des algorithmes testant des vecteurs de traits. Les résultats des tests montrent que la technique proposée est capable de détecter plus de 80% des appels de contact détectés par les opérateurs humains et de produire environ 26 fausses alarmes par 24 h d'observation.

1. INTRODUCTION

Continuous monitoring of North Atlantic right whales (NARW) presence in large areas can be accomplished by passive acoustical methods using data recordings obtained from distributed autonomous hydrophone systems [1-4]. Such systems yield enormous data sets totaling many years of potential listening time, presenting an analytical challenge. Using human operators to visually and aurally evaluate data spectrograms is impractical in projects that collect huge amounts of data. Apart from this, the human operators often provide subjective and inaccurate estimates [5] so the design of effective, automated detection techniques is of critical importance.

To reduce subjectivity and to decrease the labor costs, various NARW detection methods known from the literature can be used (see e.g., [6-11]). These methods can potentially improve the detection efficiency by rejecting a huge portion of the data that contains no signal. However, as test results demonstrate, known methods do not provide the required trade-off between the probabilities of detection and false alarm. In particular, for the detection probability of 0.8, the lowest level of false alarm probability provided by the spectrogram-based detector is from 10^{-2} to 10^{-3} , depending on the impulsive noise rate [11]. For the NARW contact calls with the typical duration of 1 s, the range of probability of false alarm corresponds to 100–1000 false detections per 24 h of observation. Since all the detection events should be

evaluated by a human operator, the labor costs are significant.

The goal of the research presented in this paper is to reduce the probability of false alarm in spectrogram-based detectors without negatively affecting the detection probability. The proposed technique is reduced to a multiple-stage hypotheses-testing process. In the initial stage, the spectrogram-based detector [11] is applied. The data segments accepted as signals in the initial stage are recognized using the proposed recognition technique. The hypothesis that the detected segment belongs to the known types of impulsive noise is tested in the second stage. If this hypothesis is rejected, a feature vector (FV) is extracted and tested in the final stage. Test results obtained using real data recordings are presented.

2. DATA MODEL AND PROBLEM FORMULATION

We use the data model similar to that considered in [11]. The NARW contact calls are modeled as polynomial-phase signals (PPS). Ambient noise is represented as a Gaussian process contaminated by unknown impulsive processes. A typical spectrogram of the input data containing a NARW contact call, background noise, impulsive noise and self-noise is shown in Fig. 1 (top frame).

We assume that the spectrogram-based detector is applied to the input data in the initial stage. For each 1 s data segment

$x(t)$, the detector tests the hypotheses H_0 ("ambient noise is present") and H ("signal and ambient noise are present"). The decision-making process is reduced to computing the statistic $z(t) = z(x(t))$ as a function of the tested data segment $x(t)$. The statistic $z(t)$ is compared with a threshold C_D and the hypothesis H is accepted if $z(t) \geq C_D$. The detector consists of a bank of P linear 2-dimensional (2D) FIR filters with frequency responses specified by the frequency modulation of NARW contact calls [11]. The frequency responses of the FIR filters maximizing the statistics $z(t)$ are shown in Fig 1 (top frame) by the red

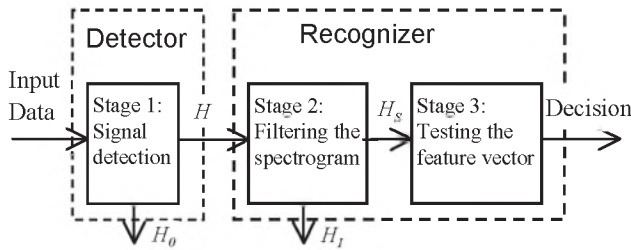


Fig. 2. A Block diagram of the proposed technique.

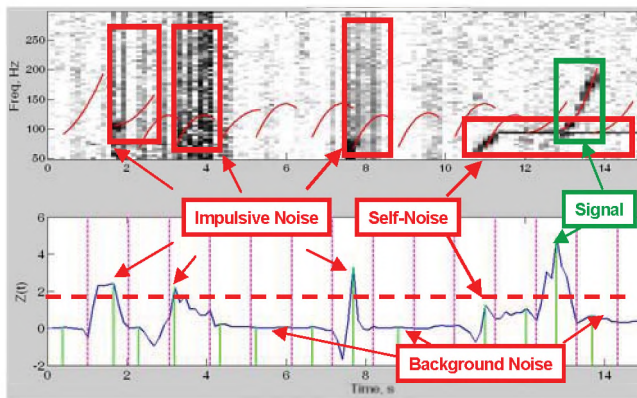


Fig. 1. A spectrogram of the input data (top frame) and the values of the statistic calculated by the spectrogram-based detector (bottom frame).

lines. The values $z(t)$ calculated from the detector output are represented in Fig. 1 (bottom frame) by green lines.

The threshold is determined by applying an optimality criterion, which is introduced based on the management goals of the detector as well as on *a priori* information. We use the Neyman-Pearson criterion, which makes it possible to minimize a false alarm probability for a given probability of detection. In practice, the threshold is set up to automatically detect 80% or more of the NARW contact calls visually detected by the human operators. The problem of choosing the threshold is beyond the scope of this paper. Instead, we focus our attention on the problem of optimizing the structure of the recognizer used in the following stages.

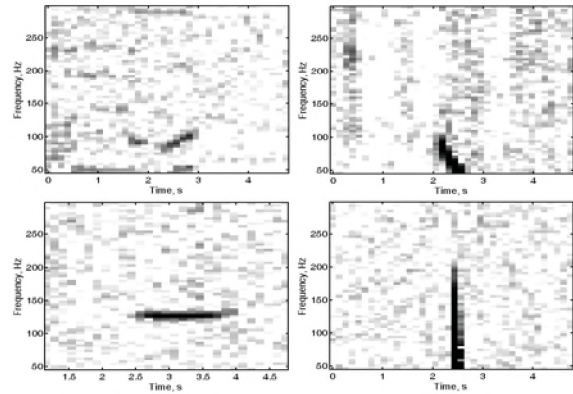


Fig. 3. The spectrograms of the typical impulsive processes: upsweep (top left), downsweep (top right), harmonic (bottom left) and wideband transient impulse (bottom right).

The data segments for which the hypothesis H_0 is accepted do not require any actions. If for a given $x(t)$ the hypothesis H is accepted, this segment is recognized in the next stages. Only these data segments are considered hereafter (For the sake of simplicity, the time index associated with those segments is omitted). Due to the presence of impulsive noise, many segments detected in the first stage may contain no signals except noise transients. As a result, the following hypothesis can be introduced:

$$H_S : X = S + W, H_I : X = Q + W \quad (1)$$

where X , S , Q and W are the matrixes representing the spectrograms of the data segment $x(t)$, signal, impulsive noise, and background noise, correspondingly.

The problem can be formulated as follows: using X , accept or reject the hypothesis H_S . We propose a solution based on a two-stage recognition technique. In the first stage, the hypothesis H_I is divided into the M sub-hypothesis $H_{I_m}, m = 1, \dots, M$. For each H_{I_m} , a parametric model of noise is used. The models are based on the spectral properties of typical kinds of impulsive noise observed in the empirical data. Based on that model, a spectrogram-based algorithm that tests the hypothesis H_S against H_{I_m} is designed. If the hypothesis H_S is accepted, the corresponding data segment is tested in the final stage. In this stage, a feature vector testing algorithm is applied. A block diagram of the proposed technique is shown in Fig. 2. The signal recognition algorithms are designed in Section 3.

3. SIGNAL RECOGNITION

Since typical noise conditions can differ for different locations, we restrict our investigations to the data

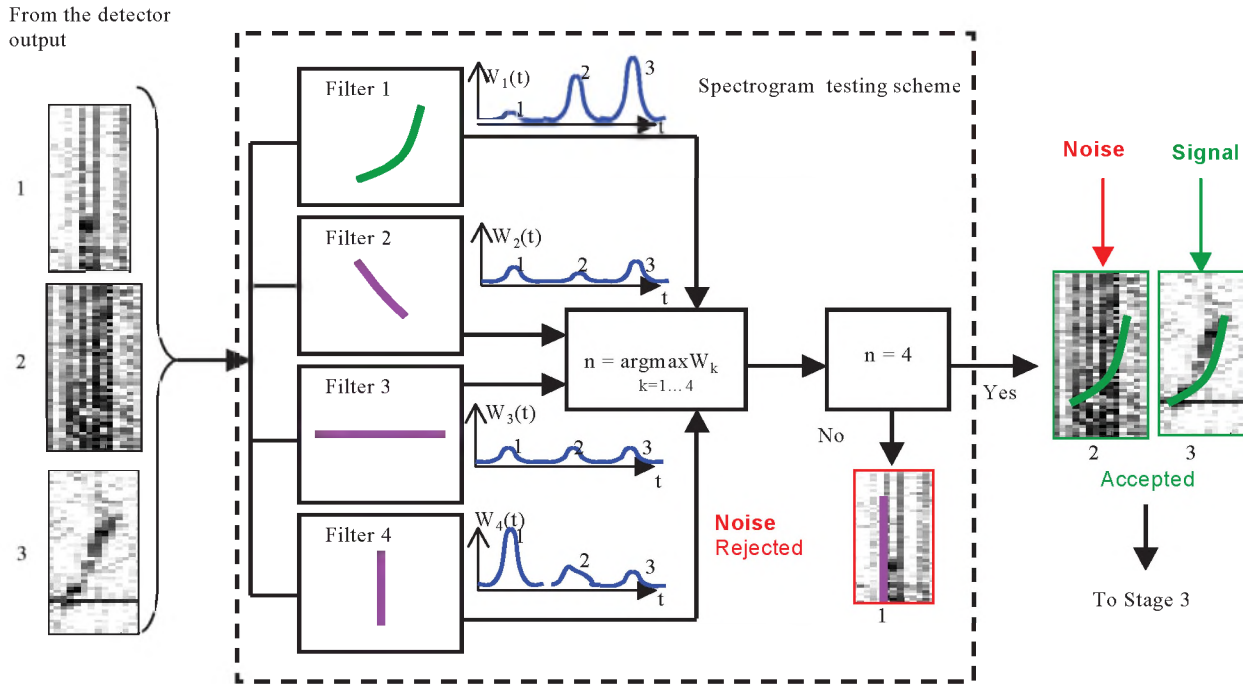


Fig. 4. Spectrogram testing scheme.

recordings collected at Cape Cod Bay, a primary habitat area for NARW. Data analysis shows that many noise transients have spectrogram-based images that are similar to the locally narrowband down-sweep and harmonic impulses as well as to wideband transients; see Fig. 3. Based on these observations, we introduce the following classes of impulsive processes: G_1 - up-sweep transient, G_2 - downsweep transient, G_3 - constant frequency tone transient and G_4 - wideband transient. The class G_1 represents the signals and the classes $G_2 - G_4$ represent impulsive noise.

To design a spectrogram-testing algorithm, we use a strategy similar to the generalized likelihood ratio test. For each class G_m , the statistic $w(m, X)$, $m = 1, \dots, 4$ is computed. If $w(1, X) < w(k, X)$, $k = 2, 3, 4$, then the hypothesis H_1 is accepted and the testing procedure is terminated for a given data segment. Otherwise the data segment is tested in the final stage.

The spectrogram testing scheme is shown in Fig. 3. It is similar to the spectrogram-based detector in the sense that it consists of a bank of 4 linear 2D FIR filters. The impulsive responses of the filters are specified by the phase structure of the typical impulsive processes, Fig. 3.

A FV is tested in the final stage. Features being used should contribute most to discrimination between signals and noise and should be easy to extract. Because of the lack of *a priori*

information regarding variability of NARW contact calls, feature selection is a difficult problem. In this paper, we use features similar to those used by human operators when visually analyzing the spectrogram. Let the symbol $\mathbf{v} = (v_1, \dots, v_K)^T$ denote the K -dimensional feature vector. We introduce the feature space with the dimension of $K = 11$ and with the features represented in Table I. The spectrogram of a NARW contact call and some features extracted from the spectrogram are also displayed in Fig. 5.

To design the FV recognition algorithm, the following approach has been used. For most of the NARW contact

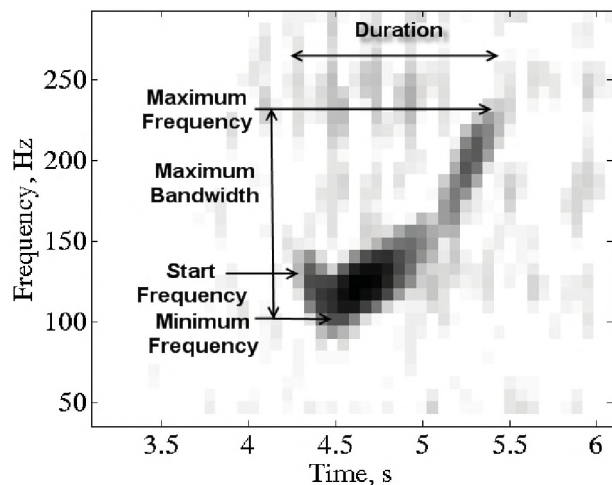


Fig. 5. Features extracted from the NARW contact call.

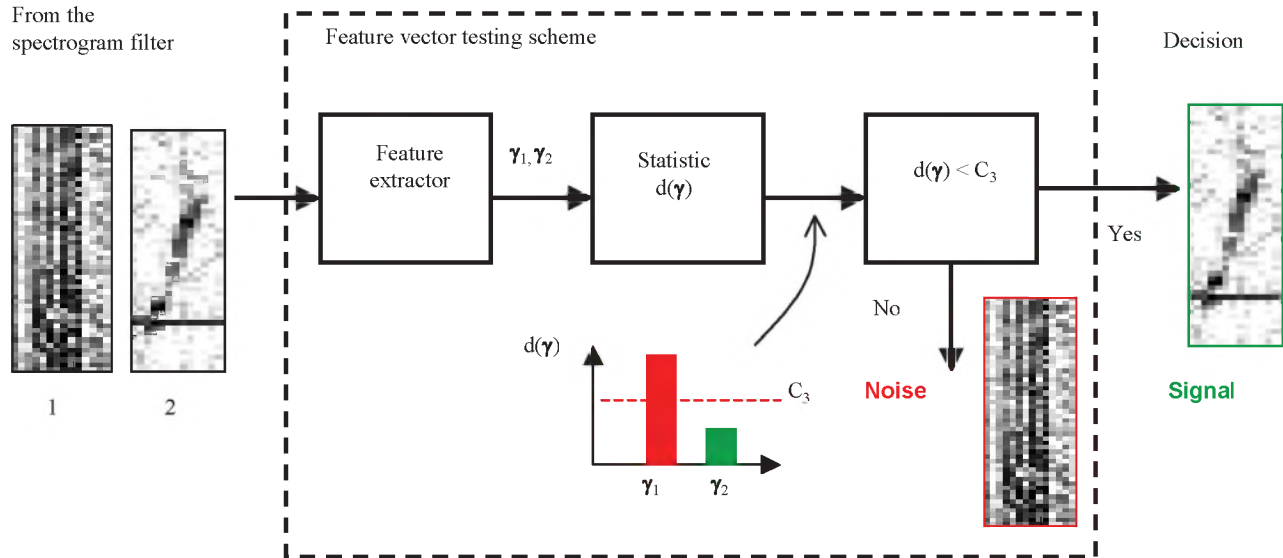


Fig. 6. Feature vector testing scheme.

calls, the FV belongs to some subspace of the K -dimensional feature space. For example, it is known that for the typical NARW contact calls, the minimum frequency is $50 \leq v_4 \leq 150$ Hz and the maximum bandwidth is $20 \leq v_3 \leq 170$ Hz. Based on the empirical observations, similar one-dimensional sets can be introduced for each element of the feature vector so that the signal subspace can be defined as:

$$V_S = \{ \mathbf{v} | v_{i \min} \leq v_i \leq v_{i \max}, i = 1, 2, \dots, K \} \subset E^K \quad (2)$$

where $v_{i \min}, v_{i \max}$ are the scalars defining the bounds of the i th feature. As a measure of discrimination between any particular FV $\mathbf{v} = \mathbf{v}(\mathbf{X})$ and V_S , we introduce the following discriminant function:

$$h(\mathbf{v}) = \sum_{i=1}^K h(v_i) \quad (3)$$

where

$$h(v_i) = \begin{cases} 0 & \text{if } v_{i \min} \leq v_i \leq v_{i \max} \\ A_{i \min} (v_i - v_{i \min})^2 & \text{if } v_i < v_{i \min} \\ A_{i \max} (v_i - v_{i \max})^2 & \text{if } v_i > v_{i \max} \end{cases} \quad (4)$$

$A_{i \min}$ and $A_{i \max}$ are the scalars. The value $h(\mathbf{v})$ is compared with a threshold C_R and the hypothesis H_S is accepted if $h(\mathbf{v}) \leq C_R$. Otherwise the hypothesis H_S is rejected. The FV testing scheme is displayed in Fig. 6. The

values $v_{i \min}, v_{i \max}, A_{i \min}$ and $A_{i \max}$ specifying the FV testing algorithm are shown in Table I.

In practice, the unknown parameters $v_{i \min}, v_{i \max}, A_{i \min}$ and $A_{i \max}$ can be determined using the training data set. It is worth noting that although the proposed FV testing algorithm is heuristic, it uses the statistical properties of signals and noise. As a result, the algorithm can provide high recognition performance.

4. TEST RESULTS

Since the signal recognition technique considered here involves a two-stage decision-making process, it is difficult to estimate the recognition performance in terms of conventional receiver operating characteristics. Therefore, the performance has been evaluated using the empirical probabilities of signal recognition and false alarm.

The empirical probability of recognition has been evaluated using NARW contact calls detected by the human operators from different testing data sets. (The testing data sets were different from the training data set.) The detector and recognizer thresholds were selected as $C_D = 0.35$ and $C_R = 1$, respectively. Under these values of the threshold, the probability of signals recognition is close to 0.8. The actual probabilities of recognition obtained for different data sets are shown in Table II.

The empirical probability of false alarm has been estimated using a data set CCB04 collected in Cape Cod Bay from December 18, 2002, to January 18, 2003. Since the ambient noise conditions may change dramatically with time, the false alarm probability was computed for chunks of data 24

Table I. Parameters of the feature space

Feature	Parameters			
	$v_{i\ min}$	$v_{i\ max}$	$A_{i\ min}$	$A_{i\ max}$
Signal duration, v_1	0.5 s	1.5 s	2.5	1
Mean value of the intermediate bandwidth, v_2	0	15 Hz	0	0.03
Start-end bandwidth, v_3	20 Hz	170 Hz	0.025	0.015
Minimum frequency, v_4	65 Hz	170 Hz	0.055	0.025
Maximum bandwidth, v_5	20 Hz	170 Hz	0.05	0.015
Duration of upsweep part of the signal, v_6	0.3 s	1.5 s	4	2
Segmentation threshold, v_7	4	10	0.3	0
Local noise level, v_8	0	0.02	0	4
Percentage of holes in the object, v_9	0	0	0	3
Percentage of downsweeps in the IF, v_{10}	0	0	0	4
Percentage of harmonicas in the IF, v_{11}	0	0.3	0	3

h in length each. The results of this test are depicted in Fig. 7. The total number of false alarms produced by the proposed technique was 826.

Test results demonstrate that the use of the proposed recognition technique essentially reduces the false alarm probability from the spectrogram-based detector [11]. In particular, the CCB04 data set contains 1,331 NARW contact calls detected by human operators (see Table II). The spectrogram-based detector was able to detect 1,289 signals so that the detection probability on the detector output was $P_D = 0.97$. The total number of false positives provided by the spectrogram-based detector was 113,341. The proposed recognizer was able to recognize 1,092 out of 1,289 signals so that the total decrease in probability of detection was $r_d = 1289/1092 = 1.18$. The corresponding decrease in probability of false alarm was $r_{fa} = 113341/826 = 137.2$.

5. DISCUSSION

For the threshold values applied, the probability of recognition of NARW contact calls ranged from 0.79 to

Table II. Probability of recognition

Data set	Observation time	Number of tested signals	Probability of recognition
CCB00	08/03/2001 – 10/04/2001	14394	0.88
CCB02	28/03/2002 – 31/05/2002	1475	0.81
CCB03	21/11/2002 – 18/12/2002	67	0.79
CCB04	18/12/2002 – 18/01/2003	1331	0.83
CCB05	18/01/2003 – 04/03/2003	1792	0.85
CCB06	04/03/2003 – 21/04/2003	313	0.8
CCB09	28/02/2004 – 17/04/2004	2220	0.86
Total		21592	0.87

0.88 (see Table II). The decrease in the probability of recognition can be explained by the influence of the following factors. First, a certain number of selected calls were hardly visible on the spectrogram and hence had relatively low SNR. Although investigation of the influence of the SNR on the recognition probability was not within the scope of this work, test results demonstrate that the recognition probability decreases as the SNR goes down. Moreover, as the results reported in [5] show, when detecting the signals with low SNR, the human operators may select up to 85 data segments with no signals on them per 24 h of observed data. The operators can also make false selections because of the similarity between contact calls and some kinds of impulsive noise. In our testing data, a certain number of selections made by the operator are questionable and are not approved by other operators. The *a priori* uncertainty regarding a signal parameter is a fundamental problem in passive bioacoustics. The actual range of signal variability is unknown to the observers. As a result, there is a nonzero probability that the selection made by the human operator is actually noise. Hence, the actual probability of recognition can be higher than that represented in Table II.

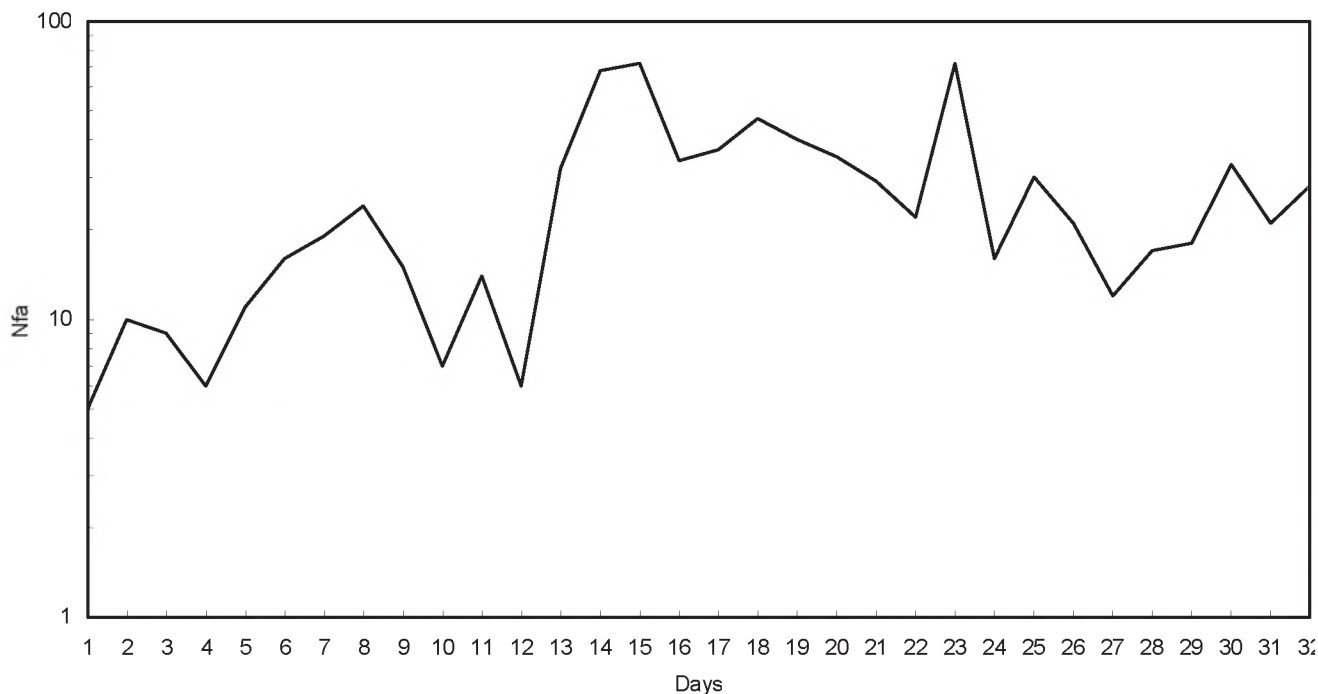


Fig. 7. The number of false alarms provided by the proposed technique. Data collected at Cape Cod Bay from December 18, 2002, to January 18, 2003.

Another factor decreasing the recognition probability is that a certain percentage of the selected NARW contact calls overlapped with transients or other calls. The proposed technique was not designed to operate under such conditions. Therefore, the problem of distinguishing partly overlapped signals should be a topic of future research.

Using the proposed technique for analyzing a long data recording, the human operator only has to inspect the automatically detected clips, thereby reducing the amount of time necessary. As test results show, an average of 26 false alarms per 24 h of observation is generated by the proposed technique. (For the duration of the false data segment equal to 1.024 s, the 26 false alarms per 24 h of observation corresponds to the false alarm probability of 3.08×10^{-4} .) Testing 26 data segments requires about 1 min per operator whereas a complete browsing of one day's worth of data by visual analysis of the spectrogram and by listening to the data requires 2–8 h per operator. Different data sets will likely have higher or lower numbers of false alarms on the recognizer output. However, practical use of the proposed technique in the Bioacoustics Research Program at the Cornell Laboratory of Ornithology shows that, on average, the human effort needed to detect more than 80% of NARW contact calls can be reduced by more than 20 times as compared with the data analysis performed by a human operator alone.

6. CONCLUSION

In the presence of a high impulsive noise rate, the probability of false alarm provided by the spectrogram-based detector can increase dramatically. To decrease false alarm probability without negatively affecting the probability of detection, a new technique proposed in this paper can be used. This technique is based on a multiple-stage decision-making process involving the spectrogram and feature vector testing algorithms.

Test results demonstrate that applying the proposed signal recognition technique to the spectrogram-based detector makes it possible to reduce the false alarm probability by more than 100 times when decreasing the probability of detection by 1.2 times as compared with the spectrogram-based detector. Correspondingly, the hours that humans need to detect 80% and more of NARW contact calls can be reduced by more than 20 times as compared with the data analysis performed by a human operator alone.

7. ACKNOWLEDGEMENTS

The author wishes to thank E. Rowland and M. Fowler for their assistance in marking right whale calls and for useful discussions. Thanks also to P. Leonard for editing the draft version of the manuscript. Research funded by NOAA grant NA03NMF4720493.

8. REFERENCES

- [1] Stellwagen Bank National Marine Sanctuary. Passive acoustic monitoring. Available: http://stellwagen.noaa.gov/science/passive_acoustics.html. Last viewed online 10 October 2007.
- [2] Cornell Lab of Ornithology, Bioacoustics Research Program. Undersea recording: Pop-Ups. Available: <http://www.birds.cornell.edu/brp/hardware/pop-ups>. Last viewed online 10 October 2007.
- [3] C. Clark, T. Calupca, D. Gillespie, K. Von der Heydt and J. Kemp, "A near-real-time acoustic detection and reporting system for endangered species in critical habitats," *J. Acoust. Soc. Am.* vol. 117, p. 2525, 2005.
- [4] D. K. Mellinger, S. L. Nieukirk, H. Matsumoto, S. L. Heimlich, R. P. Dziak, J. Haxel and M. Fowler, "Seasonal occurrence of North Atlantic right whale (*Eubalaena Glacialis*) vocalizations at two sites on the Scotian shelf," *Marine Mammal Science*, vol. 23, pp. 856-867, Oct. 2007.
- [5] I. Urazghildiiev and C. Clark, "Detection performances of experienced human operators compared to a likelihood ratio based detector," *J. Acoust. Soc. Am.*, vol. 122, pp. 200 – 204, 2007.
- [6] D. K. Mellinger and C. W. Clark, "Recognizing transient low-frequency whale sounds by spectrogram correlation," *J. Acoust. Soc. Am.* vol. 107, pp. 3518 – 3529, 2000.
- [7] D. K. Mellinger, "A comparison of methods for detecting right whale calls," *Canadian Acoustics*, vol. 32, pp. 55-65, 2004.
- [8] L. M. Munger, D. K. Mellinger, S. M. Wiggins, S. E. Moore and J. A. Hildebrand, "Performance of spectrogram cross-correlation in detecting right whale calls in long-term recordings from the Bering Sea," *Canadian Acoustics*, vol. 33, pp. 25-34, 2005.
- [9] D. Gillespie, "Detection and classification of right whale calls using an "edge" detector operating on a smoothed spectrogram," *Canadian Acoustics*, vol. 32, pp. 39-47, 2004.
- [10] I. Urazghildiiev and C. Clark, "Acoustic detection of North Atlantic right whale contact calls using the generalized likelihood ratio test," *J. Acoust. Soc. Am.*, vol. 120, pp. 1956 – 1963, 2006.
- [11] I. Urazghildiiev and C. Clark, "Acoustic detection of North Atlantic right whale contact calls using the spectrogram-based statistics," *J. Acoust. Soc. Am.*, vol. 122, pp. 769 – 776, 2007.



Photo Credit: Erin A. Falcone. Copyright: Cascadia Research, Olympia, WA, USA

DETECTION OF LEOPARD SEAL (*HYDRURGA LEPTONYX*) VOCALIZATIONS USING THE ENVELOPE-SPECTROGRAM TECHNIQUE (tEST) IN COMBINATION WITH A HIDDEN MARKOV MODEL

Holger Klinck, Lars Kindermann and Olaf Boebel

Alfred Wegener Institute for Polar and Marine Research, Am Alten Hafen 26, 27568 Bremerhaven, Germany

Homepage: www.awi.de/acoustics

Holger.Klinck@awi.de

ABSTRACT

This paper describes a technique for the automated detection of leopard seal (*Hydrurga leptonyx*) vocalizations. Automatic detection of leopard seal calls within the Antarctic underwater soundscape is difficult because (a) the calls are frequently of low amplitude (b) the call duration is highly variable and (c) the frequency band overlaps with those of many other marine mammal vocalizations. However, humans easily distinguish leopard seal vocalizations from those of other marine mammals because of the calls' distinctive sound, which is a result of the pulsed structure of the leopard seal vocalizations. To exploit the unique temporal evolution of the pulse repetition rate (PRR) in high (HDT) and low (LDT) double trills, the Envelope-Spectrogram Technique (tEST) was developed. The extracted PRR feature allows detection of the target vocalizations even against a background of other marine mammal vocalizations. To handle the high variability of the calls' duration, the tEST algorithm was combined with a Hidden Markov Model (HMM) which is particularly well adapted to handle temporal variability. The developed HMM based detection algorithm worked rather reliably. The detection rate over a 4 day test period was high (72 %) although the signal to noise ratio (SNR) was low (< 10 dB). The number of false positive detections (12 %) was tolerable. Most of the false positives occurred during the period when R/V Polarstern was approaching the recording station and the SNR was temporarily < 0 dB. The detector worked 3 times faster than real-time and is therefore suitable for both off line biological research and time critical in-the-field applications, such as the detection of the presence of leopard seals in the context of human diver operations.

SOMMAIRE

Cet article décrit une technique de détection automatique des vocalisations du Léopard de Mer (*Hydrurga leptonyx*). La détection des sons émis par le Léopard de Mer à travers le bruit de fond sous marin est difficile parce que (a) les émissions sont fréquemment de basse amplitude (b) la durée des émissions est hautement variable et (c) les vocalisations sont dans la même bande que celle utilisée par de nombreux autres mammifères marins. Cependant, l'homme est facilement en mesure d'identifier les vocalisations émises par le Léopard de Mer de celles des autres mammifères, grâce à la pulsation particulière de ces émissions. Pour exploiter cette caractéristique unique de l'évolution temporelle du taux de répétition des pulsations (PRR) des doubles trilles hauts (HDT) et graves (LDT), la technique du spectrogramme de l'enveloppe (tEST) a été développée. Les caractéristiques PRR du signal permettent la détection des vocalisations recherchées même en présence de celles d'autres mammifères marins. Pour résoudre les problèmes dus à la haute variabilité des durées d'émission, l'algorithme tEST a été combiné avec le modèle des chaînes de Markov (HMM), particulièrement bien adapté pour appréhender les variations temporelles. Cet algorithme de détection basé sur les HMM s'est révélé relativement performant. Le taux de détection sur une période d'essai de quatre jours a été élevé (72 %) malgré un faible rapport signal sur bruit (SNR) (< 10 dB). Le nombre de détections positives erronées (12 %) était tolérable. La plupart des détections erronées se sont produites lorsque le navire de recherche R/V Polarstern s'est approché de la station d'enregistrement, diminuant ainsi le SNR (< 0 dB). Le détecteur travaillant trois fois plus vite que le temps réel, il est de fait utilisable aussi bien pour les analyses de données post récolte, que pour une utilisation directe sur le terrain, comme par exemple la détection de la présence de Léopards de Mer lors d'opérations de plongée sous-marine.

1. INTRODUCTION

The leopard seal (*Hydrurga leptonyx*) represents one of three Antarctic pack ice seal species. Leopard seals are solitary living animals, feeding on krill, squid, fish, penguins and other seal species (Reeves et al., 2002). As a top predator of circumpolar distribution, the leopard seal plays an important role in the Antarctic ecosystem. The population size is estimated at around 200000 animals (Reeves et al., 2002). Research on this species is restricted due to the Antarctic pack ice region being accessible for humans usually only during the short austral summer period.



Figure 1: Leopard seal (*Hydrurga leptonyx*) on floating sea ice.

Underwater, leopard seals are known to be rather vocal - at least during polar summer when most of the research was conducted. Stirling and Siniff (1979) described high vocalization rates of male leopard seals during the breeding season (November to January). Females show above-average vocalization rates during sexual receptivity (Rogers et al., 1996). Hence, passive acoustic monitoring offers the unique possibility to investigate the species without a need of direct access. PALAOA - the Perennial Acoustic Observatory in the Antarctic Ocean, (Boebel et al., 2006) is an autonomous recording station operated by the Alfred Wegener Institute (AWI), Germany, on the Ekström Ice Shelf close to the German Neumayer Base, providing underwater recordings from the Atlantic sector of the Southern Ocean. Since January 2006, PALAOA records the Antarctic underwater soundscape quasi-continuously. The station's audio system allows broadband data acquisition with sampling rates of up to 192 kHz and 24 bit resolution. So far more than 6400 hours of acoustic data (as at September 2007) were accumulated. The recorded sounds are transmitted in real-time to the AWI in Germany, allowing real-time access and analysis of the acoustic data.

Extracting the signals of interest - in this case the leopard seal vocalizations - from the resulting 2.5 TBytes of data, is challenging. Obviously, human "observers" will not be able to manage this task, but rather, numerical detection algorithms need to be developed to perform an automated, computer based search. The resulting time series of calls will then form the data base for ecological studies with focus on diel patterns, diurnal and seasonal variability and their interrelation with the changing physical environment.

Apart from these scientific applications, the development of detection algorithms for leopard seal vocalizations can also help to increase the safeness of research divers in the Southern Ocean. Several encounters between human divers and leopard seals have been reported throughout the last decades (Muir et al., 2006). The most serious incident occurred in July 2003 at the British Rothera Station, located at the Antarctic Peninsula, when a scientist was killed by a leopard seal. As a consequence of this accident, acoustic monitoring prior to and during diving activities are used by AWI as risk mitigation method for diving activities. To this end, robust and fast (at least real-time) detection algorithms are needed to screen the hydro-acoustic recordings.

2. THE ACOUSTIC ENVIRONMENT

The Southern Ocean is among the regions least disturbed by anthropogenic noise. However, PALAOA records reveal a high degree of abiotic and biotic acoustic activity in the Southern Ocean. During austral summer in particular, the Antarctic underwater soundscape is dominated by the vocalizations of Weddell seals (*Leptonychotes weddellii*), Ross seals (*Ommatophoca rossii*), crabeater seals (*Lobodon carcinophaga*), leopard seals (*Hydrurga leptonyx*) and various baleen (Mysticeti) and toothed (Odontoceti) whale species.

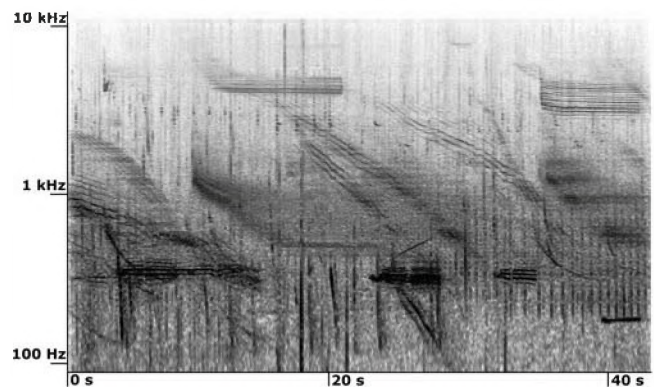


Figure 2: Spectrogram of a PALAOA sound file.

Figure 2 shows a spectrogram of a typical sound file as recorded at the PALAOA Station in austral summer. Overlapping vocalizations from different animals/species significantly complicate the detection of specific

vocalization “targets.” However, human listeners can easily distinguish leopard seal vocalizations from those of other marine mammals because of their distinctive sound. It is believed that this distinctive sound is a result of the pulsed structure of the leopard seal vocalizations. To develop a detection algorithm for leopard seal calls, this publication exploits in detail the temporal structure of the pulse repetition rate (PRR) throughout the calls. The PRR feature is exclusively linked to leopard seal vocalizations (at least in the vicinity of the PALAOA Station) and seems to be a robust feature for the detection while other marine mammal vocalizations are present.

3. LEOPARD SEAL VOCALIZATIONS

3.1 State of knowledge

A first (partial) spectrogram of a leopard seal vocalization was published by Ray (1970) while Stirling and Siniff (1979) described four different leopard seal call types quantitatively. A comprehensive description of the vocal repertoire of leopard seals was published by Rogers et al. (1995). Rogers identified twelve different call types by analyzing recordings of captive and free living animals (Prydz Bay, Antarctica). The frequency span of the analyzed call types ranges between 65 Hz and 4800 Hz. Thomas et al. (1983) recorded ultrasonic vocalizations with frequencies up to 164 kHz of leopard seals in captivity during hunting activity. However, ultrasonic vocalizations have so far not been reported from field studies.

By far the most frequent vocalizations of leopard seals are the so called high double trill (HDT – see Figure 3) and the low double trill (LDT). In the PALAOA recordings, the HDT (frequency range: 2.5 - 4.5 kHz) and the LDT call type (frequency range: 230 - 470 Hz) represent more than 70 % of all leopard seal vocalizations while Rogers et al. (1995) reported 79 % of such calls for their data set.

Due to its distinct Signal to Noise ratio (SNR) this paper focuses on the analysis and detection of HDTs.

3.2 The high double trill (HDT)

Figure 3 (top) depicts the waveform and spectrogram of a high double trill. The waveform clearly shows that the call is separated into two segments which consist of a series of short pulses. These pulses cause an amplitude modulation of the main signal. This modulation generates so called sidebands, which are revealed in the spectrogram in Figure 3 (bottom). The frequency difference between the sidebands equals the frequency of the PRR. This implies that an increasing (decreasing) PRR is causing increasing (decreasing) frequency differences between the sidebands. In general the number of sidebands is determined by the type of amplitude modulation. For sinusoidal modulation, only two sidebands are generated while the primary

frequency is rendered invisible in the spectrogram. By contrast, triangular or rectangular modulations cause multiple (> 2) sidebands. The type of amplitude modulation of the HDT is in-between a sinusoidal and triangular modulation (see Figure 5).

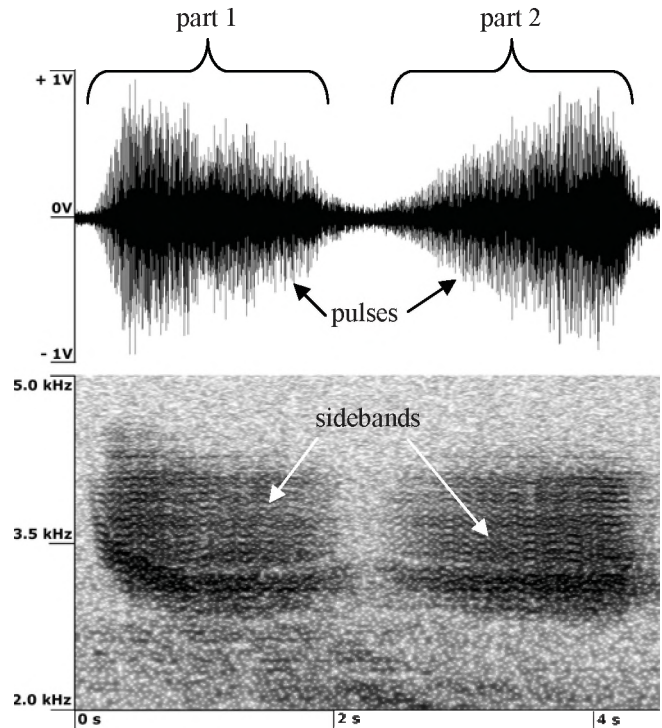


Figure 3: Waveform and spectrogram of a HDT call.

A total of 150 HDTs were analyzed to gain information about the temporal structure and the frequency characteristics of this call type (Table 1). The HDTs feature a high variability of the call duration, ranging from 1.9 s - 9.0 s. The calls cover a frequency range between 2500 Hz and 4450 Hz.

Table 1: Acoustic features of the HDTs recorded at PALAOA.

	Min	Max	Mean	Stdv
Call duration	1.9 s	9.0 s	4.5 s	1.6 s
Frequency	2500 Hz	4450 Hz	---	---

4. THE ENVELOPE-SPECTROGRAM TECHNIQUE (tEST)

So far, HDT and LDT descriptions regarded the PRR as constant for the duration of the call (Rogers, 2007; Rogers et al., 1995). By contrast, spectrograms of calls recorded by PALAOA reveal varying side-band distances over the duration of HDT and LDT calls, suggesting a variation of the PRR in the course of the call.

To accurately analyse the temporal structure of the PRR, a Matlab based algorithm was developed. The respective sound snippet was first band pass filtered with the frequency range of the target signal (HDT: 2.5 - 3.5 kHz). The envelope of the absolute values of a band passed waveform was calculated by detecting all maxima values (peaks) in the waveform and interpolating (1-D) the detected points. The resulting waveform was then down-sampled to a sampling rate of 1000 Hz and transformed into the frequency domain by means of a Fast Fourier Transformation (FFT-Parameters: Hamming window 256 points; 50 % overlap).

This algorithm, named tEST (the Envelope-Spectrogram Technique) hereinafter, provides the spectrogram of the envelope, i.e. the frequency-evolution of the PRR. If the SNR over the frequency range of the target signal is low (<6 dB), it can be helpful to use a narrower filter which covers only the frequency range of the signal of the highest energy.

4.1 Applying tEST on HDT calls

Figure 4 shows the result of tEST applied on a HDT. The signal was processed as described in the former paragraph.

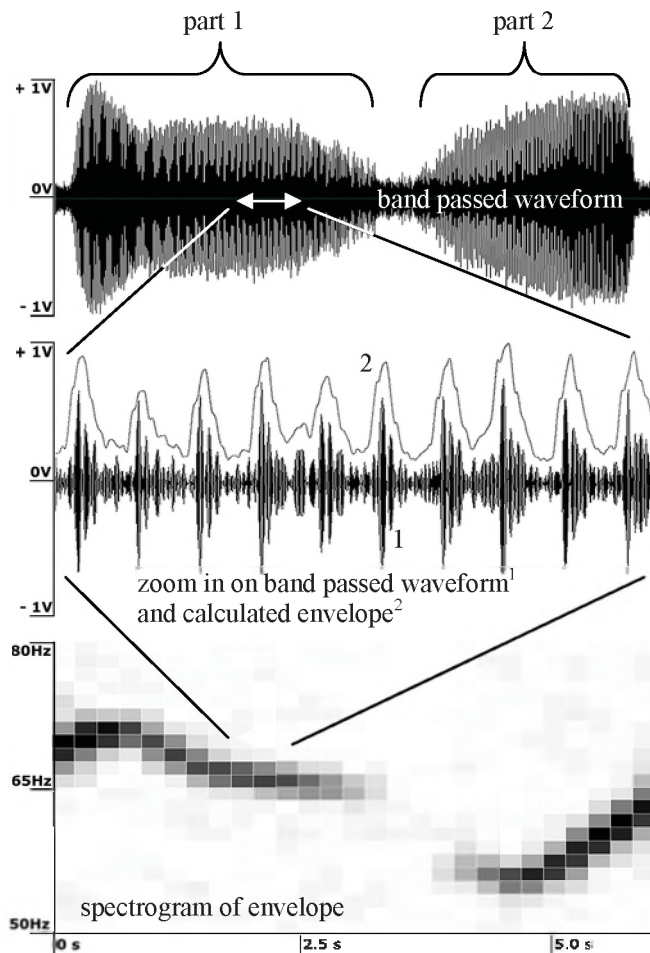


Figure 4: Results of tEST applied on a HDT.

The result of the FFT of the envelope signal is displayed in the lower part of Figure 4. For the selected sample, the pulse repetition rate of the envelope varies between 52 Hz and 72 Hz (20 Hz bandwidth @ 2 Hz resolution). The first part of the call is characterized by descending rates. In the second part the pulse repetition rate is ascending.

The same 150 HDTs as used for the spectral description were analysed with tEST. All vocalizations showed descending repetition rates in the first part of the call and ascending rates in the second part. The observed frequencies ranged between 45 Hz and 75 Hz.

5. DETECTION OF LEOPARD SEAL VOCALIZATIONS

5.1 Introduction

In summary, the previous sections showed:

(a) The analysis of the acoustic environment (Section 2) confirmed that vocalization of various whale and seal species occur simultaneously within the frequency bands of the target vocalizations. Thus the likelihood of false positive detections will be high using detection methods such as energy summation or comparing energies in different frequency bands.

(b) The call durations of the HDT calls vary widely (Section 3) which renders detection algorithms based on matched filter/spectrogram correlation difficult. Further more the detection performance of these methods is linked to the representativeness of available examples of the target vocalization. 150 calls are probably not enough samples to create an effective filter.

(c) Leopard seal calls exhibit temporal modulation of the PRR throughout HDTs providing a unique feature of this leopard seal call type (Section 4). Other marine mammal vocalizations in frequency bands overlapping with those of the target vocalizations are likely not to pass as “false positives” if the detection algorithm is to exploit this rather unique feature.

For the detection of the HDTs based on the PRR feature a Hidden Markov Model (HMM) was applied. The next paragraphs will give a short introduction to HMMs and how they are used for the detection.

5.2 Hidden Markov Models (HMM)

Hidden Markov Models (HMM) are statistical models for the detection and classification of transient patterns, representing state of the art tools in human speech

recognition (Rabiner, 1993). HMMs are particularly well adapted to this call type of variable duration, as they allow detection of temporally changing structures. For leopard seal vocalizations this implies that the detection probability is high irrespectively of the calls duration, as long as the envelope follows the specific temporal evolution (see Figure 6). A short introduction and description on how to build a Hidden Markov Model are given in the following paragraphs. Unfortunately, the scope of this paper only allows an abbreviated, qualitative description of Hidden Markov Models. For detailed information see Rabiner, 1993 and Deller et al. 2000. All model parameters used in this study are available on request.

(a) Feature extraction: To extract the call type's typical features, it is recommended to choose the best available samples. Consequently the best 100 samples (high SNR) were selected out of the available 150 samples for this process. The features are extracted by means of a time-frame based analysis (frame length: 256 ms) of the envelope signal of the 100 sample files of variable duration (2 - 9 s, depending on call duration). For each time frame, a feature vector is calculated (see Figure 5), representing the respective energy distribution as a function of frequency.

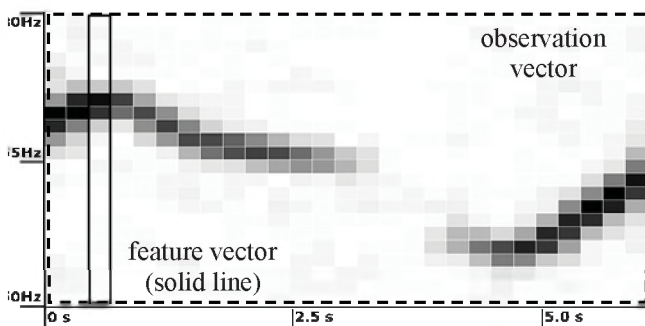


Figure 5: Feature vector (enclosed by solid box) and observation vector (enclosed by dashed box) of a HDT.

All feature vectors ($n = \text{duration of sample file} / 256 \text{ ms}$) of one sample file comprise the so called observation vector (which actually is a matrix - see Figure 7). Hence, 100 sets of spectral vectors from the spectrograms of the envelopes are extracted.

(b) Information reduction: To condense the numerous ensuing feature vectors to a set of 'most significant' feature vectors, a "k-mean (squared Euclidean distance)" cluster algorithm (Deller et al., 2000) was applied to the training set. This creates the so called codebook of 10 (number empirically chosen) codebook vectors representing the target vocalization's most significant (sub-)set of 10 feature vectors. For each feature vector of an observation vector the best matching (minimal distant) codebook vector is determined, which results in an observation sequence. Each of these consists of a series of integers, representing the

sequence of IDs of the best fitting codebook vectors. Thus, each set of the 100 spectral vectors is quantized to a one dimensional array of integers. The resulting set of 100 quantized vectors represents the quantized training set.

(c) Generate the HMM: Evaluating model parameters describing the quantized training set best. A Hidden Markov Model is a quintuple, comprising (a) the number of (hidden) states S ; (b) the state transition matrix A (transition probabilities between the states); (c) the observation probability matrix B ; (d) the state probability vector at time $t=1$, $\pi(1)$; and (e) the number of observable outputs Y (number of codebook vectors), or in short:

$$\text{HMM} = \{S, \pi(1), A, B, Y\}.$$

The number of states (S) was assigned to 5. The number of 10 observable outputs (Y) is given by the size of the codebook. The state transition matrix (A), the observation probability matrix (B) and the state probability vector at time=1 ($\pi(1)$) were determined by applying the forward/backward algorithm (Deller et al., 2000) on the quantized data set.

In the first step, the model parameters A , B and $\pi(1)$ are initialized (see below) and the algorithm calculates the match between the model and the training set. In the second step the algorithm starts to modify the model parameters. The algorithm guarantees that every iteration has a matching likelihood that is \geq to the previous one. Once the matching likelihood converges, training is done.

Critical to this process is the initial guess of the model parameters. In the case of the described target signals, meaningful parameters were unknown. If the initial guess is too far away from the optimal parameters, then the algorithm will only find a local maximum but not the global one. For this reason the initial parameters were randomized and the resulting HMMs used to analyse one sample file including a known number of target signals repeatedly. The best fitting model parameters (giving the highest detection probability for the target signals) were then chosen for the further process.

5.3 Detection of HDTs using the HMM

To detect HDTs with the optimized HMM (5 states), a 6-s window is continuously slid in steps of 1.0 s (~83 % overlap) over the data stream: the respective window content is first band pass filtered (2500 Hz - 3500 Hz). The resulting waveform is then used to calculate the envelope, which is used to derive the observation sequence as described in section 5.2. In a final step the probability of the observation sequence of each window under the assumption of the model $P_{(\text{window}|\text{model})}$ is calculated. In Figure 6 an

example for the output of the HMM detection algorithm is given. The spectrogram shows four signals - two HDTs (c) of different duration, an artificial signal (b) and a Weddell seal call (a). The detector output is shown in the lower part of Figure 6. The detection threshold is set automatically by the algorithm depending on the overall SNR. If the detector output reaches the detection threshold, a call is “detected”. In this sample the HDT calls are clearly detected by the system but not the Weddell seal call or the artificial signal present in the same frequency band.

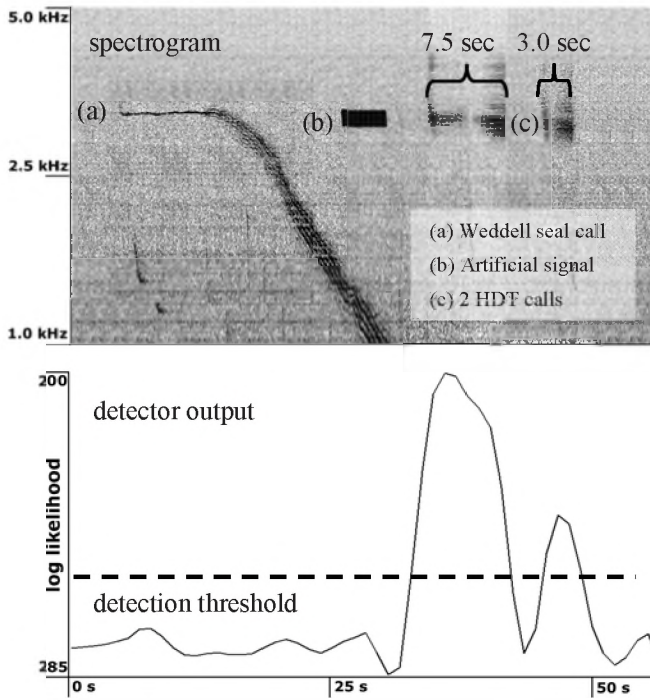


Figure 6: Detector output for a Weddell seal call, an artificial signal and two HDT calls of different duration.

5.4 Results of the detection system running over a test data set

To test the detection algorithm, independent data (i.e. not including the 100 calls used to develop the HMM) from a 4 day period of variable SNR was selected and analysed. The period started out with a good bandwidth related SNR of 10 dB (between 2.5 kHz and 3.5 kHz), which deteriorated to SNR < 0 dB during the last day of the period when R/V Polarstern approached the recording station. To deal with the low SNR the spectrogram of the envelope was manipulated using a wavelet based denoising technique (Kovesi, 2000 and Kovesi, 1999). Also an anisotropic diffusion was performed on the spectrogram to enhance the contrast at sharp intensity gradients (Kovesi, 2000). The use of the denoising and the anisotropic diffusion algorithm increased significantly the detection performance. Thus, the algorithms were directly integrated into the system.

A subset of data was selected (2 min of data every 10 min) to create a reference data set, which was used to evaluate the detection algorithm. The result of the test run is presented in Figure 7. The light line represents the manually detected calls; the dark line represents the automatically detected calls. The total number of manually detected calls in the selected files was 1527 and the detection rate of the system 72 %. The overall temporal evolution of the two curves shows a high degree of similarity. Analysing all files over the period the HMM based algorithm detected 7548 HDTs.

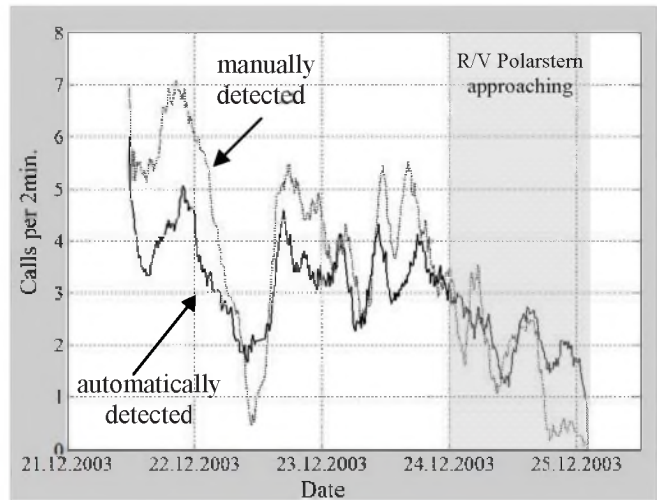


Figure 7: Results of a test run over a four day period.

Most of the false positives (in total 12 %) occurred during the period when R/V Polarstern was approaching - especially when the vessel was close to the recording station (second half of day 24.12.2003). A detailed error analysis will be included in upcoming studies, to determine the exact cause of these false detections.

The detection algorithm is rather fast. Analysing a sound file of 2 minutes duration (48 kHz - 16 bit) takes about 40 seconds (Desktop PC with single Intel Pentium IV 3.4 kHz processor and 2GB RAM). Thus the HMM based detection system is suitable for real-time applications.

6. DISCUSSION AND OUTLOOK

The tEST algorithm which was developed in the course of this study is a useful tool for analysing the temporal evolution of pulse repetition rates in animal calls. The analysis of the high double trill (HDT) of the leopard seal revealed for the first time a temporal variation in the repetition rate of the pulses.

First results of the analysis on low double trills (LDT) of the leopard seal showed also a temporal variation in the pulse repetition rate throughout this call type. The frequency range of the PRR is compared to the HDT around 4 times

lower (bandwidth between 13 Hz and 20 Hz). Future work will concentrate on the detection of LDTs using the PRR feature in combination with a Hidden Markov Model.

However, the specific PRR modulation is exclusively linked to leopard seal vocalizations (at least in the vicinity of PALAOA Station), facilitating the development of a HMM based detection system to detect the target vocalizations in a data set which was entirely overlaid by vocalizations of other marine mammals.

A problem of detection algorithms is often their validation - especially when working with huge data sets. HMM based detection systems provide the "matching probability" between the signal and the used model for each call detected. Analysing this probability over time can help to identify regions where the "matching probability" is low and a validation is necessary in particular.

In summary, it is noted that the Hidden Markov Model worked rather reliably. The detection rate over the 4 day test period was high (72 %) although the SNR was unfavourable (< 10 dB). The number of false positive detections (12 %) was tolerable, because most of the false positives occurred during the period when R/V Polarstern was approaching the recording station when the SNR was temporarily < 0 dB. The detector worked 3 times faster than real-time and is therefore suitable for time critical applications.

ACKNOWLEDGEMENTS

Many thanks to Cornelia Kreiß for pooling the data manually and creating the necessary reference data set for deriving the efficiency of the detection system. The authors especially want to thank Marie A. Roch for her helpful remarks and suggestions. Internal and external reviewers provided useful comments on previous drafts of this manuscript. Delphine Dissard, Catherine Audebert and an anonymous reviewer provided the French translation of the abstract. Setting up the PALAOA observatory would not have been possible without the extensive support of the AWI logistic department.

REFERENCES

- Boebel, O., Kindermann, L., Klinck, H., Bornemann, H., Plötz, J., Steinhage, D., Riedel, S. and Burkhardt, E. (2006): Acoustic Observatory Provides Real-Time Underwater Sounds from the Antarctic Ocean. In: EOS, 87, pp. 361 and 366.
- Deller, J. R., Hansen, F. H. L. and Proakis, J. G. (2000): Discrete-Time Processing of Speech signals. In: Wiley-IEEE Press, Chapter 12, pp. 677-744.

Kovesi, M. (2000): MATLAB and Octave Functions for Computer Vision and Image Processing. School of Computer Science & Software Engineering, University of Western Australia. Available from: <<http://www.csse.uwa.edu.au/~pk/research/matlabfns/>>.

Kovesi, M. (1999): Phase Preserving Denoising of Images. In: The Australian Pattern Recognition Society Conference (DICTA '99), December 1999, Perth, pp. 212-217.

Muir, S. F., Barnes, D. K.A. and Reid, K. (2006): Interactions between humans and leopard seals. In: Antarctic Science, 18(1), pp. 61-74.

Rabiner, L. and Juang, B. H. (1993): Fundamentals of Speech Recognition, Prentice Hall PTR, New York, 496 pp.

Ray, G. C. (1970): Population ecology of Antarctic seals, Volume 1. In: Antarctic Ecology (Ed. Holdgate, M. W.), Academic Press, New York, pp. 398-414.

Reeves, R. R., Stewart, B. S., Clapham, P. J. and Powell, J. A. (2002): National Audubon Society: Guide to Marine Mammals of the World. Alfred A. Knopf, New York, 531 pp.

Rogers, T. L. (2007): Age-related differences in the acoustic characteristics of male leopard seals, *Hydrurga leptonyx*. In: Journal of the Acoustic Society of America, 122(1), pp. 596-605.

Rogers, T. L., Cato, D. H. and Bryden, M. M. (1996): Behavioural significance of underwater vocalizations of captive leopard seals, *Hydrurga leptonyx*. In: Marine Mammal Science, 12(3), pp. 414-427.

Rogers, T. L., Cato, D. H. and Bryden, M. M. (1995): Underwater vocal repertoire of leopard seals (*Hydrurga leptonyx*) in Prydz Bay, Antarctica. In: Sensory Systems of Aquatic Mammals (Ed. Kastelein, R. A., Thomas, J. A. and Nachtigall, K.), De Spil Publishers, Woerden, pp. 223-236.

Stirling, I. and Siniff, D. B. (1979): Underwater vocalizations of leopard seals (*Hydrurga leptonyx*) and crabeater seals (*Lobodon carcinophagus*) near South Shetland Islands, Antarctica. In: Canadian Journal of Zoology, 57, pp. 1244-1248.

Thomas, J. A., Fisher, S. R. and Evans, W. E. (1983): Ultrasonic vocalizations of leopard seals (*Hydrurga leptonyx*). In: Antarctic Journal of the US, 17, page 186.

JOINT LOCALIZATION AND SEPARATION OF SPERM WHALE CLICKS

Paul M. Baggenstoss

Naval Undersea Warfare Center, Newport RI, 02841, p.m. baggenstoss@ieee.org

ABSTRACT

In this paper we consider the joint problems of separating and localizing sperm whale click trains. Click train separation is the single-sensor problem of grouping the clicks from each animal together when the clicks of more than one animal are present at a given sensor. Localization is the problem of localizing the animals based on the measurement of time delays of the same click events at multiple sensors. The two problems are inherently connected. We first consider the two problems independently using novel applications of statistical signal processing methods. For separation, we employ an algorithm inspired by the Viterbi algorithm from dynamic programming. For localization, we employ an algorithm inspired by the expectation-maximization (EM) algorithm. Finally, we use the two algorithms to “assist” each other in a joint localization/separation solution. We demonstrate the algorithm on real data.

SOMMAIRE

En cet article nous considérons les problèmes communs de séparer et de localiser des trains de clic de cachalot. La separation de train de clic est le problème de simple-hydrophone de grouper les clics de chaque animal ensemble quand les clics de plus d'un animal sont présents à une hydrophone donnée. La localisation est le problème de localiser les animaux basés sur la mesure du temps retarde des mêmes événements de clic aux sondes multiples. Le problème deux sont en soi reliés. Nous considérons d'abord les deux problèmes employant indépendamment des applications de nouveaux des méthodes statistiques de traitement des signaux. Pour la séparation, nous utilisons un algorithme inspiré par l'algorithme de Viterbi de la programmation dynamique. Pour la localisation, nous utilisons un algorithme inspiré par l'algorithme de E-M. En conclusion, nous employons les deux algorithmes pour nous aider dans une solution du joint localization/séparation. Nous démontrons l'algorithme sur de vraies données.

1 CLICK TRAIN SEPARATION

1.1 Introduction and Problem Definition

When recorded at a single hydrophone, multiple sperm whale vocalizations are difficult to separate into the click-trains of the individual whales. Previous work has utilized various clues including spectral and temporal features and inter-click correlation as well as multi-sensor time-delay [1],[2],[3]. Based on the structure of the problem, involving many interrelated clues spread over time, we believe there is significant room for improvement through the application of dynamic programming. With dynamic programming, a globally best solution can be approximated through efficient time-recursive processing. Let us assume that a series of clicks has been received at a sensor. For the purpose of this discussion we define an arbitrary error metric $E_{i,j}$ as the result of comparing clicks i and j . It is not important for the discussion how the clicks are compared. We can assume that information about the time duration, amplitude, and spectral content plus a measure of inter-click correlation has been used to develop this error metric. The goal is to arrange clicks into groups or *chains*. If the inter-click error is measured only between adjacent clicks within a chain, then the total error is minimized when the clicks are properly grouped. Of course there must be a penalty for creating a new chain. Otherwise, one could

place each click in a separate chain to minimize the errors. The total error for a *complete grouping* is the sum of the inter-click errors in each group plus the penalty value. The solution could be found by brute-force evaluation of all possible groupings, but in general this is computationally prohibitive. To obtain an efficient algorithm, we employ dynamic programming.

1.2 Dynamic Programming Algorithm

Dynamic programming is a means of solving problems, whose complexity would grow at an exponential rate with time if solved brute-force, recursively with linearly increasing complexity. For details, we refer the reader to the classic book by Bellman [4]. Although the problem we try to solve does not meet the requirements to be solved by dynamic programming, we use an algorithm inspired by dynamic programming. The goal of the algorithm is to group together clicks at a single sensor that come from a given whale and a given propagation path. Before we describe the algorithm, let us define the following terms. **Chain**: a set of associated sequential clicks. **Over-Complete Grouping**: a set of chains that make up a complete set including all clicks, but with duplicates. **Complete Grouping**: a set of chains that make up a complete set including all clicks with no duplicates. **Best Complete Grouping**: the complete

grouping with the minimum error of all complete groupings.

The algorithm operates on the clicks received at a single sensor. Assume the algorithm has processed all clicks up to index $n - 1$ and has an *over-complete grouping* list of candidate chains and chain error values. To terminate the algorithm, it would be necessary to search the list for the *best complete grouping*. If the list of chains is not very large, and since the error metrics have already been computed and the totaled up for each chain, the problem is not computationally prohibitive (since we still have additional clicks to process, we are not going to terminate). We now process the remaining clicks.

To add click n , we do the following. First, we assume that detection n may not be a member of any existing chain so we add it as a “chain of one” with error value P and retain all existing chains. Second, we assume detection n may append to an existing chain. We copy all existing chains, thereby doubling the number of chains, and append click n to the end of each copy (subject to limits on click period). Having added click n , we then proceed to click $n + 1$. The list of chains can grow substantially with each added detection, in fact it grows exponentially. Before the number of chains gets too large to manage, it is necessary to pare down the list. The act of paring down the list we call a *collapsing search* because it collapses the list down to the *best complete grouping*. Note that there is no guarantee that the collapsing search performed on clicks up to n won't lose a chain that is part of the *best complete grouping* of clicks up to $n + 1$. To minimize the chance of losing such chains, we do two things. First, the collapsing search is performed at intervals of T click updates. Second, a set of parallel lists are maintained that perform the collapsing search also at intervals of T click updates, but a different time offset (phase). The use of multiple list phases is illustrated in figure 1.

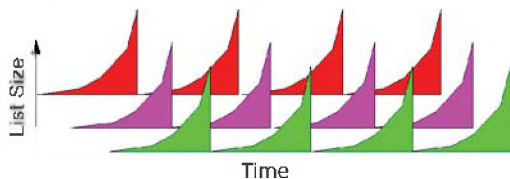


Fig. 1. Illustration of a multiple search phases.

The list size of each list phase grows exponentially until the collapsing search is performed (every T updates).

We performed an experiment to see the reduction in error rate as T increased. We performed a collapsing search at an interval of length T , and kept T phase lists, so one list was performing a search at each update. We used two test problems: (a) a benign problem with a single high-SNR whale with a direct and a reflected propagation path, and (b) a difficult case with up to three whales at low SNR. In each trial, we selected at random 12 consecutive click detections. Then we located the best complete grouping by brute-force search. We then ran the recursive algorithm, with the inter-click error metric described below, with collapsing search every T updates (and kept T phase lists). We considered it an error if the best complete grouping was lost. The error rate was the fraction of the trials in

which the best complete grouping was lost. The results are shown in figure 2. In the benign problem the error rate dropped exponentially as T increased until no errors were found above a value of $T = 2$. In the difficult problem, the error rate decreased less rapidly and never reached the point where no errors were found. This is expected because the globally minimum solution is probably not “truth” in any case. In other words, solutions found using the recursive algorithm did not appear any better to the eye as the global best solution.

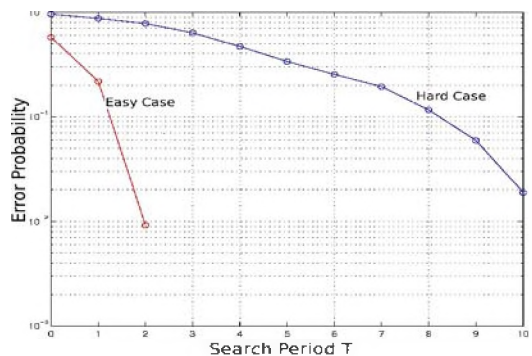


Fig. 2. Error rate as a function of search interval T . The benign case is the lower trace. There were no errors at $T = 3$ and higher.

1.3 Inter-Click Error Metric

We used a probabilistic inter-click error metric. This was accomplished by extracting 10 features from the click pair to be tested. Features included measures of spectral and temporal closeness, correlation measures, etc. We considered the binary decision: associated (H_1) versus not associated (H_0). For training, feature samples of H_1 were obtained from adjacent clicks from hand validated groupings, while feature samples of H_0 were obtained from random click pairs. A Gaussian mixture probability density function (PDF) estimate [5] was used to estimate the distribution under H_0 and H_1 . The probabilistic error metric was obtained from the log likelihood ratio using the log likelihood ratio $E_{i,j} = -\log \{p(x|H_1)/p(x|H_0)\}$. This is the negative of the log-likelihood of the probabilistic test for H_1 vs. H_0 . We used a new chain penalty value of -8 in all our experiments.

2 MULTI-SENSOR LOCALIZATION

We now describe a model-based whale localization algorithm. It is distinguished from existing algorithms in its use of the expectation-maximization (EM) algorithm and use of probabilistic “soft” association of click-pairs.

2.1 Introduction

The problem of localizing sperm whales using click-trains received at multiple sensors works primarily by measuring time delays between the click vocalizations received at multiple sensors and comparing with a propagation model. The task is made difficult by the problem of

associating clicks, that is, knowing if clicks received at separate sensors are actually from the same click vocalization. This is made even more difficult by the existence of multiple propagation paths and the existence of multiple whales. While existing approaches solve the association problem by correlating click trains [6], [7], [8], [9], we seek to employ the E-M algorithm to sort out the association problem. We explore the advantages and disadvantages of the approach and test it using real data.

2.2 Algorithm Description

The EM algorithm [10] has been successfully applied to many problems where there is inherent association ambiguity. In the application of the EM algorithm to such problems, the problem can be regarded as a mixture density and the algorithm takes a particular form [5]. There is a fixed number N of data samples, and a fixed number M of probabilistic “models”.

In our problem, a model is a potential whale location and a data sample is a measured 2-sensor inter-click time delay. Each data sample (inter-click time delay) is regarded as a possible statistical realization of one of the M solutions. Let there be set of NM click association probabilities, denoted by w_{ij} , and representing the probability that data sample i is a realization of model j . The EM algorithm consists of 2 steps:

- 1) The “E”-step: Given the localization solutions, estimate the click pair association probabilities w_{ij} , and solution weights α_j . The solution weights are a measure of validity of each solution.
- 2) The “M” step: Given the click association probabilities, estimate the parameters of each localization solution. Parameters include position and time delay variance.

The “E” and “M” steps are repeated until convergence. Solutions with low α_j are pruned. Notice that rather than associating one click to another, we associate measured click pairs to a source solution hypothesis. We allow all click pairs to exist even if they are false pairings. We add a special “error” localization solution located at the center of the sensor field and with a high location variance (σ_e^2). The algorithm should associate invalid click pairs to the “error” solution. The EM algorithm uses “soft” association - a solution is associated with a click pair with certain probability. It may have nonzero association with all click-pairs. The concept is illustrated in figure 3.

2.3 Algorithm Details

2.3.1 Solution Probabilistic Model

We assume that \square_i is a Gaussian random variable with mean $T(z_j, s_i, r_i)$ and variance σ_j^2 . Let \square_i be the time delay measurement for click pair i . Let $L_{i,j}$ be the likelihood function value for click pair i and solution j

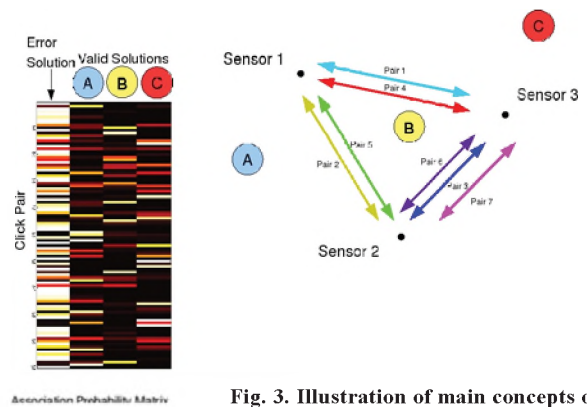


Fig. 3. Illustration of main concepts of EM algorithm.

On the left is the click-pair association probability matrix (CPAPM) represented as an intensity image. The Y-axis is the index of the click pair and the X-axis is index of the solution. On the right is a geographical representation of the sensors and solutions. The three solutions, A, B, and C are represented on the right in geographical position and on the left as columns of the CPAPM. The additional Error solution is shown as the first column. Each horizontal row of the CPAPM is the probability of a given click pair having been generated by a whale at each of the solutions, and sums to 1 over all solutions. Each vertical column can be regarded as a given solution’s probability of ownership for the click pair. Those solutions with the largest column sums are associated with more click pairs, and are therefore more important. The EM algorithm alternatively updates solutions based on the weighted click pairs, and then updates the CPAPM.

defined by,

$$L_{i,j} = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left\{-\left[\tau_i - T(z_j, s_i, r_i)\right]^2 / (2\sigma_j^2)\right\}$$

where σ_j^2 is the time delay error variance for solution j , z_j is the current position vector for solution j , (s_i, r_i) is the sensor pair from which click pair i has been obtained, and $T(z_j, s_i, r_i)$ is the model time delay.

2.3.2 Initialization

We detect individual clicks at each sensor, then create “click pairs” from every two-sensor pair of clicks (that can reasonably be associated with each other given the dimensions of the search area, the maximum range of reception, and the spacing of the hydrophones). Let there be M initial location solutions obtained by a grid-search using any model-based localization procedure. A large number of initial solutions can be used. The solution variances σ_j^2 can be initialized based on the grid-search quantization size. We include an “error” solution, located at the center with wide variance. Let $\alpha_j, 1 \leq j \leq M$ be the solution weights for the M solutions. They can be initialized to $1/M$.

2.3.3 Solution weights and Click Pair Association Probability Matrix (CPAPM) Update (E-step)

The CPAPM is estimated as follows. The un-normalized CPAPM is computed as $\tilde{w}_{i,j} = \alpha_j L_{i,j}$ where i indexes the click pairs and j indexes the solutions. Normalization is then performed so that for each i , $\sum_j w_{i,j} = 1$.

Solution weights, α_j , are obtained by summing the CPAPM along each column (fixed j) to determine the effective number of click pairs associated with solution j , then normalizing so the solution weights sum to 1.

2.3.4 Solution Estimation Procedure (M-step)

The parameters of each solution include the time delay error variance σ_j^2 and the position vector z_j . These parameters are estimated by weighted maximum likelihood (ML) by maximizing

$$Q(z_j, \sigma_j^2) = \sum_{i=1}^N w_{i,j} \log L_{i,j}(z_j, \sigma_j^2)$$

over the parameters z_j, σ_j^2 . Space does not permit a detailed description of the maximization procedure, however, ML is a well-studied method [11].

2.3.5 Solution Error Ellipse

A well-known property of ML is that the statistical random error of the parameter estimates approximates Cramer-Rao lower bound matrix [11]. This is useful for drawing solution error bound ellipses on the geographical solution plot. Let C be the 4-by-4 error covariance matrix for the 4-dimensional parameter set $\theta = \{z_1, z_2, z_3, \sigma^2\}$ for a general location solution. The Cramer-Rao lower bound [11] matrix is given by $C = I^{-1}$ where

$$I_{p,q} = -E\left\{\partial^2 Q(\theta) / \partial p \partial q\right\}$$

where p and q represent components of θ . The solution error ellipses are contours of constant value of the inner product $\theta^T I \theta$.

2.3.6 Recursion

The algorithm repeats the E-step and M-step until convergence. Solutions with very low solution weights are removed.

2.4 Algorithm Modifications

The algorithm as described above is a special case of the E-M algorithm for mixture densities. Some modifications may be necessary, however, which deviate from the E-M algorithm. In difficult problems, the majority of click pairs may be “invalid”, that is they are time delays measured between two clicks that are not both from the same acoustic click event. Normally, these should get assigned to the “error” solution. Nevertheless, because they often have time delays matching a given model solution by chance, their existence causes significant problems in the convergence of the algorithm. The algorithm can be modified to effectively deal with the problem. First, by using a power in the exponent of (1) higher than 2, the distribution is no longer Gaussian, but the effect of outlier time delays is minimized. Second, we use *click-based* solution weights. In the unmodified algorithm, the solution weights are *click-pair based*. They are proportional to the effective number of click-pairs assigned to each solution since they are derived from the CPAPM. To remove the influence of

invalid click pairs, we create *click-based* solution weights, based on a click detection association probability matrix (CDAPM). Let $\beta_{k,j}$ be the probability that detection k is assigned to solution j . We estimate $\beta_{k,j}$ by approximating the number of click pairs assigned to solution j containing click k . More precisely, $\beta_{k,j} = \sum_{i \in I_k} L_{i,j} \alpha_j$, where I_k is

set of click pairs containing click k . We then normalize $\beta_{k,j}$ in the same way as $w_{i,j}$. We estimate the solution weights α_j from $\beta_{k,j}$ instead of $w_{i,j}$.

3 COMBINING SINGLE AND MULTI-SENSOR ALGORITHMS

3.1 Assisting single-sensor separation with multi-sensor information

One way to assist the click-train separation is to develop a measure of the probability that two clicks, received at the same sensor, are from the same animal and same propagation path. This requires “support” from two additional clicks. Let clicks A_1 and B_1 be received at one sensor. If A_1 and B_1 are indeed from the same animal, then it is possible that these same two clicks have been received at another sensor. Denoting these two clicks as A_2 and B_2 , we would find that the time delays (A_1 to A_2) and (B_1 to B_2) should match to a high degree of accuracy. With this in mind, we create the match measure

$$r_{i,j} = \sum_{m,n} e^{-\frac{(\tau_{i,m} - \tau_{j,n})^2}{2\sigma_\tau^2}}, \quad (3)$$

where $\tau_{i,m}$ is the time delay between click detections i and m and σ_τ^2 is a time delay error variance parameter. Indexes i, j represent A_1 and B_1 . Indexes m, n represent all potential pairs A_2 and B_2 . The search is limited to likely candidates for A_2 and B_2 based on time delay limits. An example of matrix $r_{i,j}$ is shown in figure 4. The information is utilized in click train separation by adding $-\log r_{i,j}$ to the error metric $E_{i,j}$.

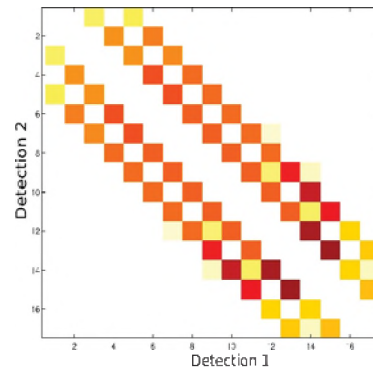


Fig. 4. Example of single-sensor inter-click error metric r_{ij} based on multi-sensor information. Darker shades indicate good match. Even numbered clicks associate with even clicks and odd with odd. This is a situation where there is one animal and each direct-path click detection is followed by a reverberation detection. Use of $r_{i,j}$ prevents association of reverberation and direct-path clicks.

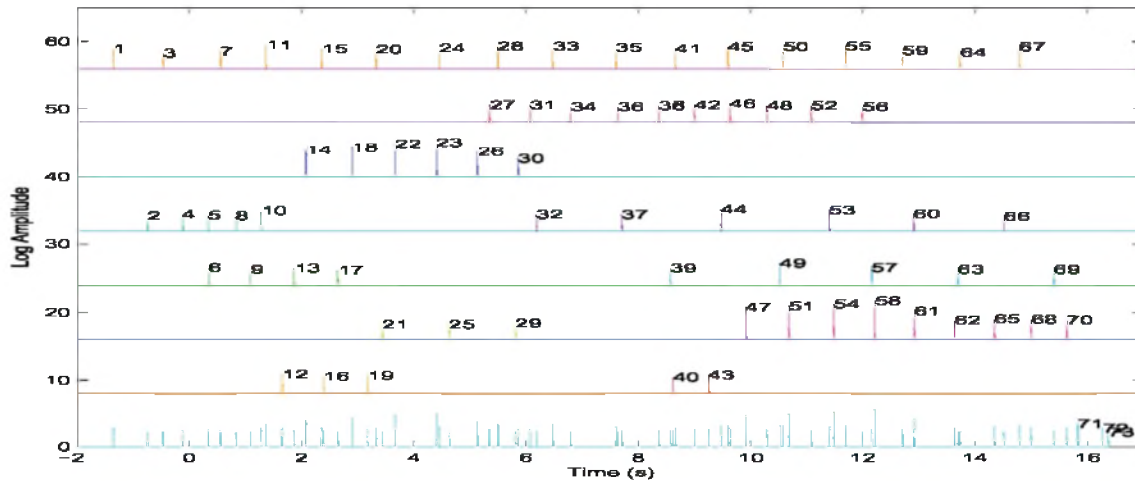


Fig. 5. Example of the sperm whale click train separation. Shown is the first 18 seconds of data from sensor 1 of 6. The lowest trace shows all clicks (log amplitude vs time). Separated click trains are displayed with vertical and/or horizontal separation. The presence of two nearby clicks of different amplitude gives the appearance of “crooked” clicks.

3.2 Assisting localization with single-sensor information

Going back to the example above, if click pair (A_1, A_2) is a valid pairing, then it stands to reason, there exists another pair (B_1, B_2) such that (a) A_1 and B_1 belong to the same click train, (b) A_2 and B_2 belong to the same click train, (c) pairs (A_1, A_2) and (B_1, B_2) have the same time delay. With this in mind, we create the click pair quality measure

$$q_i = \sum_{j \in J_i} e^{-\frac{(\tau_i - \tau_j)^2}{2\sigma_\tau^2}}, \quad (4)$$

where J_i is the set of click pairs that are from the same single-sensor click train as pair i . Click pairs with low values of q_i can be eliminated.

4 ALGORITHM SUMMARY

- 1) Make click detections on each sensor.
- 2) Develop a list of potential inter-sensor click pairings with time delay τ_i subject to constraints on time delay.
- 3) For each sensor:
 - a) For every pair of clicks (i, j) .
 - i) calculate single-sensor inter-click error metric E_{ij} (section 1.3).
 - ii) calculate inter-click match metric based on multi-sensor time delays (τ_{ij} in equation 3). Add $-\log \tau_{i,j} * E_{i,j}$.
 - b) Run the click separation algorithm described in section 1.
- 4) From the single-sensor separation results, develop the click pair metric q_i in equation (4).
- 5) Run the localization algorithm described in section 2.3:
 - a) Use a grid-search to find a set of potential whale

locations z_j . Initialize the solution variances σ_j^2 to reflect the time-delay variance corresponding to the grid-search quantization. Initialize the solution validity measures α_j to a constant $\alpha_j = 1/M$ where M is the number of solutions.

- b) Compute CDAPM $\beta_{k,j}$ and validity probabilities α_i (section 2.4). Eliminate locations with low α_i .
- c) For each j , maximize (2) over σ_j^2 and z_j and compute the Cramer-Rao lower bound covariance for z_j from which the error ellipses can be drawn. Go to step 5-b, repeat.

5 EXPERIMENTAL RESULTS

We utilized sperm whale data from bottom mounted sensors from the Monaco 2005 workshop. The data consisted of two sets which can be described as “easy” and “difficult”. Data set 2 (easy), was a 25-minute run where a single whale was present with high SNR. This data set was used to produce figure 4 and the lower trace in figure 2. Data set 1 (difficult) was a 20 minute run with two and possibly three whales. This data set was used to produce figures 5, through 9 and the upper trace in figure 2.

5.1 Click Separation Results

Using “difficult” data set, we obtained single sensor click-train separation results using the described recursive algorithm assisted by multi-sensor information using (3). Results from the first 18 seconds of sensor 1 are typical and are shown in figure 5. A total of 70 click events were grouped into 11 chains ranging from 2 to 17 clicks in length. From the first 18 seconds of data from 6 sensors, the single-sensor groupings were used to calculate q_i (equation 4). A total of 9564 multi-sensor click pairs and associated

time-delays were created. By setting a lower bound of $q_i = 0.8$, it was possible to eliminate 8643 of the 9564 click pairs. An initial set of 287 initial solutions were found by looking for local minima in the $x - y - z$ grid-search. Using the “easy” data set, the algorithm provided two very strong and tight solutions, the direct path.

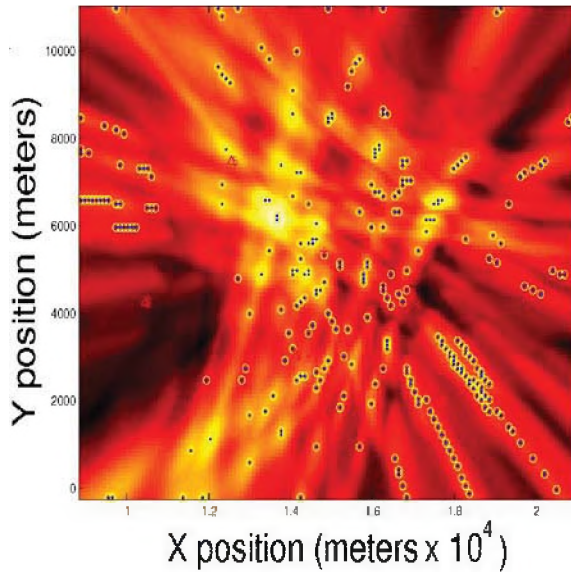


Fig. 6. Initial set of 287 solutions on a geographical plot. Initial solutions were local maxima of the function 1 searched over z_1, z_2, z_3 in a grid. Image intensity was maximized over depth (z_3).

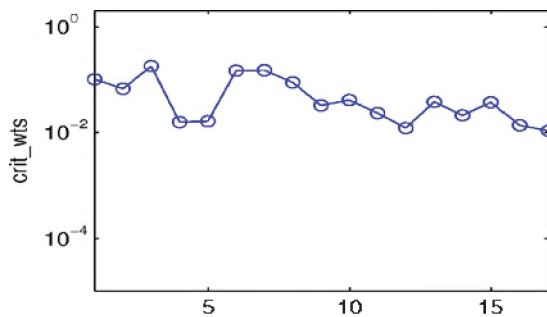


Fig. 7. Solution weights after 100 iterations. Solutions 3, 6, and 7 have appreciable weight.

5.2 Localization Results

The first 18 seconds of the “difficult” data set provides a good illustration of algorithm behavior. The data of figure 5 was one of six sensor inputs provided to the localization algorithm. The algorithm was able to positively identify the locations of two whales in the sensor range. After 100 iterations, the solution weights were as shown in figure 7. Solutions 3, 6, and 7 have appreciable weight. Solutions 6 and 7 were clearly whales and solution 3 was a potential whale, although the solution was ambiguous. The solution error ellipses are shown in figure 8. Error ellipses were obtained from the CR bound analysis of equation (2). The error ellipse for solution 3 resembles a line, an

indication that click pairs from only two sensors are available, or the positioning is unfavorable for exact localization. Ellipses for the valid whale solutions 6 and 7 are very tight. The clicks associated with each solution can be determined using the CDAPM. In Figure 9 we show all clicks, those associated with whale 1, and those associated with whale 2. The clicks have been time-aligned in accordance with the time delay from each sensor to each solution’s position.

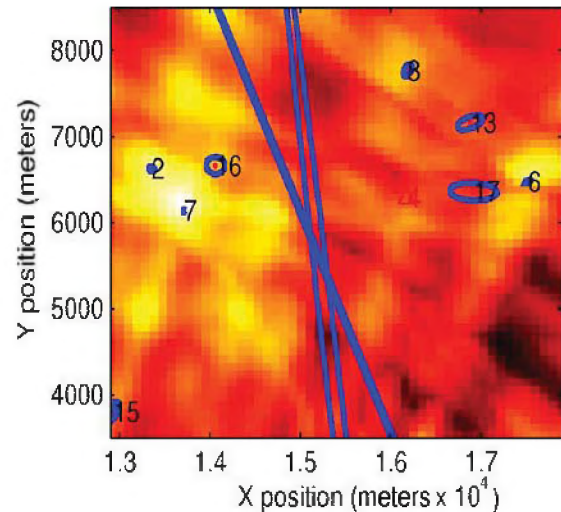


Fig. 8. Solution error ellipses. The diagonal line is actually an ellipse for solution 3 indicating wide error in one direction probably due to having information from just 2 sensors.

5.3 General comments and future work

The click separation algorithm worked perfectly in all 18-second segments of the “easy” data set, separating the direct from the multi-path chains. In the “difficult” data set, it appeared to work very well although it is difficult to validate the results without positive localization solutions for all clicks. By comparing localization and separation results from the first 18-seconds (i.e including figures 5 and 9) we were able to find no clear cases where clicks were improperly grouped together, but many cases where click trains were separated into smaller pieces. Much of this behavior can be controlled by the “new chain penalty” P .

The localization algorithm performed very well provided good initialization solutions are provided. A fine grid search was necessary causing the algorithm to be very slow, requiring up to five minutes to process the solution for 18 seconds of input data. However, as the number of solutions is reduced to a handful, a single E-M iteration requires only a fraction of a second. In the “difficult” data set, good whale locations were obtained in most of the 18-second intervals, and whales could be tracked throughout most of the data set, but it was not uncommon that invalid solutions dominated. It was clear that additional work was needed to eliminate these invalid solutions, possibly by using information from the separation algorithm.

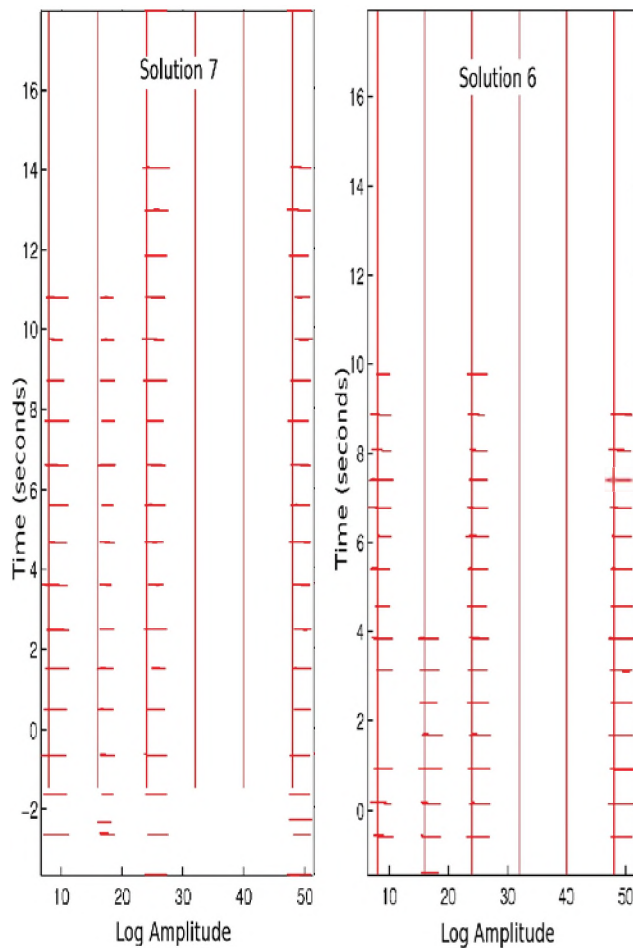


Fig. 9. Example of the sperm whale localization. Log-amplitude on the X-axis and time on the Y-axis.

For clarity, the clicks from various sensors are separated horizontally by artificially adding a different bias to the X-axis value of each sensor. Left panel: the clicks with high CDAPM value for solution 7 are shown. Sensors time axes have been time aligned with according to the respective solution to align the clicks. Right panel: clicks with high CDAPM value for solution 6 are shown. These two solutions represent less than half of the available clicks. The rest were associated with the error solution or a variety of invalid solutions, some of which appear in figure 8.

6 CONCLUSIONS

We have provided a snapshot of our research effort directed toward joint localization and separation of sperm whale click trains. We have presented two novel algorithms, one for separation, and one for localization, and demonstrated them on real data. The click-train separation algorithm uses an approach inspired by dynamic programming and efficiently seeks the globally best click grouping. The localization algorithm uses the EM algorithm to do “soft” association of click pairs to solutions. The algorithms are loosely tied together by (a) utilizing multi-sensor time delay information to assist the separation algorithm and (b) eliminating false click pairings in the localization algorithm by the use of information from the click separation algorithm. Because the localization algorithm uses the EM algorithm to make “soft” click and click-pair assignments, it may prove superior to existing approaches in many situations, such as when a large number of solution

ambiguities exist.

REFERENCES

- [1] M. van der Schaar, E. Delory, A. Catal, and M. Andr, “Neural network-based sperm whale click classification,” *J. Mar. Biol. Ass. U.K.*, vol. 87, pp. 35–38, 2007.
- [2] T. Ura, R. Bahl, M. Sakata, J. Kojima, T. Fukuchi, J. Ura, K. Mori, T. Nakatani, Y. Nose, and H. Sugimatsu, “Development of AUV-based system for acoustic tracking of diving sperm whales,” in *Proceedings of OCEANS 2004 (Kobe, Japan)*, pp. 2302–2307, November 2004.
- [3] R. Bahl and T. Ura, “Automatic real-time segregation and classification of multiple vocalizing sperm whales,” *Seisan-Kenkyu Bimonthly Journal of IIS, University of Tokyo*, vol. 55, pp. 61–64, May 2003.
- [4] R. E. Bellman, *Adaptive Control Processes*. Princeton, New Jersey, USA: Princeton Univ. Press, 1961.
- [5] R. A. Redner and H. F. Walker, “Mixture densities, maximum likelihood, and the EM algorithm,” *SIAM Review*, vol. 26, April 1984.
- [6] R. Morrissey, J. Ward, N. DiMarzio, S. Jarvis, and D. Moretti, “Passive acoustic detection and localization of sperm whales (*Physeter Macrocephalus*) in the Tongue of the Ocean,” *Applied Acoustics*, vol. 67, pp. 1091–1105, Nov 2006.
- [7] S. Jarvis and D. Moretti, “Passive detection and localization of transient signals from marine mammals using widely spaced bottom mounted hydrophones in open ocean environments,” in *International Workshop on the Application of Passive acoustics in Fisheries*, (NUWC), 2002.
- [8] P. White, T. Leighton, D. Finfer, C. Powles, and O. Baumann, “Localisation of sperm whales using bottom-mounted sensors,” *Applied Acoustics*, vol. 67, pp. 1074–1090, 2006.
- [9] C. O. Tiemann and M. B. Porter, “Automated model-based localization of sperm whale clicks,” in *OCEANS 2003. Proceedings*, vol. 2, pp. 821–827, Sept 2004.
- [10] R. Hogg, J. McKean, and A. Craig, *Introduction to Mathematical Statistics*. Upper Saddle River, NJ: Pearson Prentice Hall, 2005.
- [11] S. Kay, *Modern Spectral Estimation: Theory and Applications*. Prentice Hall, 1988.

PAIR-WISE SPECTROGRAM PROCESSING USED TO TRACK A SPERM WHALE

Eva-Marie Nosal and L. Neil Frazer

School of Ocean and Earth Science and Technology, 1680 East-West Road, Honolulu HI 96822

ABSTRACT

We developed a model-based localization method called pair-wise spectrogram (PWS) processing to track marine mammals using widely spaced hydrophone arrays [Nosal & Frazer, *IEEE J. Ocean. Eng.*, in press]. Here we use PWS to track the sperm whale from the dataset provided for the *2005 Workshop on detection and localization of marine mammals using passive acoustics (Monaco)*. This dataset provides a good opportunity to validate and explore the properties of the PWS processor. We demonstrate the relationship between pair-wise processing and a time-of-arrival method for a simple case. We also show how varying the size of windows used to create spectrograms optimizes the tradeoff between processor resolution and robustness, and how these parameters can be adjusted according to grid spacing. PWS position estimates are within tens of meters of those obtained using a careful time-of-arrival method applied to individual clicks.

RÉSUMÉ

Nous avons développé une méthode de localisation basée sur un modèle de propagation du son, appelée traitement de spectrogramme par paires (PWS : pair-wise spectrogram), pour déterminer les trajectoires de mammifères marins à l'aide de réseaux d'hydrophones très espacés [Nosal & Frazer, *IEEE J. Ocean. Eng.*, in press]. Nous présentons une application de cette méthode visant à déterminer la trajectoire d'un cachalot à partir du jeu de données fourni par l'*Atelier 2005 sur la détection et la localisation des mammifères marins à l'aide du repérage acoustique passif (Monaco)*. Ce jeu constitue une bonne occasion de valider et d'explorer les propriétés du PWS. Nous démontrons la relation entre le traitement par paire et la méthode des temps d'arrivée pour ce cas simple. Nous montrons aussi comment la modification de la taille des fenêtres utilisées pour créer les spectrogrammes permet d'optimiser le compromis entre la résolution et la robustesse du traitement et comment ces paramètres peuvent être ajustés en fonction de la taille de la grille. La différence entre les positions estimées avec PWS et celles obtenues en utilisant la méthode des temps d'arrivée appliquée aux clicks individuels est de l'ordre d'une dizaine de mètres.

1 INTRODUCTION

The most commonly used methods for tracking marine mammals are time difference of arrival (TDOA) methods [e.g. Watkins and Schevill 1972; Clark *et al.* 1986; Spiesberger and Fristrup 1990; Janik *et al.* 2000]. In TDOA methods, the difference in time of arrival between pairs of hydrophones is estimated, usually via cross-correlation of waveforms or spectrograms. Each receiver pair defines a hyperboloid, and the intersection of hyperboloids (from various receiver pairs) defines the position of the source. Depending on the receiver geometry, four or five receivers are required to localize the source in three dimensions [Spiesberger 2001]. Reflections from the bottom and surface can be treated as recordings made by virtual receivers [Urick 1983]. Using reflections improves the accuracy of estimated source positions [Mohl *et al.* 1990; Wahlberg *et al.* 2001; Thode *et al.* 2002] and reduces the number of required receivers

[Aubauer *et al.* 2000; Tiemann *et al.* 2007; Laplace 2007].

TDOA methods are usually implemented with an isospeed assumption. This has the advantage of providing closed-form solutions and rapid run-times. It is acceptable in many cases (e.g. nearly isospeed conditions, relatively short distance propagation), particularly when care is taken to account for the resulting errors [Wahlberg *et al.* 2001; Spiesberger and Wahlberg 2002]. In other cases, a depth-dependent sound speed profile can significantly improve position estimates [Chapman 2004; Tiemann *et al.* 2004; Nosal and Frazer 2006]. To remove isospeed assumptions, model-based TDOA methods can be implemented using a matched field approach [Tiemann *et al.* 2004; Nosal and Frazer 2006, 2007]. TDOAs are estimated (as before), a 3D grid of candidate source location is created, and TDOAs are modeled repeatedly for a source at each of the grid points. The modeled

TDOAs are then compared to the measured TDOAs, and the estimated source location is the one that gives the best agreement. Therefore, we can regard the TDOA method as a matched-field method in which arrival time is the only part of the field being “matched.”

Pair-wise spectrogram (PWS) processing [Nosal and Frazer in press] extends model-based TDOA methods by matching amplitude and phase information in addition to arrival times. Simulations have shown [Nosal and Frazer in press] that pair-wise processing is a promising passive acoustic localization method, but it has yet to be tested on real data. In this paper, a single sperm whale dataset is used to validate and explore the processor. Since TDOA methods give very good position estimates for this dataset, they are used to “ground-truth” the PWS position estimates.

2 DATA

The dataset was made available by Naval Undersea Warfare Center for the 2nd International Workshop on Detection and Localization of Marine Mammals Using Passive Acoustics [Adam *et al.* 2006]. It features a single sperm whale producing regular clicks (on average 1.06 clicks/s) for 25 minutes. Recordings are from 5 widely-spaced bottom-mounted hydrophones at the Atlantic Undersea Test and Evaluation Center in the Tongue of the Ocean (off Andros Island, Bahamas). Signal to noise ratios vary between receivers, and clicks, and are typically between 2 and 30 dB. The sampling rate is 48 kHz. Filtering, phone sensitivity, and directivity are unknown. Hydrophone positions are given in Table 1. The track of this sperm whale has been found using various TDOA methods [Nosal and Frazer 2006, 2007; Giraudet *et al.* 2006; Morrissey *et al.* 2006; White *et al.* 2006].

Table 1. Hydrophone positions provided by NUWC

Phone	x-pos (m)	y-pos (m)	depth (m)
G	10658.04	-14953.63	1530.55
H	12788.99	-11897.12	1556.14
I	14318.86	-16189.18	1553.58
J	8672.59	-18064.35	1361.93
K	12007.50	-19238.87	1522.54

3 PWW & PWS processing

3.1 Symbols and notation

- s source waveform (time domain)
- S source waveform (frequency domain)
- r_i received signal at phone i (time)
- R_i received signal at phone i (frequency)
- \hat{G}_i modeled impulse response at phone i (time)

- \hat{G}_i modeled impulse response at phone i (frequency)
- N number of samples in a signal
- N_r number of receivers
- $*$ complex conjugate
- ω radian temporal frequency

3.2 Overview

This section gives a brief overview of pair-wise processing. Complete details can be found in [Nosal and Frazer in press]. A 3D grid of candidate source location is created, and at a given candidate source location, the pair-wise waveform (PWW) processor is given by:

$$\rho_{pww} = \frac{\sum_{n=1}^N \sum_{i=1}^{N_r} \sum_{j \neq i}^{N_r} H_{ij}^*(\omega_n) H_{ji}(\omega_n)}{\sum_{n=1}^N \sum_{i=1}^{N_r} \sum_{j \neq i}^{N_r} |H_{ij}(\omega_n)|^2} \quad (1)$$

where $H_{ij}(\omega_n) = R_i(\omega_n) \hat{G}_j(\omega_n)$. This processor is maximized at the correct source location, since there the modeled impulse response is approximately equal to the true impulse response, *i.e.* $\hat{G}_i \approx G_i$ (not exactly equal because the model is never perfect), so that $H_{ij} = R_i \hat{G}_j = S G_i \hat{G}_j \approx S G_j \hat{G}_i = R_j \hat{G}_i = H_{ji}$.

Appendix A.3 shows that if only arrival time is used, pair-wise waveform (PWW) processing is equivalent to TDOA methods in a limiting case. Pair-wise processing assumes that all hydrophones have the same impulse response, but it makes no assumption about the spectrum of the source [Frazer and Sun 1998].

PWS processing is similar to PWW processing, except that spectrograms are processed instead of waveforms. This is useful because spectrograms are less sensitive to imperfections in the model due to uncertainties in sound speed profiles, receiver positions, and so on (see Appendix A.4 for an explanation). The PWS processor is:

$$\rho_{pws} = \frac{\sum_m \sum_n \sum_{i=1}^{N_r} \sum_{j \neq i}^{N_r} \tilde{H}_{ij}^*(T_m, f_n) \tilde{H}_{ji}(T_m, f_n)}{\sum_m \sum_n \sum_{i=1}^{N_r} \sum_{j \neq i}^{N_r} |\tilde{H}_{ij}(T_m, f_n)|^2} \quad (2)$$

where \tilde{H}_{ij} denotes the spectrogram of H_{ij} and the sums are over all spectrogram time (T_m) and frequency (f_n) channels.

In nearly all problems of interest the number of sources is unknown, hence the output of ρ_{pww} and ρ_{pws} is potentially

multimodal and should be generated on a grid of candidate locations.

3.3 Processing of data

The PWS results presented here used the raw data only (no pre-processing) and the Gaussian beam acoustic propagation model BELLHOP [Porter and Bucker 1987; Porter 2005] was used to model Green's functions. Received signals were split into 30 s segments that overlapped by 20 s and each segment was processed separately. A 30 s segment length was chosen as an optimum after trials with different length segments. Shorter segments gave less consistent position estimates, possibly because some segments didn't contain enough clicks, and longer segments reduced performance, probably due to movement of the animal.

Within each segment, an important parameter is the window length used to create spectrograms. Simulations have shown [Nosal and Frazer in press] that longer window lengths give position estimates that are not as sensitive to uncertainties in the bathymetry, sound speed

profile, receiver positions, and so on. With longer windows, the peaks of the likelihood surfaces are broader, which means that coarser grid spacing can be used. Figure 1 illustrates this using the first 30 s of the dataset. Spectrograms were all created using Hanning windows with 50% overlap. With a grid spacing of 200 m, a window length of 256 ms gives a good first estimate of the animal's position, while a window length of 32 ms does not (the animal is lost somewhere between grid points). With finer grid spacing (Fig. 2), spectrogram windows can be shortened to get increasingly precise position estimates. Trial runs indicated that a good window length is the travel time of sound between two grid points, *i.e.* window length should be roughly $2\Delta d/c$ for a grid spacing of Δd and speed of sound $c \approx 1500$ m/s.

Larger grid spacing gives faster run-times (with fewer points to process) so the tradeoff between robustness and resolution can be used to reduce computational requirements. Rough position estimates were made by running the processor for a large area with 200 m grid spacing and spectrogram window lengths of 256 ms.

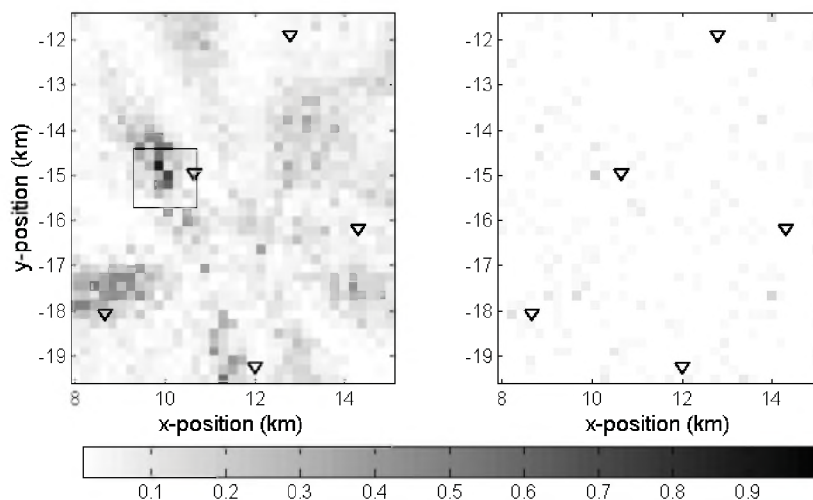


Figure 1. Plan view of the likelihood surface at a single depth of 685 m (approximately the correct depth of the animal) for the first 30 s of the dataset and a 200 m grid spacing. Receiver positions are indicated by triangles. Spectrograms use Hanning windows with 50% overlap and window lengths of (a) 256 ms, and (b) 32 ms. With such coarse grid spacing, 32 ms windows are too short to give a position estimate (unless the animal is near a grid point). The box in (a) indicates the area shown in Figure 2.

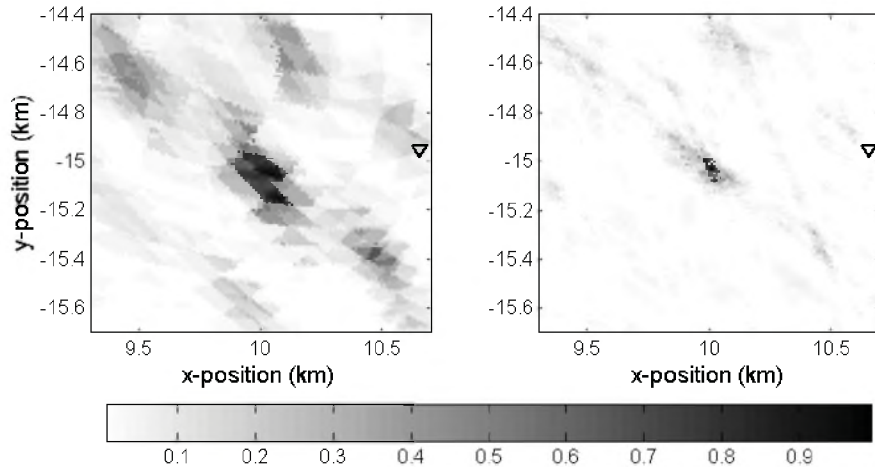


Figure 2. The boxed area shown in Fig 1 (a) processed here with 10 m grid spacing. With such fine grid spacing, 32 ms windows (b) give a more precise position estimate than 256 ms windows (a).

Rough position estimates were refined by processing smaller areas (surrounding the most promising grid point only) with 50 m grid spacing and spectrogram window lengths of 64 ms. Final estimates were obtained by further refining these with 10 m grid spacing and 16 ms windows. Resulting position estimates are shown in Figure 3. They are all within 40 m (and usually within 10 m) of those obtained with the TDOA method by Nosal and Frazer [2007], which have estimated 95% confidence interval half-widths of less than 25 m (details of this TDOA method are provided in Nosal and Frazer [2007]).

Error estimates were made as in Nosal and Frazer [2007] using conditional likelihood functions (CLFs). To get error in x at a given time step, y and z were fixed at the estimated source location and the resulting CLF was summed across x position to get the cumulative CLF for x (this can be thought of as summing along the horizontal strip that passes through the brightest point in Figure 2). 95% confidence intervals (CI) correspond to the distance between the x -positions at which the conditional CLF attains 2.5% and 97.5%, respectively. CIs for y and z were obtained analogously. For all dimensions, and over the whole track, CI half widths were within 45 m.

Note that error estimates cannot be directly compared because the TDOA method gave position estimates for every click while the PWS estimates are for 30 s intervals. It is evident, however, that the TDOA position estimates are more accurate for this dataset. On average, the whale moved 12, 40, and 9 m in the x , y , and z directions, respectively, in 30 s so that CI half widths within 45 m for the PWS methods are reasonable and may be partially explained by animal movement within a segment.

4 DISCUSSION

When deciding between localization methods, it is important to consider the tradeoff between the accuracy and power of the processor on one hand, and the computational demands and modeling complexity on the other hand. For the dataset considered here, with a single animal and very clear arrivals, the TDOA method excels since it gives better position estimates and is faster and easier to implement (to process 23 minutes of data, the TDOA method took 20 min while PWS processing took 43 minutes). Nevertheless, the PWS did successfully track the animal, which validates the PWS processor for this simple case. Future work will deal with more complicated datasets (more noise, multiple animals, shallower water, and long-duration calls) for which the PWS processor was developed [Nosal and Frazer in press].

One consideration that has not been addressed in PWS processing is source directivity. To use amplitudes, PWS assumes an omni-directional source; this does not hold for most marine mammals. Even though sperm whale clicks are highly directional [Möhl *et al.* 2003], the problem did not affect our results for this dataset, probably because there was enough information in the arrival times to overcome it. One solution for cases where directionality is a problem may be to use lower frequencies only, which are not as directional as higher frequencies. High frequencies might still contribute to the position estimates if spectrograms are thresholded to contribute only time of arrival information. Another (more difficult, but possibly more useful) approach is to include directivity in the source model and animal orientation in the search space.

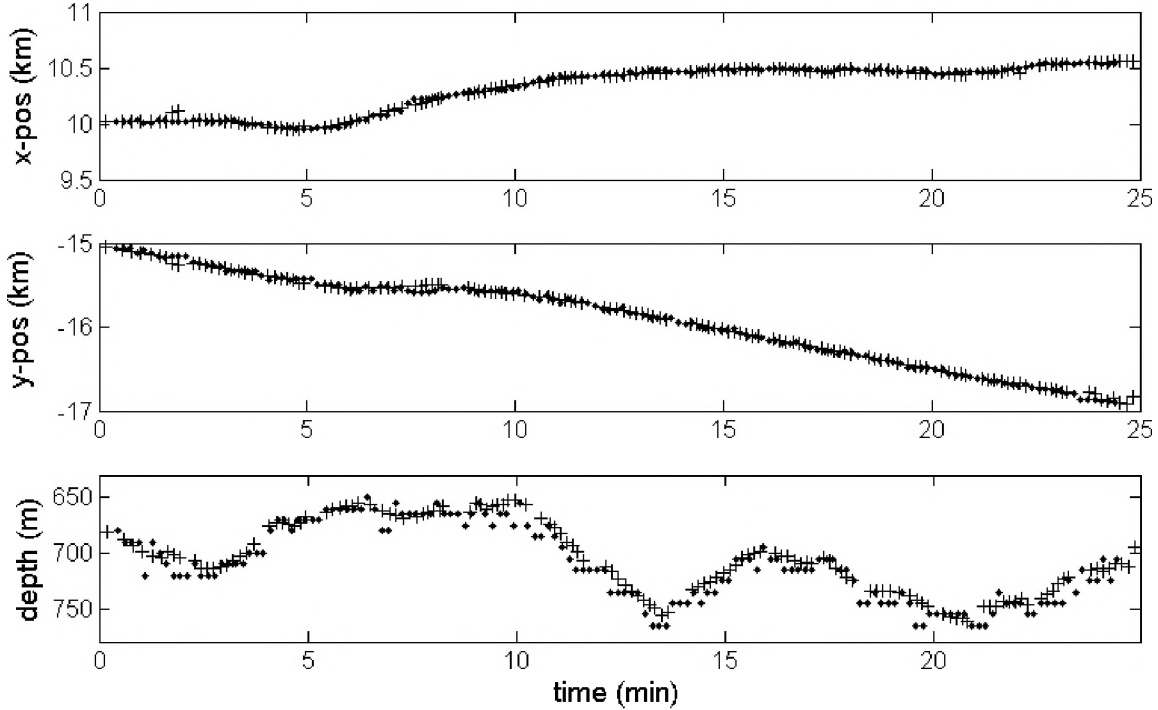


Figure 3. Whale positions estimated using PWS processing (dots) and TDOA processing (crosses).

APPENDIX

A.1 Notation

The appendix uses the same symbols and notation as in Section 3.1. The following are added:

\otimes convolution:

$$(f \otimes g)(t_m) = \sum_{n=1}^N f(t_n)g(t_m - t_n)$$

\circ cross-correlation:

$$(f \circ g)(t_m) = \sum_{n=1}^N f(t_n)g(t_m + t_n)$$

$\delta(t_n - \tau)$ unit impulse.

$$\delta(t_n - \tau) = \begin{cases} 1 & \text{when } t_n = \tau \\ 0 & \text{otherwise} \end{cases}$$

Discrete fourier transform of f , $DFT(f)$, is denoted using upper case:

$$f(t_n) \xleftarrow{DFT} F(\omega_n)$$

Inner product: $\langle f, g \rangle = \sum_{n=1}^N f(t_n)g^*(t_n)$

A.2 Required properties/relationships:

$$\delta(t_n - \tau) \otimes f(t_n) = f(t_n - \tau) \quad (P1)$$

$$\delta(t_n - \tau) \circ f(t_n) = f(t_n + \tau) \quad (P2)$$

$$f \circ g \xleftarrow{DFT} F^*G \quad (P3)$$

$$\text{Power theorem: } \langle f, g^* \rangle = \frac{1}{N} \langle F, G^* \rangle \quad (P4)$$

A.3 TDOA vs PWW for a simple case

This appendix aims to provide some intuition about the PWW processor and its relationship to the TDOA method. Consider the limiting case of infinitely fine grid spacing. Also consider unit impulses for the whale signal and the true and modeled impulse responses: respectively, $s(t_n) = \delta(t_n - \tau_0)$; $g_i(t_n) = \delta(t_n - \tau_i)$; and $\hat{g}_i(t_n) = \delta(t_n - \hat{\tau}_i)$, where τ_0 is the time at which the whale produces the impulse, τ_i is the true direct arrival time at receiver i , and $\hat{\tau}_i$ is the modeled direct arrival time at receiver i . A delayed impulse is an unrealistic Green's function, but it is useful for gaining insight.

By (P1) the received signal at receiver i is:

$$r_i(t_n) = s \otimes g(t_n) = \delta(t_n - \tau_0 - \tau_i). \quad (\text{A1})$$

The cross-correlation of the received signals r_i and r_j at hydrophone pair $i-j$ is then (by P2):

$$\begin{aligned} r_i \circ r_j(t_n) &= \delta(t_n - \tau_0 - \tau_i) \circ \delta(t_n - \tau_0 - \tau_j) \\ &= \delta(t_n + \tau_0 + \tau_i - \tau_0 - \tau_j) \\ &= \delta(t_n + \tau_i - \tau_j) \end{aligned} \quad (\text{A2})$$

Similarly, the cross-correlation of the modeled impulse responses $\hat{g}_i(t)$ and $\hat{g}_j(t)$ at hydrophone pair $i-j$ is

$$\hat{g}_i \circ \hat{g}_j(t_n) = \delta(t_n + \hat{\tau}_i - \hat{\tau}_j) \quad (\text{A3})$$

In the TDOA method, $\tau_i - \tau_j$ is found as the time that maximizes $r_i \circ r_j$. The best candidate source location is the one that minimizes the difference between $\tau_i - \tau_j$ and $\hat{\tau}_i - \hat{\tau}_j$. In our unit impulse case, this is equivalent to finding the source location that maximizes the inner product between $r_i \circ r_j$ and $\hat{g}_i \circ \hat{g}_j$ since

$$\begin{aligned} \langle r_i \circ r_j, \hat{g}_i \circ \hat{g}_j \rangle &= \sum_{n=1}^N \delta(t_n + \tau_i - \tau_j) \delta(t_n + \hat{\tau}_i - \hat{\tau}_j) \\ &= \begin{cases} 1 & \text{if } \tau_i - \tau_j = \hat{\tau}_i - \hat{\tau}_j \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (\text{A4})$$

Since the signals are impulses, the denominator in the PWW processor, Eq. (1), for a single receiver pair is N . Accordingly, for receiver pair $i-j$ the PWW processor is (by P3 and P4):

$$\begin{aligned} \rho_{pww} &= \frac{1}{N} \sum_{n=1}^N \left(R_i(\omega_n) \hat{G}_j(\omega_n) \right)^* R_j(\omega_n) \hat{G}_i(\omega_n) \\ &= \frac{1}{N} \sum_{n=1}^N \left(R_i^*(\omega_n) R_j(\omega_n) \right) \left(\hat{G}_i(\omega_n) \hat{G}_j^*(\omega_n) \right) \\ &= \frac{1}{N} \left\langle DFT(r_i \circ r_j), \left(DFT(\hat{g}_i \circ \hat{g}_j) \right)^* \right\rangle \\ &= \langle r_i \circ r_j, \hat{g}_i \circ \hat{g}_j \rangle \end{aligned} \quad (\text{A5})$$

Compare this to (A4) to see that the TDOA method and the PWW processor are equivalent for a single receiver pair $i-j$. In principle, the argument can be extended to include surface and bottom reflections and to show that TDOA methods that use cross-correlation of spectrograms to find TDOAs are equivalent to PWS processing, within the limits of the geometrical acoustics approximation to a wavefield.

In general, differences between the methods include the way in which different receiver pairs are combined.

Moreover, PWW retains amplitude and phase information from the signals, while the TDOA method keeps only time of arrival information.

A.4 PWW vs PWS processing

The simple example in A.3 illustrates the motivation for using spectrograms in pair-wise processing instead of waveforms. With finite grid spacing, $\tau_i - \tau_j$ is never exactly equal to $\hat{\tau}_i - \hat{\tau}_j$. Even for a candidate source location exactly at the correct source location, the environment and model are imperfect, *i.e.* $\tau_i - \tau_j \neq \hat{\tau}_i - \hat{\tau}_j$, so by (A4) the processor is always zero. In a real case with non-unit impulses, the processor is not zero but attains only low values. By taking spectrograms, arrivals are smeared in time, so they don't need to match as precisely. This gives poorer resolution (positions are not found as precisely) but allows for the use of coarser grids (see Section 3.3) and makes the PWS processor more robust with respect to environmental and modeling uncertainties than the PWW processor (demonstrated using simulations in Nosal and Frazer [in press]).

ACKNOWLEDGMENT

This work was funded by the Office of Naval Research. The authors thank Mélanie Abecassis for help in translating the abstract.

REFERENCES

- Adam O, J-F Motsch, F Desharnais, NA DiMarzio, D Gillespie, RC Gisiner (2006). Overview of the 2005 workshop on detection and localization of marine mammals using passive acoustics. *Appl. Acoust.* 67, 1061-1070.
- Aubauer R, MO Lammers, WWL Au (2000). One-hydrophone method of estimating distance and depth of phonating dolphins in shallow water. *J. Acoust. Soc. Am.* 107(1), 2744-2749.
- Chapman DMF (2004). You can't get there from here: Shallow water sound propagation and whale localization. *Can. Acoust.* 32(2), 167-171.
- Clark CW, WT Ellison, K Beeman (1986). Acoustic tracking of migrating bowhead whales. *Proceedings of IEEE Oceans* 1986, 341-346.
- Frazer LN, Sun X (1998). New objective functions for waveform inversion. *Geophysics* 63, 1-10.
- Giraudet P, H Glotin H (2006). Real-time 3D tracking of whales by precise and echo-robust TDOAs of clicks extracted from 5 bottom-mounted hydrophones records of the AUTECH. *Appl. Acoust.* 67, 1106-1117.

- Janik VM, SM VanParijs, PM Thompson (2000). A two-dimensional acoustic localization system for marine mammals. *Mar. Mamm. Sci.* 16, 437-447.
- Laplace C (2007). A Bayesian method to estimate the depth and the range of phonating sperm whales using a single hydrophone. *J. Acoust. Soc. Am.* 121(3), 1519-1528.
- Möhl B, A Surlykke, LA Miller (1990). High intensity narwhal clicks. pp. 295-303 in J Thomas, R Kastelein (Eds.). *Sensory ability of cetaceans*, Plenum Press, New York.
- Möhl B, M Wahlberg, PT Madsen, A Heerfordt, A Lund (2003). The monopulsed nature of sperm whale clicks. *J. Acoust. Soc. Am.* 114, 1143-1154.
- Morrissey RP, J Ward, N DiMarzio, S Jarvis, DJ Moretti (2006). Passive acoustic detection and localization of sperm whales (*Physeter macrocephalus*) in the tongue of the ocean. *Appl. Acoust.* 67, 1091-1105.
- Nosal E-M, LN Frazer (2006). Delays between direct and reflected arrivals used to track a single sperm whale. *Appl. Acoust.* 87 (11-12), 1187-1201.
- Nosal E-M, LN Frazer (2007). Sperm whale three-dimensional track, swim orientation, beam pattern, and click levels observed on bottom-mounted hydrophones. *J. Acoust. Soc. Am.* 122(4), 1969-1978.
- Nosal E-M, LN Frazer (in press). Modified pair-wise spectrogram processing for localization of unknown broadband sources. *IEEE J. Ocean Eng.*
- Porter MB, HP Bucker (1987). Gaussian beam tracing for computing ocean acoustic fields. *J. Acoust. Soc. Am.* 82(4), 1349-1359.
- Porter MB (2005). *BELLHOP*. Last accessed on Dec 9, 2007 at < <http://oalib.hlsresearch.com/>>.
- Spiesberger JL (2001). Hyperbolic location errors due to insufficient numbers of receivers. *J. Acoust. Soc. Am.* 109(6), 3076-3079.
- Spiesberger JL, KM Fristrup (1990). Passive localization of calling animals and sensing of their acoustic environment using acoustic tomography. *Am. Nat.* 135, 107-153.
- Spiesberger JL, M Wahlberg (2002). Probability density functions for hyperbolic and isodiachronic locations. *J. Acoust. Soc. Am.* 112(6), 3046-3052.
- Thode A, DK Mellinger, S Stienessen, A Martinez, K Mullin (2002). Depth-dependent acoustic features of diving sperm whales (*Physeter macrocephalus*) in the Gulf of Mexico. *J. Acoust. Soc. Am.* 112(1), 308-321.
- Tiemann CO, MB Porter, LN Frazer (2004). Localization of marine mammals near Hawaii using an acoustic propagation model. *J. Acoust. Soc. Am.* 115(6), 2834-2843.
- Tiemann CO, AM Thode, J Straley, V O'Connell, K Folkert (2007). Three-dimensional localization of sperm whales using a single hydrophone. *J. Acoust. Soc. Am.* 120(4), 2355-2365.
- Urick RJ (1983). *Principles of underwater sound*, 3rd ed. Peninsula, Los Altos, CA.
- Wahlberg M, B Möhl, PT Madsen (2001). Estimating source position accuracy of a large-aperture hydrophone array for bioacoustics. *J. Acoust. Soc. Am.* 109(1), 397-406.
- White PR, TG Leighton, DC Finfer, C Prowles, C Baumann (2006). Localisation of sperm whales using bottom-mounted sensors. *Appl. Acoust.* 67, 1074-1090.



Photo Credit: Greg Schorr. Copyright: Cascadia Research, Olympia, WA, USA

WHALE COCKTAIL PARTY: REAL-TIME MULTIPLE TRACKING AND SIGNAL ANALYSES

Hervé Glotin^A, Frédéric Caudal^A, Pascale Giraudet^B

A-System & Information Sciences Laboratory (LSIS - UMR CNRS 6168)

B-Department of Biology

Université du Sud Toulon Var - BP 20132 - 83957 La Garde Cedex - France.

{glotin, caudal, giraudet}@univ-tln.fr

ABSTRACT

This paper provides a real-time passive acoustic method to track multiple vocalizing whales using four or more omni-directional widely-spaced bottom-mounted hydrophones. Since the interest in marine mammals has increased, robust and real-time systems are required. To meet these demands, a real-time tracking algorithm was developed. After non-parametric Teager-Kaiser-Mallat signal filtering, rough Time Delays Of Arrival are calculated, selected and filtered, and used to estimate the positions of whales for a constant, linear or estimated sound speed profile. The complete algorithm is tested on real data from NUWC¹ and AUTE². Our model is validated by similar results from the US Navy³ and SOEST⁴ University of Hawaii Laboratory in the case of one whale, and by similar results from the Columbia University ROSA⁵ Laboratory for the case of multiple whales. At this time, our tracking method is the only one which provides typical speed and depth estimates for multiple vocalizing whales.

RÉSUMÉ

Ce papier propose une méthode temps-réel de trajectographie par acoustique passive de plusieurs cétacés émettant simultanément en utilisant un réseau d'au moins 4 hydrophones espacés de quelques centaines de mètres. Étant donné l'intérêt accru pour les mammifères marins, des systèmes temps-réel et robustes sont nécessaires. Pour répondre à cette demande, un algorithme temps-réel de trajectographie multiple a été développé. Après un filtrage non paramétrique Teager-Kaiser-Mallat du signal, les différences de temps d'arrivée aux hydrophones sont estimées, sélectionnées, filtrées, et permettent d'estimer les positions des baleines pour un profil de célérité constant, linéaire ou estimé. L'algorithme est testé sur des données réelles du NUWC¹ et de l'AUTE². Notre modèle est validé par des résultats similaires de l'US Navy³ et du laboratoire SOEST⁴ de l'université d'Hawaii dans le cas d'émissions simples, et par une estimation du nombre de baleines du laboratoire ROSA⁵ de l'université de Columbia dans le cas de plusieurs émissions simultanées. Actuellement, notre méthode de trajectographie est la seule donnant, dans le cas de plusieurs baleines, des vitesses et des profondeurs vraisemblables.

1 INTRODUCTION

Processing of Marine Mammal (MM) signals for passive oceanic acoustic localization is a problem that has recently attracted attention in scientific literature and in some organizations like AUTE² and NUWC. Motivation for processing MM signals stems from increasing interest in the behavior of endangered MM. One of the goals of current research in this field is to develop tools to localize the vocalizing and clicking whale for species monitoring. In this

paper we propose a low cost time-domain tracking algorithm based on passive acoustics. The experiments of this paper consist of tracking an unknown number of sperm whales (*Physeter catodon*). Clicks are recorded on two datasets of 20 and 25 minutes on an open-ocean widely-spaced bottom-mounted hydrophone array. The output of the method is the track(s) of the MM(s) in 3D space and time. This paper deals with the 3D tracking of MM using a widely-spaced bottom-mounted array in deep water - two main requirements for the localization technique presented here. It focuses on sperm

¹Naval Undersea Warfare Center of the US Navy

²Atlantic Undersea Test & Evaluation Center - Bahamas

³NUWC

⁴School of Ocean and Earth Science and Technology

⁵Recognition and Organization of Speech and Audio

whale clicks; detection and classification are not a concern. There were previous algorithms developed in the state of art [3, 12, 11] but none are able to have satisfying results for multiple tracks. Most of them are far from being real-time. The main goal is to build a robust and real-time tracking model, despite ocean noise, multiple echoes, imprecise sound speed profiles, an unknown number of vocalizing MM, and the non-linear time frequency structure of most MM signals [7]. Background ocean noise results from the addition of several noises: sea state, biological noises, ship noise and molecular turbulence. Propagation characteristics from an acoustic source to an array of hydrophones include multipath effects (and reverberations), which create secondary peaks in the Cross-Correlation (CC) function that the generalized CC methods cannot eliminate. Here we improve the algorithm from [3] to build a robust 3D tracking algorithm. In Section 2 we propose a time-domain algorithm for MM transient call localization. In Section 3 we show and compare results of tracks estimates with results from other specialists teams.

2 MATERIAL AND METHOD

The signals are records from the ocean floor near Andros Island – Bahamas⁶, provided with celerity profiles and recorded in March 2002. Datasets are sampled at 48 kHz and contain MM clicks, whistles, and background noises like distant engine boat noises. Dataset1 (D1) is recorded on hydrophones 1 to 6 with 20 min length while dataset2 (D2) is recorded on hydrophones 7 to 11 with 25 min length. We will use a constant sound speed with $c = 1500 \text{ m s}^{-1}$ and estimated celerity profile, or a linear profile with $c(z) = c_0 + gz$ where z is the depth, $c_0 = 1542 \text{ m s}^{-1}$ is the sound speed at the surface and $g = 0.051 \text{ s}^{-1}$ is the gradient. Sound source tracking is performed by continuous localization in 3D using Time Delays Of Arrival (noted T) estimation from four hydrophones.

2.1 Signal filtering

A sperm whale click is a transient increase of signal energy lasting about 20 ms (Figure 1-a). Therefore, we use the Teager-Kaiser (TK) energy operator on the raw data. The TK operator is defined for a discrete time signal as [8]:

$$\Psi[x(n)] = x^2(n) - x(n+1)x(n-1), \quad (1)$$

where n denotes the sample number. An important property of the TK energy operator in Eq.(1) is that it is nearly instantaneous given that only three samples are required in the energy computation at each time instant. Considering the raw signal as:

$$s(n) = x(n) + u(n),$$

where $s(n)$ is the raw signal, $x(n)$ is the signal of interest (clicks), $u(n)$ is an additive noise defined as a process realization considered wide sense stationary (WSS) Gaussian during a short time, by applying the TK operator to $s(n)$, $\Psi[s(n)]$ can be expressed as [9]:

$$\Psi[s(n)] \approx \Psi[x(n)] + w(n),$$

where $w(n)$ is a random gaussian process whose parameters are in [9]. The output is dominated by the clicks energy. Then, the sampling frequency is reduced to 480 Hz by the mean of 100 adjacent bins to reduce the variance of the noise and the data size. We apply the Mallat algorithm [10] with the Daubechies wavelet (order 3). We chose this wavelet for its great similarity to the shape of a decimated click [2]. The signal is denoised with a soft universal thresholding. This thresholding is defined as $D(uk, \lambda) = \text{sgn}(uk) \max(0, |uk| - \lambda)$, with uk the wavelet coefficients, $\lambda = \sqrt{(2 \log_e(Q)) \sigma_N \sigma_{\tilde{N}}}$, and Q is the length of the resolution level of the signal to denoise [1]. The noise standard deviation σ_N is calculated on each 10s window on the raw signal with a maximum likelihood criterion. $\sigma_{\tilde{N}}$ is the standard deviation of the wavelet coefficients on a resolution level of a generated, reduced and 0-mean Gaussian noise. This filtering step is very fast and does not need any parameter. Figure 1-c and 1-f are the filtered signals on single (Figure 1-b) and multiple (Figure 1-d) emitting MM recordings.

2.2 Rough TDOA (\tilde{T}) estimation

First, T estimates are based on MM click realignment only. Every 10 s, and for each pair of hydrophones (i, j), the difference between times t_i and t_j of the arrival of a click train on hydrophones i and j is referred as $T(i, j) = t_i - t_j$. Its

estimate $\tilde{T}(i, j)$ is calculated by CC of 10-s chunks (overlap of 2s) of the filtered signal for hydrophones i and j [3, 2]. We keep the 35 (NbT) highest peaks on each CC to determine the corresponding $\tilde{T}(i, j)$ (see Figure 1 for detail). The filtered signals give a very fast rough estimate of \tilde{T} (precision ± 2 ms). Figure 1-e shows the CC with the raw signal and Figure 1-g with the filtered signal. The red circles highlight the 35 \tilde{T} . Without filtering, CC generates spurious delay estimates and the tracks are not correct. The raw CC shows more \tilde{T} produced by noise than the filtered CC.

2.3 Echo identification and elimination

Each signal shows echoes for each click (Figure 1-b), maybe due to the reflection of the click train off the ocean surface or bottom or different water layers. Echoes may be responsible

⁶ Hydrophones positions (X(m),Y(m),Z(m)) are: H1=(18501,9494,-1687);H2=(10447,4244,-1677);H3=(14119,3034,-1627);H4=(16179,6294,-1672);H5=(12557,7471,-1670);H6=(17691,1975,-1633);H7=(10658,-14953,-1530);H8=(12788,-11897,-1556);H9=(14318,-16189,-1553);H10=(8672,-18064,-1361);H11=(12007,-19238,-1522)

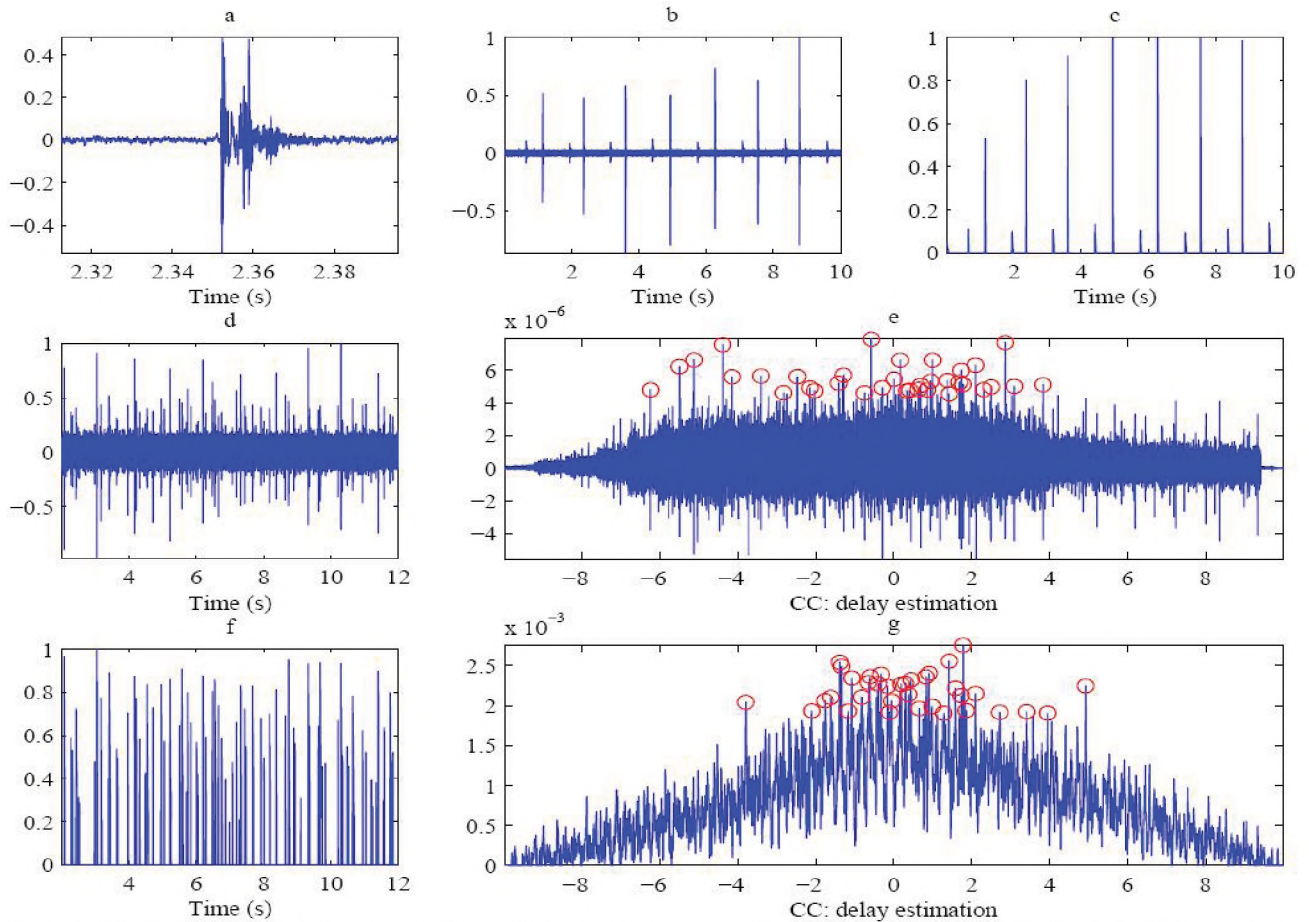


Figure 1: (a): detail of a click on the normalized .wav file format. (b): raw signal (D2) of hydrophone 7 (H7) during the first 10 seconds of recording, containing 7 clicks and their echoes. (c): (b) after filtering. (d): raw signal (D1) of H3 during the first 10 seconds of recording showing multiple emission. (e): CC between (d) and corresponding raw signal chunk of H1. (f): (d) after filtering. (g): CC between (f) and corresponding filtered signal chunk of H1.

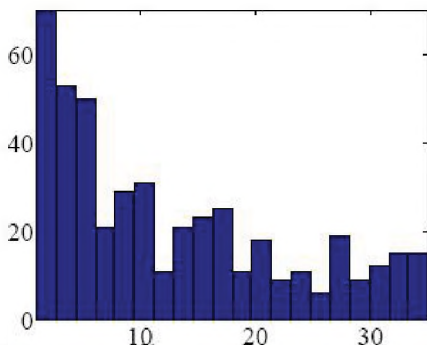


Figure 2: Maximum \tilde{T} CC rank histogram for each triplet

for the detection of additional \tilde{T} in the previous step. We use a method based on autocorrelation [3, 4, 5, 2] to compute echoes $E(i)$ on each 10s chunk and each hydrophone and then eliminate \tilde{T} correspond to a multiple of the echo. For each pair of hydrophones (i, j) , all $\tilde{T}_a(i, j)$ satisfying one of the following equations are removed, $k \in \{1..4\}$, $a \in \{2..NbT\}$:

$$\begin{aligned} \tilde{T}_a(i, j) - \tilde{T}_1(i, j) &= k * E(i) \pm \xi, \\ \tilde{T}_a(i, j) - \tilde{T}_1(i, j) &= -k * E(j) \pm \xi. \end{aligned}$$

where $\xi = 2ms$.

2.4 \tilde{T} transitivity and filtering

Once many \tilde{T} for each pair of hydrophones have been eliminated, the remaining \tilde{T} are combined every 10s to select all quadruplets of hydrophones whose \tilde{T} independent triplet correspond to the same source. Thus we consider that a quadruplet of hydrophones (i, j, k, h) localized the same source with the $\tilde{T}_{a, b, c, d, e, f}$ if the 4 following equations are verified [3, 2] for each time t :

$$\begin{aligned} \tilde{T}_a(i, j) + \tilde{T}_b(j, k) &= \tilde{T}_d(i, k) \pm \delta, \\ \tilde{T}_a(i, j) + \tilde{T}_c(j, h) &= \tilde{T}_f(i, h) \pm \delta, \\ \tilde{T}_d(i, k) + \tilde{T}_e(k, h) &= \tilde{T}_f(i, h) \pm \delta, \\ \tilde{T}_b(j, k) + \tilde{T}_e(k, h) &= \tilde{T}_c(j, h) \pm \delta. \end{aligned}$$

\tilde{T} has been estimated with 2 ms precision, moreover \tilde{T} transitivity only works for an isospeed model which means sound rays move in a straight line. We consider the error $\delta = 6\text{ms}$. The distribution of the maximum \tilde{T} rank for each triplet (Figure 2) in D1, is not negligible near the 35th rank.

2.5 Source localization with a constant profile

Tracks positions are:

$$\{X_t, \forall t\} \text{ with } X_t = (x_t, y_t, z_t)^T.$$

$X_{\{i,j,k,h\}}$ are the known coordinates of hydrophones i, j, k, h .

The three independent \tilde{T} of each hydrophones (i, j, k, h) quadruplet measured on the windows t are noted:

$$\{\tilde{T}_a(i, j, t), \tilde{T}_d(i, k, t), \tilde{T}_f(i, h, t)\}.$$

The modeled delays are:

$$\begin{aligned} T_a(i, j, t) &= \frac{\|X_t, H_i\| - \|X_t, H_j\|}{c}, \\ T_d(i, k, t) &= \frac{\|X_t, H_i\| - \|X_t, H_k\|}{c}, \\ T_f(i, h, t) &= \frac{\|X_t, H_i\| - \|X_t, H_h\|}{c}, \end{aligned} \quad (2)$$

where $\| \cdot \|$ denotes the Euclidian norm. We assume that the precision errors of the T due to the decimation are modelled with a Gaussian, centered, additive, and uncorrelated noise between sensors, noted \mathcal{E} considered the same on each of the windows t and with a variance $\sigma^2 = (\xi/3)^2$ (σ contains 68% of the Gaussian distribution).

$$\begin{aligned} \tilde{T}_a(i, j, t) &= T_a(i, j, X_t) + \mathcal{E}_{i,j,t}, \\ \tilde{T}_d(i, k, t) &= T_d(i, k, X_t) + \mathcal{E}_{i,k,t}, \\ \tilde{T}_f(i, h, t) &= T_f(i, h, X_t) + \mathcal{E}_{i,h,t}, \end{aligned} \quad (3)$$

X_t is estimated with a least square method. The least square criteria to minimize is given by:

$$\begin{aligned} Q(X_t) &= \frac{1}{2} \left[\frac{\tilde{T}_a(i, j, t) - T_a(i, j, X_t)}{\sigma^2} \right]^2 \\ &+ \frac{1}{2} \left[\frac{\tilde{T}_d(i, k, t) - T_d(i, k, X_t)}{\sigma^2} \right]^2 \\ &+ \frac{1}{2} \left[\frac{\tilde{T}_f(i, h, t) - T_f(i, h, X_t)}{\sigma^2} \right]^2. \end{aligned}$$

This case is a non linear criteria minimization. Indeed,

$Q(X_t)$ contains the non linear function $\| \cdot \|$ (Eq.(2)). To solve this problem, the classic recursive minimization method like Gauss-Newton with the Levenberg-Marquardt technique can be applied with an initialization to the middle of the

hydrophones array. X_t estimate is noted \hat{X}_t . After \hat{X}_t estimation, the LMS error is $Q(\hat{X}_t)$. It is adequate when it is inferior to a threshold [14].

2.6 Joint celerity profile optimisation

It is possible, by adding a degree of freedom to Eq.(2), to estimate an optimal celerity profile that will best fit the positions estimates [16]. Five hydrophones are necessary, which is the case in D2, to calculate four independent \tilde{T} . The fourth adds a degree of freedom to the system and permits the estimation of \hat{X}_t ,

$$X_t = (x_t, y_t, z_t, c_t)^T$$

where x_t, y_t, z_t are the source coordinates and c_t the optimal sound speed in windows t . After this \hat{X}_t estimation, we inject the c_t numeric values in the equations system (2) and the system is solved with the least square function

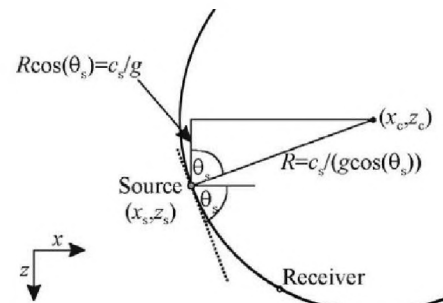


Figure 3: Geometry for a source and receiver in a linear sound speed profile [13]

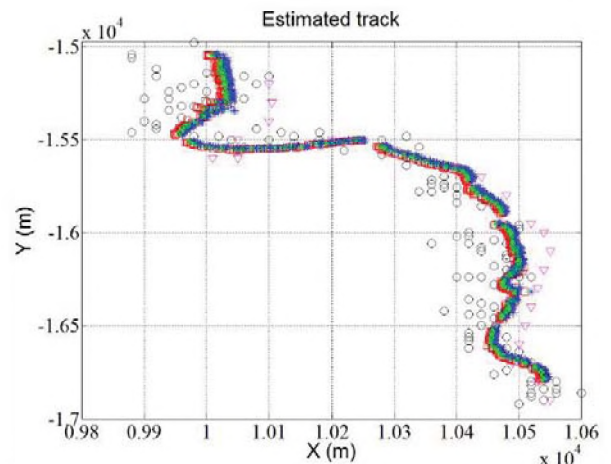


Figure 4: Plan view of the MM in D2, our estimates with a linear (*), a constant profile (□) and an estimated profile (◇), threesome are almost merged; and estimates from Morrissey's [11] (▽) and from Nosal's [12] methods (o). Note the variance of the positions with Nosal's method. The whale direction is opposed to the Y axis. Track and recording duration: 25min. The breaks in the track are due to a temporary cessation of clicking or to engine boat noises.

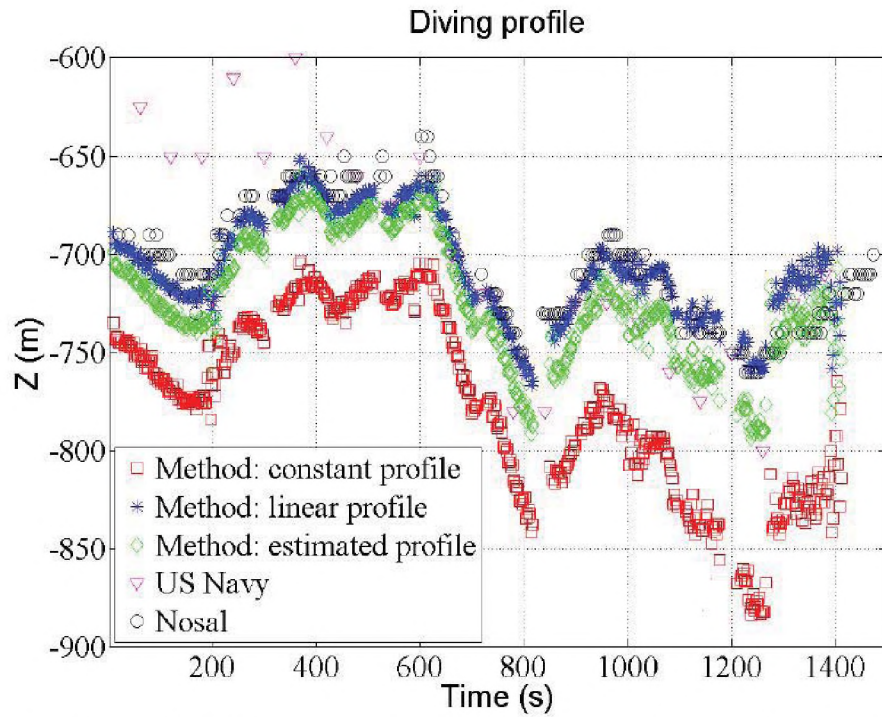


Figure 5: Diving profile of the MM in D2, our estimates with a linear (*), a constant profile (□) and an estimated profile (◇); and estimates from Morrissey's [11] (▽) and from Nosal's [12] methods (○).

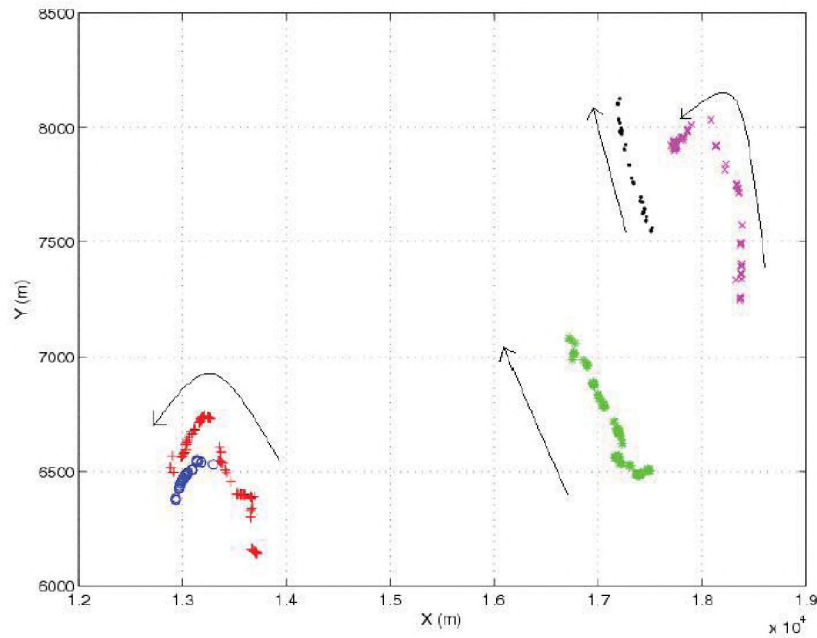


Figure 6: Plan view in D1. Each symbol correspond to one of the five whales. The arrows stress the directions of each whale. See Figure 7 for their diving profile. Whale 1:(o), 2:(+), 3:(*), 4:(-), 5:(x). Recording duration: 20min.

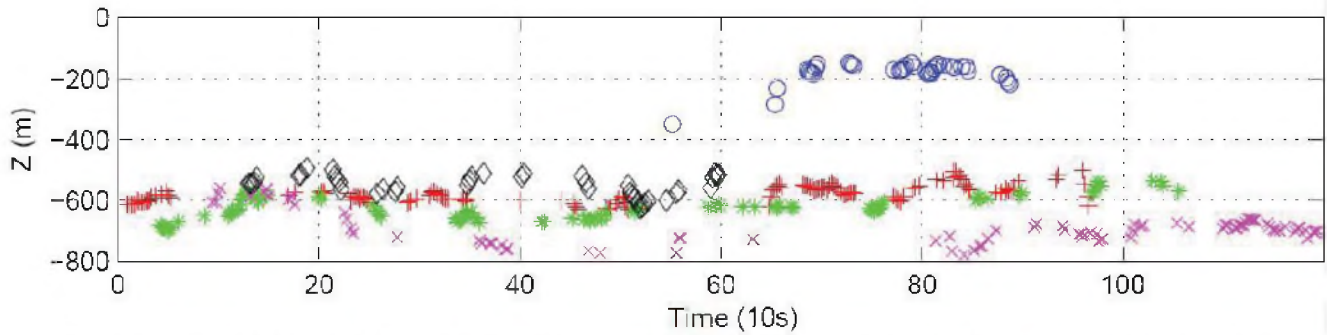


Figure 7: Averaged diving profile in D1. Each symbol correspond to one of the five whales. Whale 1:(o), 2:(+), 3:(*), 4:(-), 5:(x).

5min chunks	0-5	5-10	10-15	15-20
ROSA Lab	4.3	5.3	4	3.6
PIMC	4	4	4	3
Δ	+0.3	+1.3	+0	+0.6

Table 1: Counting number estimations of whales in D1. First row is the five minutes chunks of D1, second is the averaged number of whales estimations from ROSA Lab, third is our estimations (PIMC) and the last row is the difference between PIMC and ROSA Lab estimations.

2.7 Source localization with a linear profile

It is well known that the ray paths in a medium with linear sound speed profile are arcs of circles and further the radius of the circle can be computed [13]. Figure 3 illustrates the appropriate geometry. c_s is the sound speed at the source and θ_s is the launch angle of the ray at the source, measured relative to the horizontal. Note one seeks to determine the launch angle of the ray θ_s which will pass through the receiver located at (x_r, z_r) . From the geometry shown in Figure 3, the center of the circle, (x_s, z_s) , along which the ray path is an arc, can be shown to be:

$$\begin{aligned} x_c &= \frac{x_s + x_r}{2} + \frac{(z_s - z_r)}{2(x_s - x_r)} \left(z_r - z_s + \frac{2c_s}{g} \right), \\ z_c &= z_s - \frac{c_s}{g}. \end{aligned} \quad (4)$$

For a linear sound speed profile, the course time τ , of the ray can be evaluated to yield [13]:

$$\tau = \frac{1}{g} \left\{ \log \left(\frac{z_c - z_s}{z_c - z_r} \right) - \log \left(\frac{R + x_c - x_s}{R + x_c - x_r} \right) \right\}. \quad (5)$$

Using Eqs.(4)-(5) allows one to compute the propagation time from the source to any receiver and hence allows one to compute the predicted delays and then the whale position.

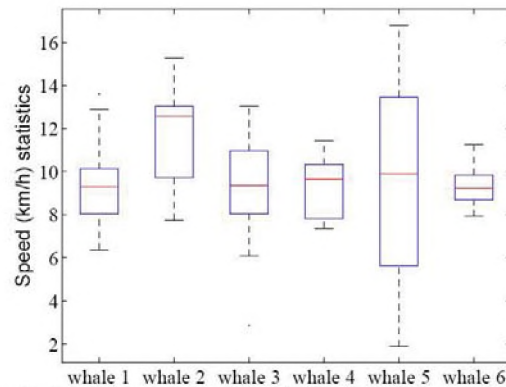


Figure 8: Speed (averaged on 30s windows) statistics on the whole set for each whale in D1 (whales 1 to 5) and D2 (whale 6). The central line of the box is the median of speed and the lower and higher lines are the quartiles. The whiskers show the extent of the speed. Whale 5 seems to stop a moment at the end of the track (See Figure 7).

3 RESULTS

For D2, three sound speed profiles were used: a constant; or an estimated; or a linear. The results are compared with the Morrissey [11] and Nosal [12] methods. In Figure 4, there is one whale, the results with the different methods are similar. In Figure 5, the diving profile underlines a bias of about 50 to 100m between the linear - estimated and the constant profiles results, which emphasizes the importance of the chosen profiles. Moreover with the linear sound speed, the results are about the same as Morrissey's and Nosal's, who used profiles corresponding to the period and place of the recordings. Results for D1 are shown in Figure 6 and 7 for a linear sound speed profile. We thus localize 5 MM. Moreover, according to ROSA Lab estimation based on click clustering (Tab.1), averaged number of MM for each 5min chunks in D1 (A)[6] is similar to ours (B).

3.1 The confidence regions

In section 2.5, because we consider a Gaussian distribution, the standard deviation of the noise is $\xi/3$. Then, we apply a Monte Carlo method. For each T realization, the source

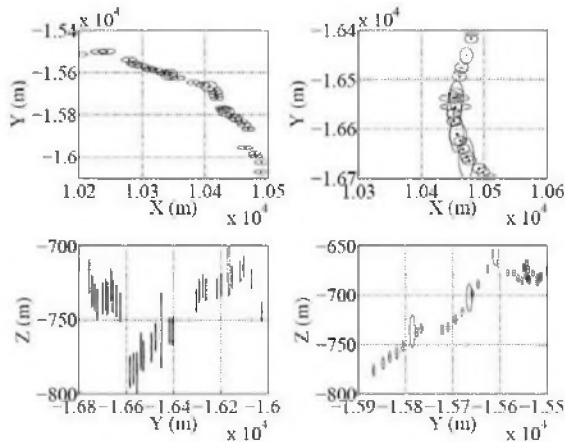


Figure 9: Confidence regions projection on X and Y and on Z and Y axes for D2 trajectory.

position is calculated. We deduce the variance and the mean for each position to plot the confidence regions with a confidence level of 0.95, which means that there is 0.95 probability for the whale to be in the ellipse centered on the position. In D2, the estimated celerity profile described in section 2.6 was used. The mean values of the confidence intervals on X, Y, Z axes are about 18, 16 and 30 m (Figure 9). This justifies the decimation on the raw signal, because the error on X and Y axes are close to the sperm whale length (20m). The results confirm that the errors on the vertical axis are meaningfully higher than the other axes because the distance between each hydrophone in this direction (maximum difference on the Z axis between hydrophones is 200m) is smaller. The D1 results obtained with a linear profile (Figure 6) indicate five trajectories. The farthest whales in D1 from the hydrophones array center have a larger uncertainty with an error of about 20 to 30m on X and Y axes, while the whales close to the center exhibit an error of about 10 to 20m like for D2 (Figure 6). Those uncertainties are reasonable according to the sperm whale length.

4 DISCUSSION AND CONCLUSION

The tracking algorithm presented in this paper is non parametric and real-time on a standard laptop and works for one or multiple emitting sperm whales. The results compared an isovelocity water column and a linear sound speed profile. Depth results with constant speed contains a bias errors due to the refraction of the sound paths from the MM to the receivers what the linear speed profile or the joint celerity optimisation correct. Our algorithm has no species dependency as long as it processes all transients. At this time, only our algorithm gives localization results with typical speed (Figure 8) and depth estimations for multiple emitting whales. In D2, results indicate that only one sperm whale was present in the area, unless other whales in the area were quiet during the selected 25-min period. Moreover, according to ROSA Lab, the estimation number of MM for each 5min chunk in D1 is similar to ours. Our method provides robust online passive acoustics detecting/counting system of clicking MM groups in

open ocean [15]. Further studies will be conducted for click labeling and inter click analyses.

5 ACKNOWLEDGMENTS

We thank AUTECH and NUWC for recording and distributing the dataset. This research was conducted within the international sea *pôle de compétitivité* at Toulon-France, and is a part of the project "Platform of Integration of Multimodal data for Cetology (PIMC)".

REFERENCES

- [1] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81:425–455, 1994.
- [2] P. Giraudet and H. Glotin. Echo-robust and real-time 3d tracking of marine-mammals using their transient calls recorded by hydrophones array. *IEEE ICASSP*, 2006.
- [3] P. Giraudet and H. Glotin. Real-time 3d tracking of whales by echo-robust precise tdoa estimates with a widely-spaced hydrophone array. *Appl acoustics*, 67:1106–1117, 2006.
- [4] H. Glotin. Dominant speaker detection based on harmonicity for adaptive weighting in audio-visual cocktail party asr. *Int. ISCA wksp Adaptive methods in speech recognition, Nice*, Sept 2001.
- [5] H. Glotin, D. Vergyri, C. Neti, G. Potamianos, and J. Luetin. Weighting schemes for audio-visual fusion in speech recognition. *IEEE ICASSP*, 2001.
- [6] X.C Halkias and D. Ellis. Estimating the number of marine mammals using recordings from one microphone. *ICASSP*, 2006.
- [7] C. Ioana and A. Quinquis. On the use of time-frequency warping operators for analysis of marine mammal signals. *IEEE ICASSP*, 2004.
- [8] J.F. Kaiser. On a simple algorithm to calculate the energy of a signal. *IEEE ICASSP*, 1990.
- [9] V. Kandia and Y. Stylianou. Detection of sperm whale clicks based on the teager-kaiser energy operator. *Appl acoustics*, 67:1144–1163, 2006.
- [10] S.G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. volume 11, pages 674–693. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1989.
- [11] R.P. Morrissey, N. DiMarzio J. Ward, S. Jarvisa, , and D.J. Moretti. Passive acoustics detection and localization of sperm whales in the tongue of the ocean. *Appl Acoustics*, 62:1091– 1105, 2006.
- [12] E.M. Nosal and L.N. Frazer. Delays between direct and reflected arrivals used to track a single sperm whale. *Appl Acoustics*, 62:1187–1201, 2006.
- [13] P.R. White, T.G Leighton, D.C Finfer, C. Powles, and O.N Baumann. Localisation of sperm whales using bottommounted sensors. *Appl Acoustics*, 62:1074–1090, 2006.
- [14] F. Caudal, H. Glotin. Multiple real-time 3d tracking of simultaneous clicking whales using hydrophone array and linear sound speed profile. *ICASSP*, 2008.
- [15] F. Caudal, P. Giraudet, H. Glotin, "Automatic Click Labelling using robust 4D tracking estimation of marine mammals : application to Inter-Click Intervals and head foraging estimations ", *Int. wksp on Detection and Classification of Marine Mammals using Passive Acoustics*, BOSTON, July 2007.
- [16] P. Giraudet, H. Glotin, " Joint underwater sound celerity estimation and source localisation using 5 widely-spread hydrophones", *Int. wksp on Detection and Classification of Marine Mammals using Passive Acoustics*, Ed. Nav. Under sea Warfare, Federal RB Massachusetts, BOSTON, July 2007.

INTRODUCTION TO PARTICLE FILTERS FOR TRACKING APPLICATIONS IN THE PASSIVE ACOUSTIC MONITORING OF CETACEANS

P. R. White and M. L. Hadley

Signal Processing and Control Group, Institute of Sound and Vibration Research (ISVR), Highfield, Southampton SO17 1BJ, UK. Emails: (prw,mhl)@isvr.soton.ac.uk.

ABSTRACT

The application of particle filters to two tracking problems in passive acoustic monitoring are discussed. Specifically we describe algorithms for extracting the contours of delphinid whistles and the localization of vocalizing animals in three dimensions using a distributed sensor array. The work is focused on highlighting the potential of particle filters in the analysis of bioacoustic signals. The discussion is based on one particular form of particle filter: the sequential importance resampling filter.

SOMMAIRE

Cette étude porte sur l'application des filtres particuliers à deux problèmes d'extraction d'information en acoustique passive. La description concerne plus spécifiquement deux algorithmes ayant pour objectif l'extraction de contour des sifflements de dauphins et la localisation en trois dimensions d'animaux vocalisant à partir d'un jeu de capteurs répartis localement. L'objectif de ce travail est de mettre en lumière le potentiel des filtres particuliers pour l'analyse de signaux bioacoustiques. Parmi les filtres particuliers, l'accent est mis dans cette étude sur forme particulière de filtre: le filtre à rééchantillonnage d'importance séquentiel.

1. INTRODUCTION

Real-time Passive Acoustic Monitoring (PAM) systems for cetaceans require the integration of many elements. Several of these elements can be cast as tracking problems. In particular this paper considers two such aspects: extracting whistle contours and the estimation of source location using a sensor array. The objective of this work is to highlight the potential of particle filters within the application area as a real-time tracking solution, so the paper is framed in a somewhat pedagogical manner. We avoid details of the theoretical principles underpinning particle filters, rather we aim to convey the fundamental steps common to particle filters.

The definition of a tracking problem is simply a parameter estimation problem in which the parameter estimates are continually updated; such tasks are also formerly referred to as sequential estimation problems. They have been widely studied in a large range of application areas, including sonar, radar and biomedicine. The classical tool for performing sequential estimation is the Kalman filter and its

variants. These methods have been widely, and often successfully, exploited. However their applicability is limited by the underlying assumptions they require.

2. BACKGROUND

The general framework for sequential estimation problems can be expressed as follows. The true value of the parameter vector to be tracked, at time step n , is denoted θ_n . The evolution of this parameter vector is described through a system function, F , such that

$$\theta_n = F(\theta_{n-1}, \mathbf{w}_n) \quad (1)$$

where \mathbf{w}_n is a vector of random variables specifying the random component of the parameter evolution, it is referred to as either the process or the system noise. Similarly the function G defines the measurement process, where \mathbf{x}_n contains the measured data

$$\mathbf{x}_n = G(\theta_n, \mathbf{v}_n) \quad (2)$$

in which \mathbf{v}_n represents the measurement noise process. In the general case the functions F and G are non-linear, the noise processes \mathbf{w}_n and \mathbf{v}_n are not necessarily additive and are not distributed according to a Gaussian distribution. Our goal is to estimate the parameter vector $\boldsymbol{\theta}_n$ on the basis of the set of measurements \mathbf{x}_k , $k=0,1,\dots,n$. In order to avoid increasing memory requirements and computational load as n increases, it is natural to seek a recursive solution. That is to say we seek a solution in which the parameter estimate at time n is derived only from knowledge of the parameter estimate at the preceding time step, $n-1$, and the current measurement \mathbf{x}_n . It should be noted that when $k=0$, \mathbf{x}_0 is the only information available. This is typically provided by a suitable detection algorithm.

2.1 The Kalman Filter

The Kalman filter is a recursive algorithm which is optimal under simplifying assumptions on the system and measurement models (Arulampalam *et al.*, 2002). Specifically (1) and (2) are simplified so that the system and measurement models are linear and the noise processes are additive and Gaussian. Leading to a model of the form

$$\begin{aligned}\boldsymbol{\theta}_n &= \mathbf{A}_n \boldsymbol{\theta}_{n-1} + \mathbf{w}_n \\ \mathbf{x}_n &= \mathbf{B}_n \boldsymbol{\theta}_n + \mathbf{v}_n\end{aligned}\quad (3)$$

in which \mathbf{A}_n and \mathbf{B}_n are the system and measurement matrices, note that whilst in (3) temporal dependence of these matrices has been assumed, in many applications they are constant. The update equations for the classic Kalman filter are (Bozic, 1979; Zarchan & Musoff, 2005)

$$\begin{aligned}\mathbf{T}_n &= \mathbf{A}_n \mathbf{P}_{n-1} \mathbf{A}_n^t + \mathbf{Q}_n \\ \mathbf{K}_n &= \mathbf{T}_n \mathbf{B}_n^t (\mathbf{B}_n \mathbf{T}_n \mathbf{B}_n^t + \mathbf{R}_n)^{-1} \\ \mathbf{P}_n &= \mathbf{T}_n - \mathbf{K}_n \mathbf{B}_n \mathbf{T}_n \\ \hat{\boldsymbol{\theta}}_n &= \mathbf{A}_n \hat{\boldsymbol{\theta}}_{n-1} + \mathbf{K}_n (\mathbf{x}_n - \mathbf{B}_n \mathbf{A}_n \hat{\boldsymbol{\theta}}_{n-1})\end{aligned}\quad (4)$$

in which \mathbf{T}_n is a temporary matrix (but can be regarded as an a priori estimate of \mathbf{P}_n) used to ease the computational load, \mathbf{Q}_n and \mathbf{R}_n are the covariance matrices for the process and measurement noises respectively, \mathbf{K}_n is the Kalman gain matrix, \mathbf{P}_n is the error covariance matrix and $\hat{\boldsymbol{\theta}}_n$ is the vector of parameter estimates at time n .

The Kalman filter is a highly flexible and computationally efficient scheme. But its application is limited to cases where (3) can be regarded as suitable approximation of (1) and (2). Variants on the Kalman filter have been proposed which extend its range of applicability, common examples of these are the extended Kalman filter (EKF) (Zarchan &

Musoff, 2005) and unscented Kalman filter (UKF) (Wan & van der Merwe, 2000).

2.2 The Particle Filter

Particle filters provide a general solution to tracking problems of the form described by (1) and (2), without the need to invoke the inherent assumptions associated with the Kalman filter. There are a wide variety of versions of particle filters that have been defined (Arulampalam *et al.*, 2002; Ristic *et al.*, 2004; Doucet *et al.*, 2001). However the objective of this work is to communicate the opportunities afforded by the use of particle filters in PAM systems, rather than a review of particle filters per se. So we shall concentrate on a simple form of particle filter, specifically we shall discuss Sequential Importance Resampling (SIR) filters. These do not represent the state-of-the-art particle filtering algorithms, but they do provide a good basis for the introduction of the concepts of particle filtering and offer good performance in the examples presented herein.

Consistent with the review character of this publication we provide a mechanistic description of the SIR filter and choose to omit the under-pinning principles, these principles are widely available elsewhere, e.g. (Arulampalam *et al.*, 2002; Ristic *et al.*, 2004; Doucet *et al.*, 2001). The objective here is to provide some insight into how to construct a particle filter and to highlight the flexibility and power that they provide.

Particle filters are also referred to as sequential Monte Carlo algorithms (Doucet *et al.*, 2001) and, as is characteristic of Monte Carlo schemes, they exploit samples drawn from the underlying distributions. Given a set of M parameter estimates at time $n-1$ which are denoted $\hat{\boldsymbol{\Theta}}_{n-1} = \{\hat{\boldsymbol{\theta}}_{n-1,k}\}_{k=1,\dots,M}$, the basic steps involved in the SIR particle filter are:

- i. Update each of the estimates using $\tilde{\boldsymbol{\theta}}_{n,k} = F(\hat{\boldsymbol{\theta}}_{n-1,k}, \mathbf{w}_{n,k})$ where $\mathbf{w}_{n,k}$ is a sample from the process noise distribution.
- ii. Use the measured data to score each of the new estimates $\tilde{\boldsymbol{\theta}}_{n,k}$ using the likelihood computed via (2) and normalize these scores so that they sum to unity.
- iii. Create $\hat{\boldsymbol{\Theta}}_n$ by drawing M samples, with replacement, from $\tilde{\boldsymbol{\Theta}}_n$ according to the scores allocated in step ii.
- iv. From the samples $\tilde{\boldsymbol{\Theta}}_n$ form an estimate of the

parameter vector.

In the first step the existing estimates are perturbed, in a manner which mimics the effect of process noise, so producing a set of candidate parameter estimates. The particle filter then considers these estimates and scores them according to how well they predict the sample that has just been measured, \mathbf{x}_n . This may be explained in the specific case of an additive noise measurement noise model, i.e. in the case where (2) can be expressed in the form:

$$\mathbf{x}_n = G(\boldsymbol{\theta}_n) + \mathbf{v}_n \quad (5)$$

In such cases the scoring is realized by evaluating the likelihood $p_v(\mathbf{x}_n - G(\tilde{\boldsymbol{\theta}}_{n,k}))$ in which p_v is the probability density function of the measurement noise process \mathbf{v} . Consequently parameter estimates close to the true value should yield values of $G(\tilde{\boldsymbol{\theta}}_{n,k})$ which are close to the measured data, so they have relatively large likelihood. Whereas estimates significantly different from the true value, will (probably) yield measurement estimates very different from the measured value, so yield a low likelihood. The scores are derived from the likelihood by scaling them so that they sum to unity.

Step iii, is realized by selecting the estimates using random sampling according to the estimate's scores. Uniform random variables are used and the probability of selecting a particular estimate is given by its score. The sampling is implemented with replacement, so that estimates with high scores are typically replicated many times. The new samples constitute the set of parameter estimates for starting the next iteration.

The final step is to construct the final parameter estimate. This can be done using one of several principles including: MAP (maximum a priori probability) and minimum mean squared error (MMSE).

3. WHISTLE CONTOUR EXTRACTION

One way in which species classification for delphinids can be achieved is through analysis of their whistles (Oswald *et al.*, 2007). Specifically the contours of the whistles in the time-frequency domain are used as the key features and these contours need to be estimated (extracted) in order to successfully realize such a system. The extraction of such whistles is normally achieved through use of the spectrogram (Oswald *et al.*, 2007; Datta & Sturtivant, 2002; Lepretre & Martin, 2002) although alternative approaches can prove successful (Johansson & White, 2004). The extraction process can be hindered by the presence of overlapping whistles and echolocation clicks

as well as potentially low Signal to Noise Ratios (SNRs).

In this work we demonstrate how particle filters can be used as one way to automate this contour extraction process. Other workers have considered applying particle filters to similar problems based on the spectrogram (Dubois *et al.*, 2005; Nagappa & Hopgood, 2006). The work described here takes advantage of the Short Time Fractional Fourier Transform (STFrFT). The STFrFT for the analysis of whistle has been considered elsewhere (Capus & Brown, 2003). It is worth noting that the method we adopt we refer to as a STFrFT, largely in deference to previous work in this application, but it should be noted that the method could be regarded under a number of other signal processing paradigms, most obviously it also exploits the principles behind adaptive basis decomposition methods (Mallat & Zhang, 1993).

Our tracking scheme is based on detecting the maxima of the STFrFT. Consider the k^{th} windowed data segment of the incoming data stream, $x(n)$, denoted \mathbf{x}_k and defined as

$$\mathbf{x}_k = [x(kP), x(kP+1), \dots, x(kP+L-1)]^T \quad (6)$$

where P is the number of samples by which the window is shifted between successive analysis windows and L is the window length. The elements of \mathbf{x}_k are denoted $x_k(m)$, $m=0, \dots, L-1$. The STFrFT is defined as

$$S(k, \alpha, f) = \left| \sum_{m=0}^{L-1} x_k(m) e^{-2\pi i(f + \alpha t_m/2)t_m} \right|^2 \quad (7)$$

where f denotes centre frequency [Hz], α is the frequency sweep rate [Hz/s] and the local time index, t_m , is defined through $t_m = \left(m - \frac{L}{2}\right) / f_s$, in which f_s denotes the sampling frequency. The STFrFT can be loosely regarded as representing the energy in a signal at a particular time and at a frequency associated with a particular sweep rate.

Evidently the classical short time Fourier transform (the spectrogram) is a special case of (7) in which $\alpha \equiv 0$. The flexibility offered by the STFrFT allows the processing scheme to more accurately model the underlying process. By accommodating linear sweeps the STFrFT can increase the SNR of received signal, assuming that a sweep rate, α , is chosen that is close to that in the received data. The rapid sweep rates that can be observed in odontocete whistles make the use of the STFrFT an attractive option. The use of the STFrFT intrinsically provides estimates of the sweep rate for each analysis window; this additional information can be used to improve tracking performance.

3.1 Particle Filter for Whistle Contour Extraction

The use of particle filters to extract whistle contours requires one to define the system and measurement functions, i.e. (1) and (2). The parameter vector we seek to estimate contains both the frequency and sweep rate and is defined as $\theta = [f, \alpha]^t$. The system model we employ is:

$$\theta_n = \begin{bmatrix} 1 & \frac{P}{f_s} \\ 0 & 1 \end{bmatrix} \theta_{n-1} + \mathbf{w}_n = \mathbf{A}\theta_{n-1} + \mathbf{w}_n \quad (8)$$

\mathbf{w}_n is a zero mean Gaussian noise with a diagonal covariance matrix, so that system noise on the frequency and sweep rate are uncorrelated with difference variances. This is a standard linear model of the form of (3).

There are several candidate measurements one can use for this system. The one adopted herein, based on the STFrFT, is

$$X_n = \left| \sum_{m=0}^{L-1} x_k(m) e^{-2\pi i(\theta_n(1) + \theta_n(2)t_m/2)t_n} \right|^2 \quad (9)$$

Note the distinction between the data, $x_k(m)$, and the measurement associated with a parameter X_n . The measurement X_n is a scalar value.

The processing scheme adopted here applies a robust pre-whitening step (Leung & White, 1998) to the incoming data stream, to create $x(n)$; this ensures that the background noise has an approximately flat spectrum of a known level. This pre-whitening allows one to use the value X_n as a proxy for the (unscaled) probability $p(\theta_n | x_k)$: large values of X_n relate to highly probable events, whereas small values of X_n relate to events of low probability. This argument is a simplification of the principles lucidly described in detail in (Brethorst, 1988). The non-linear character of (9) favors the use of a particle filter solution.

3.2 Results for Whistle Contour Extraction

This method is applied to a short (1.9 s) section of a whistle recorded from *Tursiops truncatus*. This recording contains features common to many similar recordings. There are trains of echolocation clicks, e.g. between 0.2 and 0.4 s, there are other whistles of varying strength and the primary whistle varies rapidly in frequency and level throughout the recording. The results are shown in Figure 1. The upper frame of this figure shows the original spectrogram, whilst the lower frame depicts the same spectrogram overlaid with a white line showing the

estimated frequency contour.

The algorithm has successfully tracked the whistle. There is an initial period, before 0.1 s, where the algorithm provides an estimate which deviates somewhat from the visual track of the whistle. The signal is weak here, but the primary cause for this behavior is the fact that the algorithms require some time to initialize, to “burn in”. The estimated track also deviates at around 0.6 s when the whistle’s amplitude temporarily reduces significantly. Accepting these minor deviations it is encouraging to note that the algorithm has successfully tracked the whistle even during the rapid frequency jump occurring shortly after 0.2 s and ending shortly before 0.4 s. This is despite the signal being partly obscured by a click train. This is particularly gratifying since such jumps are characteristic of *T. truncatus* and one would seek to avoid classifying such a jump as two separate whistles.

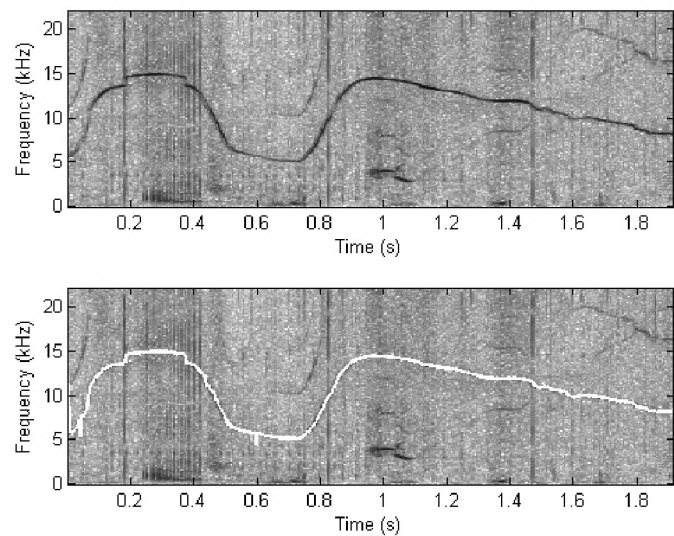


Figure 1: Results of contour extraction for a *Tursiops truncatus* whistle based on a particle filter. Upper frame shows the spectrogram, the lower frame shows the same plot with a white line overlaid to show the contour estimate.

4. SOURCE LOCALISATION

The second problem this paper considers is that of tracking the location of a vocalizing animal in three dimensions using a hydrophone array. The most suitable signals for performing such localizations are echolocation clicks, but other vocalizations can be used. In this example we use echolocation clicks from a sperm whale, *Physeter macrocephalus*.

The localization problem can be described as follows. A set of acoustic sensors (in this case hydrophones) are located at

known positions, \mathbf{r}_m , in an environment. The source, is at an unknown location \mathbf{s} , and emits a sound (a vocalization) which propagates through a known medium. The phrase “known medium” is intended to highlight the assumption that the propagation time from one point in the medium to a second point can be computed, implying knowledge of (at least) the sound speed profile. This model we shall denote M . Using the received signals one is able to compute the delays, $\tau(m,n)$, observed between the vocalizations being detected on the pair of hydrophones m and n . Collecting all of these delays into a vector $\boldsymbol{\tau}$ allows one to express the problem as: given the measured delays $\boldsymbol{\tau}$, a model of the acoustic environment, M , and the sensor locations, \mathbf{r} , can one infer the source location?

The choice of the acoustic model M is an important factor controlling the accuracy of the resulting estimated source locations. The use of a simple model should result in an efficient algorithm, but the estimates may be subject to considerable error. The use of models that accurately capture the propagation of sound in the ocean is clearly advisable, but their use is often limited by the absence of complete knowledge of the physical parameters required to specify such a model. The form of the model used does not directly impact the following discourse. This problem has been treated by a large number of authors as a non-sequential estimation problem, e.g. (Spiesberger, 2001; Thode, 2004; White *et al.*, 2006). In the following we translate the problem into a tracking, sequential estimation, task and present a particle filter based solution.

4.1 Particle Filter for Source Localization

The underlying model when using a particle filter for localization is relatively straightforward. The unknown parameter vector, $\boldsymbol{\theta}$, contains the source co-ordinates, for example expressed in Cartesian co-ordinates. The system matrix aims to model how the animal moves through the medium. Various methods can be used to impose models that are appropriate for the known behavioral parameters. For example one can seek to impose maximum swim rates, or rates of ascent and descent. With the objective of retaining simplicity we stick with a simple random walk model:

$$\boldsymbol{\theta}_p = \boldsymbol{\theta}_{p-1} + \mathbf{w}_p \quad (10)$$

where \mathbf{w}_p represents a vector of independent, zero mean, Gaussian white noise. The underlying assumptions in (10) are very limited, it assumes that the current location is the just a random perturbation from the preceding location; Further note that the vocalizations typically occur at irregular intervals. The subscript p denotes data associated with the p^{th} such vocalization. A shortcoming of the

random walk model described by (10) is that it does not account for this irregular sampling. It is reasonable to increase the standard deviation of \mathbf{w}_p in proportion to the interval between irregular vocalizations, reflecting the fact that an animal is likely to have moved a greater distance in longer intervals than short ones.

The measurement model is simply:

$$\boldsymbol{\tau}_p = M(\boldsymbol{\theta}_p, \mathbf{r}) + \mathbf{v}_p \quad (11)$$

The function M is typically a highly non-linear function. It is this that again makes the particle filter an attractive processing option. The measurement noise \mathbf{v}_p is modeled using a long-tailed distribution, such as a Laplacian distribution. The advantage of this is that it models the occasional failure of the delay estimation operation. The presence of strong reflectors can lead to some isolated delay estimates with large errors. By employing a measurement noise model with a long-tailed distribution such outliers are penalized less than would be the case if a Gaussian model was used for the noise.

4.2 Results for Source Localization

The algorithm outline in the preceding subsection has been applied to data obtained from an echo-locating sperm whale using bottom mounted hydrophones. Specifically, the data used in this study was the second data set supplied for the 2005 Workshop on Detection and Localization of Marine Mammals using Passive Acoustics held in Monaco (Adam, 2006).

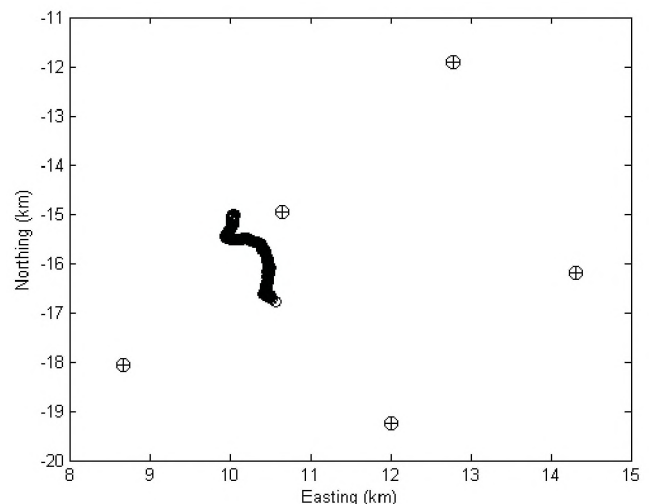


Figure 2: Estimated source locations using particle filter. Open circles ‘O’ indicate the individual estimates. Crossed circles ‘⊕’ indicate the sensor locations

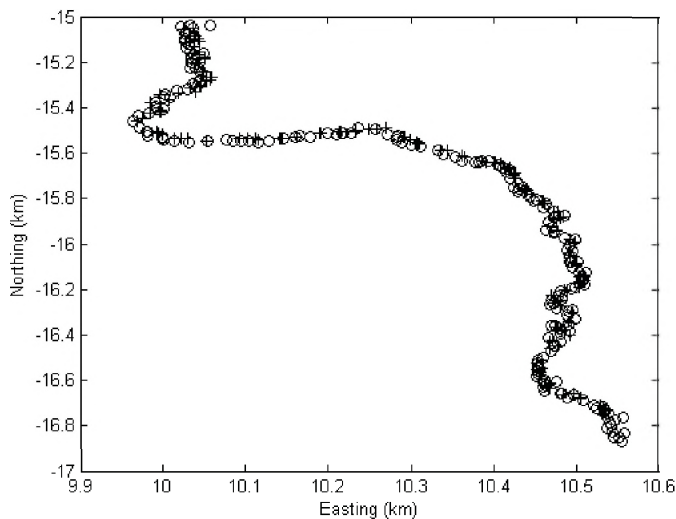


Figure 3: Comparison of estimated source locations using particle filter and non-sequential estimation.

Open circles 'O' indicate the results of particle filtering. Crosses '+' indicate estimates from non-sequential estimator

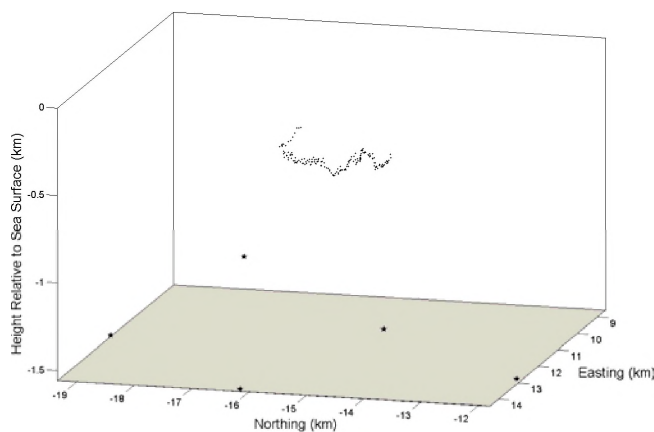


Figure 4: Estimated source location in three dimensions. Sensor location locations are indicated by the black stars (★)

The results depicted here are based on the same delay estimates as those employed in (White *et al.*, 2006). The results are obtained assuming a linear sound speed profile. Figure 2 depicts the result of applying the particle filter to this data set. Whilst Figure 3 shows the same results on an expanded scale, also shown in this plot are the results of the algorithm presented in (White *et al.*, 2006) obtained on the same data set using a propagation model with a constant sound speed. Comparing the results from the particle filter with those from the estimation method, in (White *et al.*, 2006), demonstrates the potential of the particle filter as a real-time tracking solution of comparable performance given the same time delay estimates.

The results shown in Figures 2 and 3 are consistent with those obtained in (Adam, 2006) and from Figure 3 we see that the particle filter results and those from the non-sequential scheme are very close to each other. In Figure 4 these results are shown in three dimensions.

5. DISCUSSION

Particle filters provide a general, powerful and flexible tool for solving tracking problems. The results herein demonstrate that the solutions achieved are of high quality, despite the rather crude nature of the particle filter algorithms used and the fact that we have, in general, avoided including all of the available prior information in the interests of retaining simplicity.

The good performance of particle filters is commonly realized at the cost of a large computational burden being incurred. A key parameter in controlling this cost is the number of particles M employed. The larger the number of particles, the better the solution but the greater the computational burden imposed. A second key parameter that affects performance, but is not related to computational burden, is suitability of the choice of the distribution width. This should be chosen to be representative of the change expected to occur in the state vector between measurements. In the source localization application this would be relative to the typical swim speed of the species of animal to be tracked.

The initial estimate of the state vector is derived depending on the application. For the whistle contour extraction once the presence of a whistle is detected each STFrFT bin is weighted according to a uniform distribution. Here this is possible because the states are discrete and therefore the number is relatively limited. In the localization application the number of possible initial states is much greater, therefore a cost function minimization estimation scheme was utilized to provide the first estimate.

The implementations employed herein both used 1000 particles. In the case of the localization the computational cost that implied was consistent with a real-time implementation, even with the algorithm implemented in MATLAB[®]. This is in part because the filter only needs updating approximately once per second. The computational load would also escalate significantly if a more detailed propagation model is used.

The real-time implementation of the contour extraction algorithm requires modification of the algorithm presented here as in its current form it is probably too computationally demanding for simple real-time implementation. One could exploit the potential for parallel implementation inherent in particle filters, but this dramatically increases the issues

associated with implementation.

REFERENCES

1. Adam, O. (2006) Special issue: Detection and localization of marine mammals using passive acoustics. *Applied Acoustics*, **67**, 1057-1058.
2. Arulampalam, M.S., Maskell, S., Gordon, N. & Clapp, T. (2002) A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, **50**, 174-188.
3. Bozic, S.M. (1979) *Digital and Kalman Filtering*. Edward Arnold, London.
4. Brethorst, G.L. (1988) *Bayesian spectrum analysis and parameter estimation*. Springer-Verlag, Berlin.
5. Capus, C. & Brown, K. (2003) Short-time fractional Fourier methods for the time-frequency representation of chirp signals. *Journal of the Acoustical Society of America*, **113**, 3253-3263.
6. Datta, S. & Sturtivant, C. (2002) Dolphin whistle classification for determining group identities. *Signal Processing*, **82**, 251-258.
7. Doucet, A., de Freitas, N. & Gordon, N. (2001) *Sequential Monte Carlo Methods in Practice*. Springer Verlag.
8. Dubois, C., Davy, M. & Idier, J. (2005) Tracking of time-frequency components using particle filtering. In: *International Conference on Speech and Signal Processing 2005*, p. IV-9-IV-12.
9. Johansson, A.T. & White, P.R. (2004) Detection and characterization of marine mammal calls by parametric modelling. *Canadian Acoustics*, **32**, 83-92.
10. Lepretre, B. & Martin, N. (2002) Extraction of pertinent subsets from time-frequency representations for detection and recognition purposes. *Signal Processing*, **82**, 229-238.
11. Leung, T.S. & White, P.R. (1998) Robust estimation of oceanic background spectrum. In: *Mathematics in signal processing IV*, McWhirly J.G. & Proudler I.K. (eds.), pp. 369-382. Clarendon Press, Oxford.
12. Mallat, S.G. & Zhang, Z.F. (1993) Matching Pursuits with Time-Frequency Dictionaries. *Ieee Transactions on Signal Processing*, **41**, 3397-3415.
13. Nagappa, S. & Hopgood, J.R. (2006) Frequency Tracking of Biological Waveforms. In: *IMA Seventh International Conference On Mathematics In Signal Processing*, p. 12.
14. Oswald, J.N., Rankin, S., Barlow, J. & Lammers, M.O. (2007) A tool for real-time acoustic species identification of delphinid whistles. *Journal of the Acoustical Society of America*, **122**, 587-595.
15. Ristic, B., Arulampalam, M.S. & Gordon, N. (2004) *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House.
16. Spiesberger, J.L. (2001) Hyperbolic location errors due to insufficient numbers of receivers. *Journal of the Acoustical Society of America*, **109**, 3076-3079.
17. Thode, A. (2004) Tracking sperm whale (*Physeter macrocephalus*) dive profiles using a towed passive acoustic array. *Journal of the Acoustical Society of America*, **116**, 245-253.
18. Wan, E.A. & van der Merwe, R. (2000) The unscented Kalman filter for nonlinear estimation. In: *IEEE Symposium on Adaptive Systems for Signal Processing, Communications and Control*, pp. 153-158.
19. White, P.R., Leighton, T.G., Finfer, D.C., Powles, C. & Baumann, O.N. (2006) Localisation of sperm whales using bottom-mounted sensors. *Applied Acoustics*, **67**, 1074-1090.
20. Zarchan, P. & Musoff, H. (2005) *Fundamentals of Kalman filtering: a practical approach*, 2nd edn. AIAA.

ROBUST 2D LOCALIZATION OF LOW-FREQUENCY CALLS IN SHALLOW WATERS USING MODAL PROPAGATION MODELLING

C. Gervaise¹, S. Vallez¹, Y. Stephan², Y. Simard^{3&4}

1 : E3I2, EA3876, ENSIETA, GIS Europôle Mer, 2 Rue François Verny, 29200 Brest, France

2 : CMO/SHOM, 13 Rue Chatellier, 29200, Brest, France

3 : Institut des sciences de la mer, université du Québec à Rimouski, 310 allée des Ursulines, Rimouski, Québec G5L 3A1, Canada

4 : Marine Mammal section, Fisheries and Oceans Canada, Maurice Lamontagne Institute, P. O. Box 1000, Mont-Joli, Québec GH 3Z4, Canada

ABSTRACT

We propose a new method to localize low-frequency calls in 2D in shallow waters from a sparse array of hydrophones using modal propagation modelling. An analysis of modal propagation modelling of transients signals in shallow water environment shows that the dispersive behaviour of the waveguide can be exploited to design a robust localization scheme without requiring any knowledge of the acoustics properties of the environment (bottom and water column) nor any simulation of propagation. The localization scheme also does not require synchronization of the array and is therefore independent of any clock drift. Promising results are obtained for Northern right whale gunshot calls from 'Bay of Fundy data set of the 2003 Workshop on Detection and Localization of Marine Mammals Using Passive Acoustics.'

RÉSUMÉ

Dans ce papier, un algorithme robuste de localisation 2D à partir des émissions transitoires dans des milieux petits fonds est proposé. Il s'appuie sur un modèle de propagation modale. Une analyse des phénomènes de dispersion induits par la propagation montre qu'il est possible, à partir d'un réseau lâche d'hydrophones, de proposer une méthode de localisation ne nécessitant ni la connaissance du milieu, ni l'exécution d'un code de propagation. L'algorithme de localisation ne nécessite pas la synchronisation du réseau et est par conséquent indépendant des dérives d'horloges. Des résultats encourageants sont obtenus pour localiser les émissions « gunshot » des baleines franches à partir du jeu de données de la Baie de Fundy, de l'Atelier de 2003 sur la détection, la localisation et la classification de mammifères marins par acoustique passive'.

1. INTRODUCTION

Localizing marine mammals in large ocean basins is needed to assess their use of the habitat in time and space and study the impact of global changes on ecosystems [Tho86] [Win04] [Sta07]. Localization may become crucial for some endangered species in relation with anthropogenic activities such as airgun seismic surveys, low-frequency military applications and collisions with ships [And01]. Even if visual observations from ships and planes may be used during daytime, passive acoustics localization methods can increase the spatial extent of localization, besides of being still active during night and bad weather conditions [Spi90] [Tho86]. Passive acoustics monitoring (PAM) appears to be suitable for integrated autonomous, real-time and long term alert systems to prevent collisions with ships [Sim06] if the animals

produce sounds regularly enough and over a range of behaviours. After being emitted, marine mammal calls propagate along paths from the animal's position to one or several hydrophones. Then features such as time difference of times of arrival (TDoA) at each hydrophone [Lau03] [Spie90] or time-frequency dispersive pattern [Win04], are extracted from the measurements using signal processing techniques and used to estimate the source location. Passive acoustic localization techniques require a model of acoustic propagation in the environment, the knowledge of ocean acoustics properties at the emission time and a localization algorithm. Accuracy and robustness of the estimates depend on the emitted signal (bandwidth and level), the noise level, the adequacy between the propagation model and the reality [Cha04] and on monitoring of ocean acoustic properties. Existing methods range from direct-ray path propagation

assumption associated with hyperbolic fixing [Lau04] [Des04] to more elaborated modal propagation and localization processing [Ebb06] [Win04]. Figures of merit of a localization scheme include localization's accuracy, robustness to weak knowledge about environment properties and real-time implementation capability for alert systems.

The productive shallow waters of continental shelves are intensively used by low-frequency calling baleen whales (< 1 kHz) [Ric95]. Also, most of the documented collisions appeared to be there or near the continental shelf [Lai01]. Considering these facts, normal mode modelling seems to be an adequate model to deal with acoustic propagation of these whale calls [Jen00] [Win04]. Real data application of our contribution focuses on the localization of North Atlantic right whales gunshot calls in the Bay of Fundy, Canada. Nowadays, North Atlantic right whales (*Eubalaena glacialis*) population is less than 350 individuals and is in decline due to high human induced mortality [Van03]. Indeed, ship strikes accounted for 35.5% (16/45) of the documented North Atlantic right whale mortality between 1970 and 1999 [Kno01]. North Atlantic right whales sounds have been recently described [Mat01] [Van03] [Par05]. In general they are low-frequency sounds (< 1 kHz) with various waveforms (constant low-frequency, moan, upsweeping and downsweeping modulations and gunshot). Gunshot calls are a loud impulsive sounds (duration ~30 ms, bandwidth ~[10Hz,20kHz], [Par05]) frequently used by right whales in the Bay of Fundy, Canada [Van03]. They are produced by lone males or males in a social active group at or near the surface and seem to have some implication in reproductive display. The Canadian right whale Conservation Area in the Bay of Fundy is close to an internationally designated shipping lane used by numerous large carriers [Lau03] which was recently changed to minimise collision risk. Efficient localization of gunshot calls through PAM systems can help improving right whale conservation. Moreover, internal waves taking place in the Bay of Fundy produce large and rapid variations of sound speed profiles [Des04_a] [Cla06] that must be taken into account by appropriate localization algorithms.

In the present paper, we propose a method to localize (in a 2D horizontal plane) low-frequency transient signals in shallow water environments. Our scheme relies on a normal mode propagation modelling and a targeted area of emissions surrounded by a sparse network of hydrophones. By exploiting the dispersive behaviour of the acoustic channel and time-frequency signal processing, our method allows localizing the source without any knowledge of the ocean acoustics properties of the channel and without any requirement to run simulations from acoustic propagation models. This method can use exactly the same recording device as that used for TDOA localisation schemes and can advantageously replace them when shallow water and

very low-frequency sounds are encountered. Our method is tested on three recordings from the dataset provided in support of the 2003 Workshop on Detection and Localization of Marine Mammals Using Passive Acoustics. Satisfactory results are obtained and the present paper aims at presenting this new potential localization scheme to the community.

The first part of the paper briefly presents the experimental material used in the field, the second part recalls the main features of normal mode propagation and the third part describes our localization scheme. Then a fourth part applies the method to real data, including comparisons with other classical methods. The last part discusses the results.

2. ACOUSTIC DATA SET

The data set that we used in this paper is the one provided in support of the 2003 Workshop on Detection and Localization of Marine Mammals Using Passive Acoustics. This dataset contains North Atlantic right whale sounds recorded in the Bay of Fundy during 2000 and 2002 [Des04_a]. None of these calls have an *in situ* visual ground truth. Among 16 recordings (9-10 September 2002), the dataset provides 5 30-s recordings containing gunshot calls. Recordings were performed by five OBH autonomous hydrophones moored on the bottom. A single hydrophone was located in each corner of a 14-km square, with the fifth located in the middle (c.f. table 1).

OBH	Deployment position		Water depth (m)
	Latitude (N)	Longitude (W)	
C	44.60073	66.49723	210
E	44.60237	66.31591	134
L	44.66203	66.40453	183
H	44.73051	66.31556	123
J	44.73038	66.49619	170

Table 1 : Dataset OBH positions

The OBH network was deployed in shallow waters with bathymetry varying from 100 to 200 meters. Sound speed profiles during the experiment were downward refractive or had a local minimum. They showed notable short-term variations. The sub-bottom structure in the area is mainly composed of a first *Lahave* clay layer over a thick layer of *Scotian drift* [Map77]. The weak compression sound speed in *Lahave* clay, which was smaller than the sound speed in water, implies a high level of dispersion for normal mode propagation. The OBH recordings were digitized using a 12-bit A/D converter with a sampling frequency of 1200 Hz. Figure 1 presents the recordings S035-2 at hydrophones H, L, E, C with a spectrogram representation in 10 to 100 Hz bandwidth (Kaiser window, $\beta=0.1102(180-8.7)$, length : 512 samples). One can clearly see:

- on the time-amplitude plots (panels C, D, E, F), the gunshot time of arrival on each OBH,
- on the time-frequency plots (panels G, H, I, J), a typical pattern of dispersive normal mode propagation. Each received gunshot has a multi-component structure and each component has its own time of arrival which depends on frequency (as the time delay between each echo increases with the range between the gunshot and the hydrophones, the time frequency structure of the arrival can be attributed without ambiguity to propagation and not to the gunshot itself).

3. DIRECT MODELLING: NORMAL MODE PROPAGATION

Considering that the energy in the vocalization is concentrated in the low-frequency band and propagating in shallow water waveguide, the normal mode propagation theory seems appropriate for the analysis. In a range independent environment, the transfer function between a receiver and an emitter r meters apart can be written as given in equation 1 (c.f Section 8) [Jen00] where $g_m(z)$ represents the modal function of index m , $k_r(m,f)$ the radial wave number of index m and frequency f , Rk_r the real part of $k_r(m,f)$ and Ik_r the imaginary part of $k_r(m,f)$, z_s is the source depth and z_r the receiver depth. The term $A(m,f,r,z_s,z_r)$ includes the attenuation between source and receiver; the term $P(m,f,r)$ includes the propagation time and propagation speed between source and receiver. From $P(m,f,r)$, phase speed v_ϕ and group speed (propagation speed of the energy) v_g of mode m at frequency f are defined by equation 2:

$$v_\phi(m,f) = \frac{2\pi f}{Rk_r} \quad v_g(m,f) = 2\pi \frac{\partial f}{\partial Rk_r} \quad Eq 2.$$

In shallow water environments, v_ϕ and v_g depend on both the frequency and the index m . That's why, different modes at the same frequency propagate with different speeds and one mode at different frequencies propagates with different speeds. So, if a source emits an impulse signal, the received signal after propagation in the channel contains several echoes, and for each echo, its frequencies arrive at different times, in that sense, normal mode propagation is said to be dispersive.

If one source emits a transient signal with a time-frequency modulation $t_e(f)$ (where t_e is the emission date of frequency f), theoretical received time-frequency structure RTF_s after a normal mode propagation of range r between source and receiver is given by equation 3:

$$RTF_s(t,f) = \sum_{m=1}^{+\infty} A(m,f,r,z_s,z_r) \delta\left(t - t_e(f) - \frac{r}{v_g(m,f)}\right) \quad Eq 3$$

where $\delta(t)$ stands for the impulse distribution.

To illustrate the dispersive propagation and previous formula, simulation is carried out. Using normal mode

propagation code ORCA [Wes96], a synthetic gunshot emitted a time 0s is propagated in the *prior* Bay of Fundy waveguide (described in Section 2) over a 10-km range. Figure 2 gives the simulated received signal (noise free) and its spectrogram with the theoretical time-frequency arrival structure. The figure clearly illustrates the dispersive behaviour of the waveguide underlined in equation 3.

4. LOCALIZATION SCHEME

Our idea consists of using the dispersive behaviour of the waveguide to localize transient emissions. To design a localization scheme, we assume that:

- a sparse network of hydrophones is used to measure propagated signals,
- a source emits a transient signal with an unknown time-frequency modulation $t_e(f)$,
- from the recordings, we dispose of time-frequency processing that allows us to extract the time of arrival of any modal arrival for each frequency.

Times of arrival of mode with index m at frequency f measured at hydrophones n and n' are given by equation 4 (c.f Section 8) where $r(s,n)$ is the range between source s and receiver n . This implies that the TDoAs of modes with index m and m' at frequency f measured at hydrophones n and n' are given by equation 5 (c.f Section 8). If the ratio of $d(n,f,m',m)$ over $d(n',f,m,m')$ is computed, one obtains equation 6 (c.f Section 8). We can note that this ratio does not depend on the waveguide properties. So, if a geographical set of coordinates is defined with the origin set half way between hydrophones n and n' , the x coordinate along a line between n and n' and the y coordinate perpendicular to this line, it is easy to

show that the set of positions which satisfy $\frac{r(s,n)}{r(s,n')} = Q$

(where Q is a constant) is:

- if $Q \neq 1$, a circle with centre coordinates equal to $\left(-\frac{1-Q^2}{1+Q^2} \frac{L}{2}, 0\right)$ and radius equal to $\frac{QL}{|1-Q^2|}$ where L is the range between hydrophones n and n' ,
- if $Q=1$, the median between hydrophones n and n' .

$R(n,n',f,m,m')$ constrains the source to lie on a circle or on a line (under the assumptions that the channel is isotropic and range independent on the array area), it does not depend on waveguide acoustic properties, so it can be used to locate the source without any requirement about monitoring channel's properties, thus offering a robust localization scheme.

Then, our localization scheme follows these steps:

- step 1: perform a preliminary analysis of recordings to identify the hydrophones, the bandwidth and the modal indexes for which the modal arrivals are clearly resolved in a time-frequency plane,

- step 2: with a given time-frequency tool, extract the times of arrival $t_r(m,n,f)$ for any m, n and f selected in step 1 (in this paper, we look for a *prior* number of local maxima on the spectrogram computed with a Kaiser window, $\beta=0.1102(180-8.3)$, $L=512$ samples),
- step 3: for each quintuplet (m,m',n,n',f) with m',m,n,n',f selected at step 1 and $m \neq m', n \neq n'$, compute $R(n,n',f,m,m')$,
- step 4: estimate the source's location by solving the optimization problem described in equation 7 (c.f. Section 8) where (x_n, y_n) are the coordinates of hydrophones n , and N is the number of quintuplets (m,m',n,n',f) selected at step 3 (in this paper, we evaluate $J(x,y)$ on a discrete grid in x and y with a step of 5 m and search the global maximum of J on the grid but we can also envisage to use global or local optimization techniques).

5. APPLICATION ON REAL DATA

The localization scheme above designed is applied to 'S035-2' Bay of Fundy data set recordings which contain a gunshot call. Figure 1 illustrates the results. To obtain these localization results, our method was applied using hydrophones C, H, L, E, modal arrivals 1 and 2 and frequencies from 30 to 50 Hz (c.f. above step 1). Table 2 summarizes the localization results. Standard deviation of our method was assessed via 100 Monte-Carlo simulations with synthetic signals simulated with ORCA and with a signal to noise ratio similar to the real one (note that these Monte Carlo simulations help us to quantify the impact of recordings noise on localization accuracy but not the impacts of propagation and gunshot instabilities).

Method	x_{GS} (m)	y_{GS} (m)	Standard deviation x_{GS} (m)	Standard deviation y_{GS} (m)
Gervaise et al.	9225	-1248	420	110
Laurinelli et al. [Lau04]	8950	-970	760	620
Desharnais et al. [Des04]	8884	-848	---	---

Table 2 : S035-2, gunshot call localization results

6. DISCUSSION

The application of our method on real data from S035-2 recordings indicates that the gunshot localization is compatible with the solutions obtained by Laurinelli et al.'s ([Lau04]) and Desharnais et al.'s ([Des04]) methods, which somewhat confirms the validity of our approach. The precision of our scheme seems to be better than Laurinelli et al.'s approach but one must note that errors are not assessed the same way (Monte Carlo simulations with a statistical mean in our case, whereas Laurinelli et al.'s is obtained from the spread of hyperbole's

intersections in the hyperbolic fixing method and does not have any statistical meaning). Similar good results were obtained on S070-3 and S013-1 recordings. For S093-4 and S110-5 recordings, signal to noise ratio were too low to be able to clearly separate the modal arrival on a spectrogram and our method fails.

Compared to other methods, ours takes into account true propagation model that really exists in the waveguide while Laurinelli et al.'s and Desharnais et al.'s methods assume an acoustic direct ray path (straight line or not) propagation. In this sense, the link between time of arrivals and source position is better explained in our scheme. Although gunshot calls contain frequencies from 20 Hz to 20 kHz, our method exploits only the low-frequency band while Laurinelli et al.'s and Desharnais et al.'s may use the full bandwidth. Therefore, time of arrival estimates are more precise for Laurinelli et al.'s and Desharnais et al.'s schemes. Without ground-truth it is difficult to establish which method is the most accurate. However, we underline the fact that our approach does not require any knowledge about acoustic properties of the waveguide, which is a major advantage for robustness. Wiggins et al. [Win04] proposes a localization scheme based on normal mode propagation with a single hydrophone while our approach requires several hydrophones. To succeed, Wiggins et al.'s scheme needs a normal mode propagation code and the knowledge of acoustic properties of the waveguide. When applying his method on Bering Sea calls, the waveguide structure was simple (a Pekeris waveguide) and group speed weakly depended on waveguide properties. This was not the case in Bay of Fundy waveguide whose bottom presents a multi-layer structure with a poorly compacted first layer and a time-space variable sound speed profile. Thus a localization approach that is robust to poor knowledge of acoustic properties of the environment offers significant additional advantages even if it requires several hydrophones.

Our localization scheme can advantageously be used in a real-time anti-collision (between whales and ships) alert system in situations where low-frequency calls are frequent because it does not require the monitoring of acoustic properties of the waveguide, which simplifies the experimental PAM implementation and it does not require to run a propagation code, so a fast real-time localization may be achieved. However, it requires real-time implementation of time-frequency processing.

Because our scheme works on TDoAs between the arrivals of modes on a same hydrophone (see Eq 5), it is not sensitive to clock's drifts, which is a major difficulty with non-cabled hydrophone arrays (c.f. Simard and Roy, 2008 [Sim08])

In this paper, our approach was applied to right whale calls, but it can be used with any low-frequency calls with clear time-frequency modulation in shallow waters, for example with North Pacific right whales, Humpback

whales in the Bering sea [McD02], eastern North Pacific blue whales [Ole07], blue whale (*Balaenoptera musculus*) in the St. Lawrence Estuary [Ber06].

- include this localization scheme in a larger acoustic perspective to perform passive geoacoustic inversion of waveguide properties using marine mammal calls [Ger07].

6. FUTURE WORK

As a perspective for future work, we plan to:

- look for larger dataset to test the proposed approach (any contributions are welcomed),
- estimate the source's depth using the normal mode propagation assumption and the time-frequency analysis of the recorded signals,

7. ACKNOWLEDGEMENTS

The present work was done in collaboration with DGA (Délégation Générale pour l'Armement) under research contract n° CA/2003/06/CMO. The authors want to acknowledge N. Roy (Fisheries and Oceans Canada) for her support to analyse acoustic properties of Bay of Fundy waveguide.

8. EQUATIONS

$$H(f) \approx (2\pi)^{1/2} \sum_{m=1}^{+\infty} A(m, f, r, z_s, z_r) P(m, f, r)$$

Equation 1 $A(m, f, r, z_s, z_r) = g_m(z_s)g_m(z_r)\exp(-Ik_r(m, f)r) / \sqrt{Rk_r(m, f)r}$
 $P(m, f, r) = \exp(-j(Rk_r(m, f)r))$

Equation 4 $t_r(m, n, f) = t_e(f) + \frac{r(s, n)}{v_g(m, f)}$ $t_r(m, n', f) = t_e(f) + \frac{r(s, n')}{v_g(m, f)}$

Equation 5 $d(n, f, m, m') = t_r(m, n, f) - t_r(m', n, f) = r(s, n) \left(\frac{1}{v_g(m, f)} - \frac{1}{v_g(m', f)} \right)$
 $d(n', f, m, m') = t_r(m, n', f) - t_r(m', n', f) = r(s, n') \left(\frac{1}{v_g(m, f)} - \frac{1}{v_g(m', f)} \right)$

Equation 6 $R(n, n', f, m, m') = \frac{d(n, f, m, m')}{d(n', f, m, m')} = \frac{r(s, n)}{r(s, n')}$

Equation 7 $(x_s, y_s) = \arg \min_{x, y} (J(x, y))$
 $J(x, y) = \sum_{i=1}^N [(x - x_n)^2 + (y - y_n)^2 - R^2(m, m', n(i), n'(i), f)((x - x_{n'})^2 + (y - y_{n'})^2)]^2$

9. FIGURES

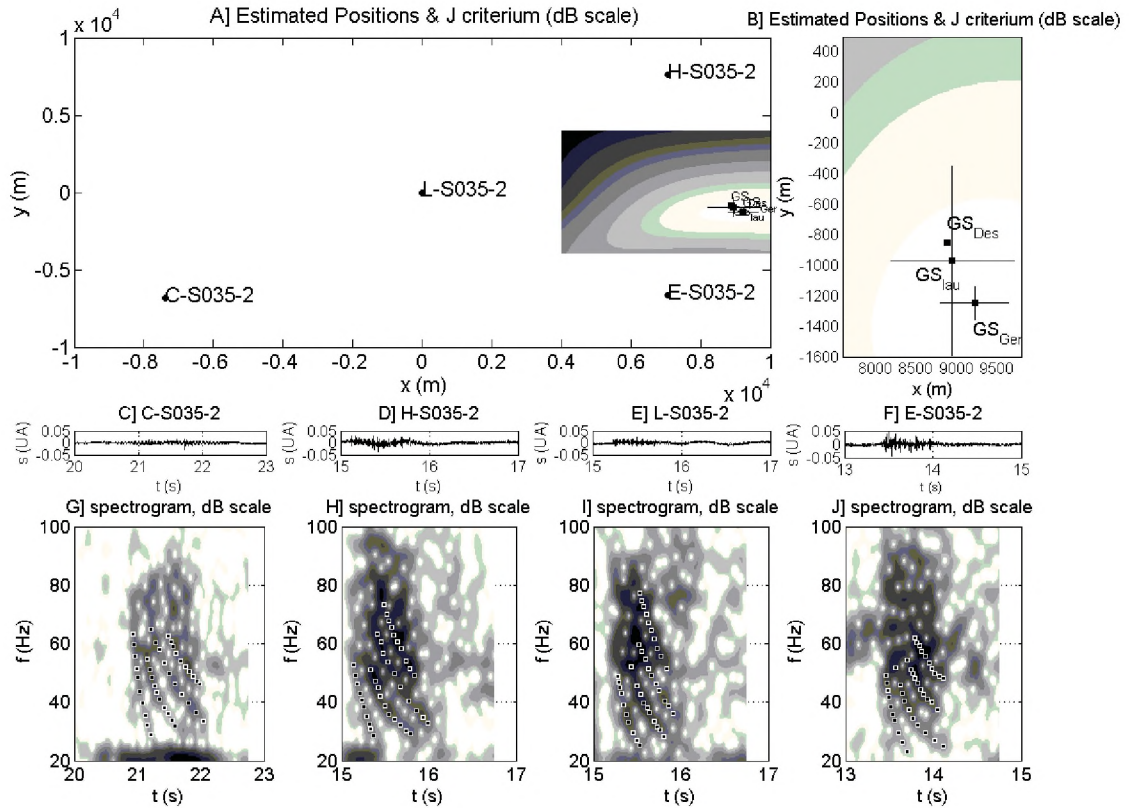


Figure 1) A : Estimated gunshot position and mapping of criteria J, B : zoom around the estimated positions, black square GS_{Ger} : estimated position with standard deviation from our method, black square GS_{Des} : estimated position with Desharnais et al.'s method, black square GS_{Lau} : estimated position with standard deviation from Laurinelli method; Panels C, D, E, F : received signals by OBHs C, L, E, H; panels G, H, I, J : received signal spectrogram (Kaiser window, $\beta=0.1102(180-8.3)$, $L=0.5s$) with estimated time-frequency law of modal arrivals (black square)

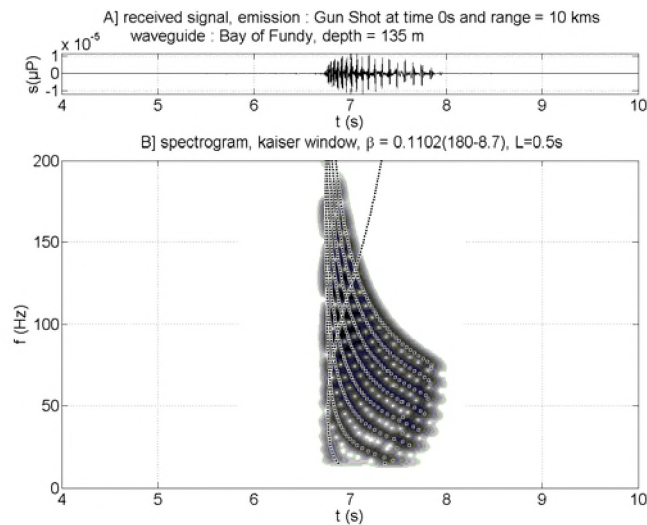


Figure 2) A] received waveform; B] gray scale map: received waveform spectrogram in dB scale, black squares: theoretical time-frequency structure of arrivals

10. REFERENCES

- [And01] M. André, Potential mitigation of fast-ferry acoustic & direct physical impact on cetaceans: toward a sustainable development of modern shipping, Workshop on Shipping Collision, 15th Annual conference of the European Cetacean Society, Roma, 2001
- [Ber06] C. L. Berchok, D. L. Bradley and T. B. Gabrielson, St. Lawrence blue whale vocalization revisited: characterization of calls detected from 1998 to 2001, *J. Acoust. Soc. Am.* 120(4), October 2006 p 2340-2354
- [Cha04] D. M. F. Chapman, You can't get there from here: shallow water sound propagation and whale localization, *Canadian Acoustics*, vol 32, no. 2, 2004 pages 167-171
- [Cla06] J.E.H. Clarke, Fundy oceanographic transect, CCGS F.G. Creed, August 3rd 2006, Ocean Mapping Group, University of New Brunswick, <http://www.omg.unb.ca/>
- [Ebb06] G. R. Ebbeson, F. Desharnais, Localization of right whales using matched correlation processing, *Canadian Acoustics*, vol.34 No 3 (2006), p 70-71
- [Des04] F. Desharnais, M. Côté, C. J. Calnan, G. R. Ebbeson, D. J. Thomson, N.E.B. Collison, C. A. Gillard, Right whale localisation using a downhill simplex inversion scheme, *Canadian Acoustics*, vol 32, no 2, 2004, p 137-145
- [Des04_a] F. Desharnais, M.H. Laurinelli, D.J. Schillinger, A.E. Hay, A description of the workshop dataset, *Canadian Acoustics*, vol.32 No 2 (2003), p 33-38
- [Ger07] C. Gervaise, S. Vallez, C. Ioana, Y. Stephan, Y. Simard, Passive acoustic tomography: review, new concepts and applications using marine mammals, *Journal of marine Biology Association of United Kingdom*, Vol. 87, p. 5-10, 2007
- [Jen00] F.B. Jensen, W.A. Kuperman, M.B. Porter, and H. Schmidt. Computational ocean acoustics. Springer-Verlag, New York, 2000.
- [Kno01] A. R. Knowlton, S. D. Kraus, Mortality and serious injury of Northern Right Whales (*Eubalaena glacialis*) in the Western North Atlantic ocean, *J. Cetacean Res. Manag.*, no2, pp 193-208, 2001
- [Lau03] M. H. Laurinelli, F. Desharnais, C. T. Taggart, Localization of North Atlantic right whale sounds in the Bay of Fundy using a sonobuoy array, *Marine Mammal Science*, 19(4):708-723, (October 2003)
- [Lau04] M. H. Laurinelli, A. Hay, Localisation of right whale sounds in the workshop Bay of Fundy dataset by spectrogram cross-correlation and hyperbolic fixing, *Canadian Acoustics*, vol 32, no 2, 2004, p 132-136
- [Lai01] D. W. Laist, A. R. Knowlton, J. G. Mead, A. S. Collet, M. Podesta, Collisions between ships and whales, *Marine Mammal Science*, 17(1):35-75; January 2001
- [Mat01] J. N. Matthews, S. Brown, D. Gillespie, M. Johnson, R. McLanaghan, A. Moscrop, D. Nowacek, R. Leaper, T. Lewis and P. Tyack, Vocalisation rates of the North Atlantic Right Whale (*Eubalaena glacialis*), *J. Cetacean Res. Manage.* 3(3): 271-282, 2001
- [Map77] Surficial geology – Eastern Gulf of Maine and Bay of Fundy. Map 4011-G published by the Canadian Hydrographic Service, 1977.
- [McD02] M. A. McDonald, 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
- [Par05] S. E. Parks, P. K. Hamilton, S. D. Kraus, P. L. Tyack, The gunshot sound produced by male North Atlantic Right Whales (*Eubalaena glacialis*) and its potential function in reproductive advertisement, *Marine Mammal Science*, 21(3):458-475 (July 2005)
- [Ole07] E. M. Oleson, J. Calambokidis, W. C. Burgess, M. A. McDonald, C. A. LeDuc, J. A. Hildebrand, Behavioral context of call production by eastern North Pacific blue whales, *Marine Ecology Progress Series*, vol.330:269-284,2007
- [Ric95] W. J. Richardson, C. R. Greene, C. I. Malme, D. H. Thomson, *Marine mammals and noise*, 1995, Academic Press, New York.
- [Spi90] J. L. Spiesberger, K. M. Fristrup, Passive localization of calling animals and sensing of their acoustic environment using acoustic tomography, *The American Naturalist*, vol. 135, No. 1, January 1990, p 107-153.
- [Sim06] Y. Simard, M. Bahoura, C.W Park, J. Rouat, M. Sirois, X. Mouy, D. Seebaruth, N. Roy, R. Lepage, Development and experimentation of a satellite buoy network for real time acoustic localization of whales in the St. Lawrence, in *Proceedings of IEEE/MTS Oceans'2006*, Boston, IEEE Cat. No. 06CH37757C Piscataway, NJ, USA. 6 p. 2006
- [Sim08] Y. Simard and N. Roy. Detection and localization of blue and fin whales from large-aperture autonomous hydrophone arrays: a case study from the St. Lawrence Estuary. *Canadian Acoustics 00:000-000 This issue*. 2008
- [Sta07] K.M. Stafford, D.K Mellinger, S.E Moore, and C.G. Fox. Seasonal variability and detection range modeling of baleen whale calls in the Gulf of Alaska, 1999–2002. *J. Acoust. Soc. Am.* 122: 3378-3390, 2007.
- [Tho86] J. A. Thomas, F. A. Awbrey, Use of acoustic techniques in studying whale behavior, *Rep. Int. Whal. Comm. (Special issue)* 8:121-38, 1986
- [Van03] A. S. M. Vanderlaan, A. E. Hay, C. T. Taggart, Characterization of North Atlantic right whale (*Eubalaena glacialis*) sounds in the Bay of Fundy, *IEEE JOE*, VOL. 28, No. 2, April 2003, p 164 – 173
- [Wes96] E. K. Westwood, C. T. Tindle, N. R. Chapman, A normal mode model for acousto-elastic ocean environments, *J. Acoust. Soc. Am.* Vol. 100, 6, December 1996
- [Win04] S. M. Wiggins, M. A. McDonald, L. M. Munger, S. E. Moore, J. A. Hildebrand, Waveguide propagation allows range estimates for North Pacific Right Whales in the Bering sea, *Canadian Acoustics*, vol. 32 No.2 (2004), p 146-154

PERFORMANCE OF THREE ACOUSTICAL METHODS FOR LOCALIZING WHALES IN THE SAGUENAY—ST. LAWRENCE MARINE PARK

Nathalie Roy¹, Yvan Simard^{1,2} and Jean Rouat³

¹Maurice Lamontagne Institute, Fisheries and Oceans Canada, P.O. Box 1000, Mont-Joli, Québec G5H-3Z4, Canada

²Marine Sciences Institute, University of Québec at Rimouski, P.O. Box 3300, Rimouski, Québec G5L-3A1, Canada

³Department of electrical and computer engineering, University of Sherbrooke, 2500 boul. de l'Université, Sherbrooke, Québec, J1K 2R1, Canada

ABSTRACT

Three algorithms are explored to localize fin whale calls recorded from a large-aperture hydrophone array deployed in the Saguenay—St. Lawrence Marine Park. The methods have to cope with varying sound speed in space and time, errors in time differences of arrival (TDoA) measurements in a noisy environment, and often a limited number of hydrophones having recorded a particular event. The array was composed of 5 AURAL autonomous hydrophones with a total aperture of about 40 km, coupled with 2 hydrophones from a small-aperture cabled coastal array. The autonomous hydrophones clock drifts were estimated with a level of uncertainty from timed sources and the coastal array time reference. The calls were then localized by constant-speed hyperbolic fixing, variable-speed isodiachron Monte-Carlo simulations, and a ray-tracing propagation model. The Monte-Carlo simulations generate clouds of possible localizations from the uncertainty in hydrophone positions, TDoAs and the effective horizontal sound speeds along the different source-hydrophone paths. The ray-tracing model produces a fixed grid of TDoAs which can then be consulted to find the likeliest positions of the whales. Results from the different methods are compared and their relative advantages or limitations are discussed.

RÉSUMÉ

Trois algorithmes sont explorés pour la localisation de vocalises de rorqual commun enregistrées par un réseau d'hydrophones à large ouverture déployé dans le Parc Marin du Saguenay–Saint-Laurent. Les méthodes doivent composer avec une vitesse du son variable dans l'espace et le temps, des erreurs dans les mesures des différences de temps d'arrivée (DTA) avec un environnement bruyant, et souvent un nombre limité d'hydrophones ayant capté un événement donné. Le réseau était composé de 5 hydrophones autonomes AURAL avec une ouverture totale d'environ 40 km, couplé avec 2 hydrophones d'un petit réseau côtier. La dérive des horloges des hydrophones autonomes a été évaluée avec un niveau d'incertitude à l'aide de sources aux temps connus ainsi que de la référence temporelle du réseau côtier. Les vocalises ont ensuite été localisées par la méthode à vitesse constante des hyperboles, par celle à vitesse variable des isodiachrones avec simulations de Monte-Carlo, et par un modèle de propagation de rayons. Les simulations de Monte-Carlo produisent des nuages de localisations possibles à partir des incertitudes sur les positions des hydrophones, sur les DTAs et sur les vitesses horizontales effectives du son le long des différentes trajectoires source-hydrophone. Le modèle de propagation des rayons produit une grille fixe de DTAs qui est ensuite consultée pour trouver les positions les plus probables des baleines. Les résultats des différentes méthodes sont comparés et leurs avantages ou limites relatives sont discutés.

1. INTRODUCTION

Continental shelf marine environments are challenging for acoustic localization methods because of varying complex bathymetry, 3D oceanographic processes affecting temperature and sound speed time-space structures, especially at tidal and seasonal frequencies. Localization from TDoAs on short distances also requires high time precision with special care given to clock synchronization of the array and time drifts during mid- and long-term deployments. The most used localization method, hyperbolic fixing (Spiesberger and Fristrup 1990), assumes constant speed over the 2D or 3D localization space. The source location is assumed to be a linear function of travel

time differences, speed of sound and receiver locations. Errors in localization can be large when these assumptions are not satisfied and uncertainties in the input data are present (e.g. Spiesberger and Wahlberg 2002). These conditions generally prevail in the study area at the head of the Laurentian Channel in the St-Lawrence Estuary, where the summer sound speed profile is characterized by a well-defined channel at intermediate depths (e.g. Fig. 1a), 3D physical processes including semi-diurnal tidal upwelling and higher frequency of internal waves or fronts resulting from the interaction of tidal currents with the complex bathymetry combining with the confluence of several estuarine water masses (e.g. Saucier and Chassé 2000).

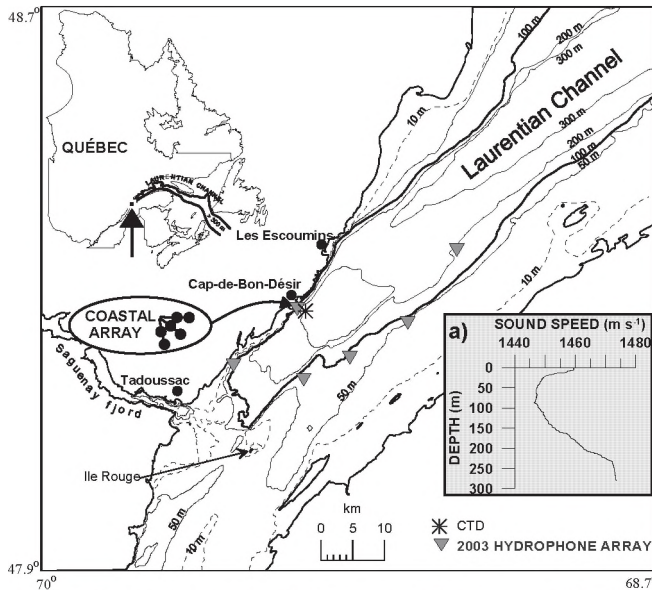


Figure 1. Study area in the Saguenay–St. Lawrence Marine Park, with bathymetry, locations of the 5 AURAL M1 autonomous hydrophones and the 6-hydrophone coastal array, CTD station and sound speed profile.

Three methods have been explored here to compare their localization performance using an array which includes autonomous hydrophones, each having its own clock, under this general context of considerable uncertainty in input data. They are the hyperbolic fixing (Spiesberger and Fristrup 1990, Spiesberger 1999, 2001), the isodiachron method with Monte-Carlo simulations (Spiesberger and Whalberg 2002, Spiesberger 2004) and the use of an acoustic propagation model (Tiemann and Porter 2004).

2. MATERIAL AND METHODS

Data collection

The hydrophone arrays were deployed in the study area during summer 2003 (Fig. 1). All hydrophones were HTI 96-min with a nominal receiving sensitivity (RS) in the low frequency band (< 2 kHz) of -164 dB re 1 V/ μ Pa. Five AURAL autonomous hydrophone systems (Multi-Electronique Inc, Rimouski, Qc, Canada) programmed to sample continuously 16-bit wave data over the 1 kHz band were deployed as oceanographic moorings. The hydrophones were placed at intermediate depths in the water column close to the summer sound channel axis. Special care was taken to minimize possible noise sources from the moorings. The outer 2 hydrophones from a 650-m aperture cabled coastal array deployed along Cap-de-Bon-Désir (Fig. 1,) completed the 7-hydrophone data base used in this study. The acquisition system for the coastal array was a 16-bit ChicoPlus Servo-16 data acquisition board (Innovative Integration, Simi Valley, CA, U.S.A.). After analysis of the recordings, this beta version of AURAL M1 was found to have a clock drift of about 18 s per day, consistent on all 5 instruments. The coastal array's PC clock had a drift of about 10 s per day, which was checked and corrected every weekday morning. The coastal and AURAL arrays were

synchronized using simultaneous recording of the same acoustic signals such as motor boats and whale vocalizations, and linear time interpolations assuming constant drift. Timing errors are inherent to such a procedure and the localization method must be robust enough to deal with such uncertainties as well as the non spatially homogenous effective sound speed.

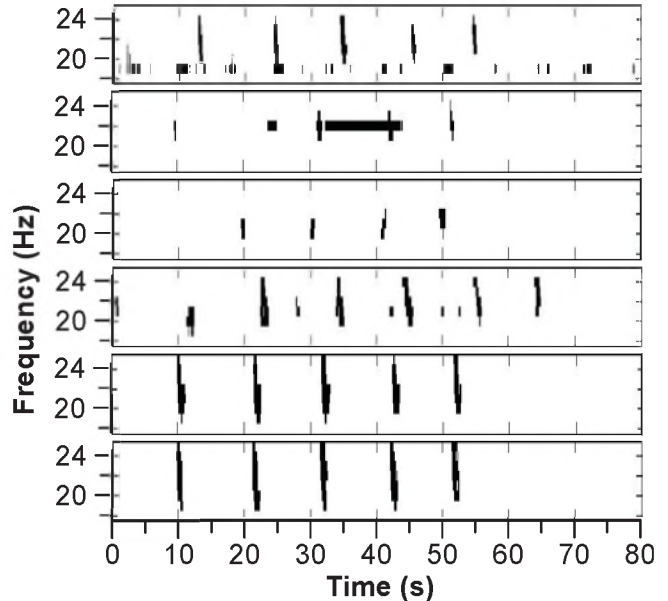


Figure 2. Binarized image of spectrogram from a fin whale series of calls from 6 hydrophones of the array.

A typical sound speed profile from a CTD cast made in the area is shown in Fig. 1a.

Data analysis

The 80-s sample used for localization is a series of fin whale pulse calls in the 18-25 Hz frequency band (Fig. 2) recorded Sept. 24 at 4:49 local time. This sample was detected on all hydrophones except one where background noise was masking the call. The TDoAs between hydrophones is determined by spectrogram cross

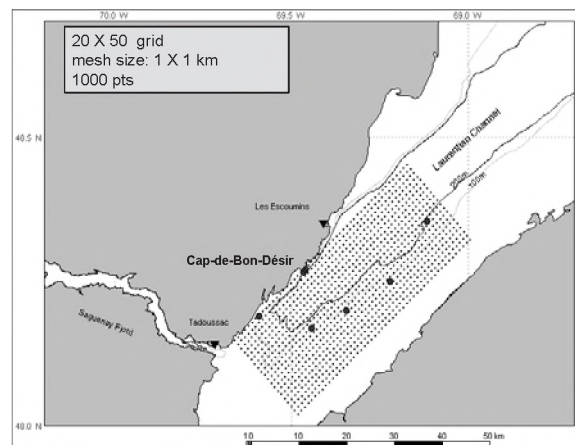


Figure 3. Ray model grid of 1000 points covering 20 × 50 km in the area of study and associated with the 7 receiver positions.

coincidence (e.g. Simard et al. 2004), after bandpass filtering to [18 25] Hz.

A constant sound speed of 1450 m s^{-1} , corresponding to the average speed in the sound channel where the hydrophones were deployed, was used for hyperbolic analysis and as the central speed of the interval used in isodiachron Monte-Carlo applications.

The 1-km resolution grid used for the ray-tracing Bellhop model (Porter and Liu 1994) covered an area of $20 \times 50 \text{ km}$ enclosing all hydrophones (Fig. 3). The typical sound speed profile of Fig. 1a was used for the modeling. For each grid point, a set of TDoAs was calculated assuming a source at a depth of 10 m, a frequency of 500 Hz, and using 21 rays in the $\pm 20^\circ$ directions along the propagation path plane. The bathymetry profile between source and receiver was extracted from a high-resolution multibeam bathymetry dataset provided by the Canadian Hydrographic Service. The mean times of ray arrivals, weighted by ray amplitudes, provided by Bellhop for the array configuration were then used for TDoA calculations. The sound source was located at the minimum Euclidean distance between measured and modeled TDoAs (Tiemann and Porter 2004).

Isodiachrons are an extension of the hyperbolic location method but where the effective sound speed along the path between source and each receiver is allowed to vary from path to path. Receivers are combined in groups of three, which gives a total of 20 groups from 6 receivers. For each group, sound speed, receiver positions and TDoAs are treated as uniformly distributed random variables within a chosen interval that is the best educated guess of data uncertainty. Monte-Carlo simulations are then applied to produce a probability density function (pdf) of source location, called a constellation, for each group of 3 receivers. The actual source is then located within the intersection of all the constellations at the most probable

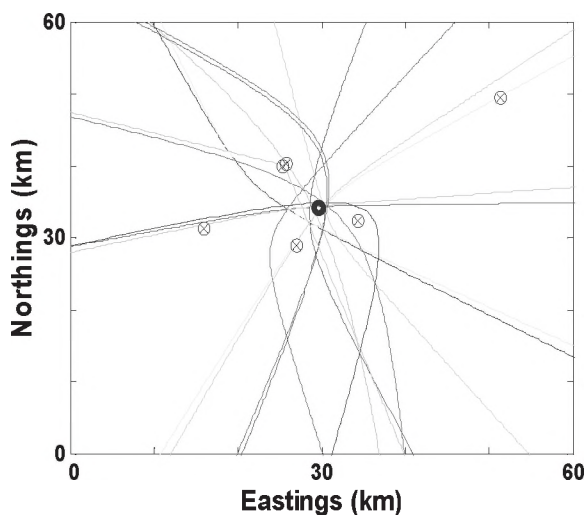


Figure 4. 2D hyperbolic localization of the fin whale calls. Hydrophone positions are illustrated as crossed circles. Whale position is the bold circle.

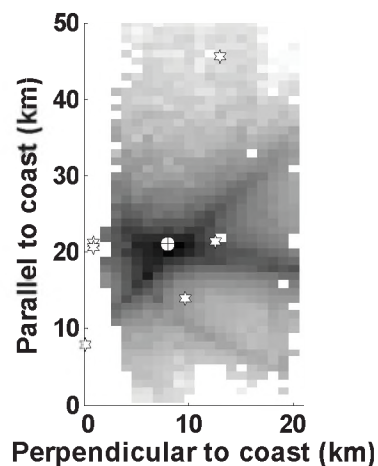


Figure 5. Localization of fin whale on the ray-tracing model grid. The receivers are illustrated by stars; the whale position is a crossed circle in the darker zone of the figure. Background grayscale image represents difference between measured and modeled TDoAs; smaller differences are darker. White patches are areas where a source would not be heard by all receivers.

position from the joint probability distribution function (pdf) along X and Y dimensions (i.e. 2D histogram) of all solutions.

3. RESULTS

Resulting hyperbolic fixing from the set of TDoAs is presented in Fig. 4. The position found for the fin whale is 48.214° N , 69.408° W , within the center of the array configuration. Estimated fixing error on this result is 1690 m, derived from the norm of the differences between

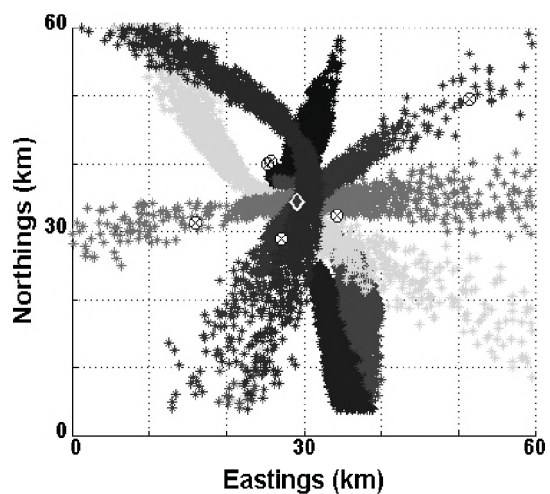


Figure 6. Localization process using isodiachron technique with Monte-Carlo simulations. The receivers are illustrated by crossed circles. Different constellations are shown in shades of gray. The estimated whale position is represented by a diamond.

estimated and measured time along every path (Simard et al. 2004).

The same set of TDoAs was applied to the ray tracing TDoA estimation grid to produce results shown in Fig. 5. The background grayscale image represents difference between measured and modeled TDoAs; smaller differences are darker. The whale position is in the darkest area of the grid at a position of 48.220° N, 69.384° W. Grid size implies an uncertainty of at least ± 1 km.

Fig. 6 illustrates results obtained from the isodiachronic Monte-Carlo method. Each constellation was obtained from 4000 different estimates of source location from a varying set of input values for receiver location, sound speed and TDoAs. Assumed errors for these variables were ± 20 m, ± 5 m s⁻¹ and ± 1.0 s respectively. The 1.0 s error is the minimal value needed to obtain intersection of all constellations. The region of intersection is a 600×800 m rectangle, the presumed position of the source being where the density is highest. The estimated whale position is 48.216° N, 69.421° W. Confidence intervals for 95% of the pdf are 580 m for x and 770 m for y.

Figure 7 shows the whale positions obtained from the three methods. The isodiachron result is at 1.01 km from the hyperbolic position and at 2.24 km from the result of the ray-tracing model grid.

4. DISCUSSION

The whale was localized in one of the 3 intensive feeding spots found at the head of the Laurentian channel (Mingelbier and Michaud 1996), where tidal upwelling along the slope concentrates krill in dense demersal patches

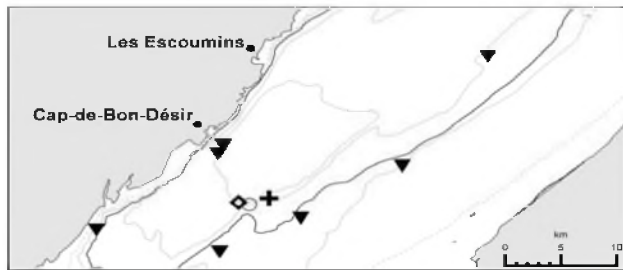


Figure 7. Locations of fin whale from the three methods: diamond for isodiachron, circle for hyperbolic result and cross for ray-tracing model grid. The receivers are illustrated by triangles.

(Cotté and Simard 2005) that are exploited by fin whales from tagging experiments (e.g. Simard et al. 2002 Fig. 3b1). We observed strong whale blows (either from fin or blue whales) from the coastal array location at Cap-de-Bon-Désir on the north shore on the same morning during daylight, which were about 4 km from the localized fin whale, 6 hours later; so the localization found by the three methods is very likely given the fidelity to the feeding site over several hours.

This localization example illustrates the need for high precision in the set of variables involved in the localization problem and an accurate propagation model, which is difficult to satisfy because of technical constraints due to complex water mass structure combined with complex bathymetric characteristics. The isodiachron Monte-Carlo technique indicated that the error on TDoAs was in the order of 1 s. Even if the whale is favourably positioned at the center of the array, both the hyperbolic and ray-tracing solutions have uncertainties exceeding 1 km, given a 1-s travel time at 1450 m s⁻¹. In both cases, an estimate of the confidence interval of the localization reflecting uncertainties in the input values would be needed (c.f. Spiesberger and Wahlberg 2002), but is not formally provided by the methods. As mentioned by Tiemann and Porter (2004), a Monte-Carlo approach could be used to evaluate uncertainties in localization by incorporating the measurement error in TDoAs. However, the uniform sound speed profile over the entire grid is an unrealistic model condition that would also require special attention in such variable environments. This would add substantial modeling efforts. Monte-Carlo simulations can also be applied to the hyperbolic method to find the optimal effective homogeneous sound-speed, which corresponds to the isodiachron particular case when the sound speed is constant. Indeed, what we found most useful in the isodiachron Monte-Carlo method is the application of pdfs to evaluate input error magnitude.

Our estimation of source location is done by finding the area of highest density on a 2D histogram of all possibilities in the overlapping constellation area. Spiesberger and Wahlberg (2002) used separate pdfs along X and Y dimensions to compute the confidence interval of the solution. We found the joint 2D pdf more practical in a context of comparing method precision and also for whale tracking purposes, where a best estimate of whale position has to be extracted. However, the 2D histogram becomes useless if very few points are present in the constellation intersection area, or when two or more equivalent peaks are found in the distribution. Then the computation of the center of gravity by principal component analysis might be more adapted to estimate the source localization.

Another difficulty is estimating confidence limits for the source localization. Spiesberger et al. (2002) proposed to use the x and y sizes of the constellations with the smallest confidence limits, defined as two standard deviations. We found this approach misleading when constellations are spread out in long elliptic shapes, and distributions can stretch out to 100 km or more. Our approach uses the pdf of the localizations in the constellation intersecting area. However, in both cases, the limitation of this Monte-Carlo method is the strong dependence of output confidence interval on input error bounds. They should ideally be independent.

In our test, the ray-tracing model grid's precision is limited by mesh size. Building the grid is demanding in computation time, although some degree of interpolation can be applied to reduce the number of grid points to compute. Computation time can be reduced by inverting the

source-receivers in computing the TDoAs along equally spaced directions for each hydrophone and then interpolating on the grid from the 7 propagation times (M.B. Porter pers. comm.). However, the grid has to be regenerated when a new configuration of receivers is used. Another drawback is that a position is always found on the grid even if the source is outside the domain or if measurement errors are high, which requires position validation by other means. To be fair, the grid method should be tested with measured data with a known source; it might actually outperform other methods in cases where input data have relatively small errors and where sound speed and ray trajectories are not uniform. In our case study, calls detected on more than three hydrophones were rare, due to the large aperture of the array and low number of hydrophones combined with high shipping noise (e.g. Simard et al 2006). Testing the localization from a low-frequency source would be suitable, notably for evaluating the effective sound speed and for synchronizing clocks, but such large-size and high-power sources are specialized equipments that are not easily available and deployable besides posing ethical problems because of their potential negative impact on fauna.

For future work perspective, the clock synchronization problems encountered in this initial deployment of the array in 2003 were clearly the main source of localization error. Then, at an order of magnitude lower, come the sound speed variation and the precision and accuracy of TDoA estimates. Increasing the number of hydrophones to form a denser array may improve probability of detection and reduce the localization error. Then regular sound speed profile measurements over the area could further improve the localization. Acquiring frequently updated profiles would need the regular use of a ship equipped with a CTD or sound speed profiler. Alternatively, sound speed profiles could eventually be estimated from the recorded acoustic data on the array by passive acoustic tomography techniques using the transiting merchant ships as sources of opportunity to monitor the environment in a non-invasive way (C. Gervaise, Ensieta, Brest, France, personal communication).

5. ACKNOWLEDGEMENTS

This work was supported by the Fisheries and Oceans Canada (DFO) Chair in applied marine acoustics at ISMER-UQAR, DFO Maurice Lamontagne Institute species at risk program and FQRNT Quebec research fund. We thank the crews of the M/V Coriolis II, NGCC Isle Rouge and all the technicians, students and assistants involved in preparing the material, its deployment at sea and the data acquisition. We also thank Parks Canada Saguenay–St. Lawrence Marine Park for their steady collaboration. The computer program for applying the isodiachronic technique was developed at Sherbrooke University as part of a Master thesis. We would like to thank students Diya Sebaruth and Hansa Devi Gukhool for their invaluable support.

6. REFERENCES

- Cotté, C., and Simard, Y. 2005. The formation of rich krill patches under tidal forcing at whale feeding ground hot spots in the St. Lawrence Estuary. *Mar. Ecol. Progr. Ser.* 288: 199-210.
- Mingelbier, M., and Michaud, M. 1996. Étude des activités d'observation en mer des cétacés de l'estuaire maritime du Saint-Laurent: compte rendu de l'échantillonnage de 1995 et synthèse des données prélevées de 1984 à 1995. Final report to Parks Canada, Canadian Heritage Department, Ottawa. GREMM, 108 de la Cale Sèche, Tadoussac, QC G0T 2A0.
- Porter, M.B., and Liu, Y.C. 1994. Finite-Element Ray Tracing. *Proceedings of the International Conference on Theoretical and Computational Acoustics*, Eds. D. Lee and M. H. Schultz (World Scientific, Singapore 1994): 947-956.
- Saucier, F.J., and Chassé, J. 2000. Tidal circulation and buoyancy effects in the St. Lawrence estuary. *Atmosphere-Ocean* 38: 505-556.
- Simard, Y., Bahoura, M. and Roy, N., 2004. Acoustic detection and localization of whales in Bay of Fundy and St. Lawrence estuary critical habitats. *Canadian Acoustics* 32 (2):107-116.
- Simard, Y., Lavoie, D. and Saucier, F.J. 2002. Channel head dynamics: Capelin (*Mallotus villosus*) aggregation in the tidally-driven upwelling system of the Saguenay–St. Lawrence Marine Park's whale feeding ground. *Can. J. Fish. Aquat. Sci.* 59: 197-210.
- Simard, Y., Roy, N. and Gervaise, C. 2006. Shipping noise and whales: World tallest ocean liner vs largest animal on earth. OCEANS'06 MTS/IEEE – Boston, IEEE Cat. No. 06CH37757C Piscataway, NJ, USA.
- 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. 2004. Geometry of locating sounds from differences in travel time: Isodiachrons. *J. Acoust. Soc. Am.* 112: 3046-3052.
- 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., and Porter, M.B. 2004. Localization of marine mammals near Hawaii using an acoustic propagation model. *J. Acoust. Soc. Am.* 115: 2834-2843.

For
Digital Recorders

Introducing

For
USB A/D Systems

PHANTOM POWER

7052PH

Measurement Mic System

7052H Type 1.5™

Titanium Diaphragm

3Hz to >20 kHz

<20 dBA > 140 dB SPL

MK224 (<14 dBA to >134 dB SPL) Optional

4048 Preamp

Superior
IEC 1094 Type 1
Long-term Stability
Temperature and Humidity
Performance

Now in Stock



**Phantom
to IEPE/ICP
Adaptor
Supplies 3-4 mA
Power
Accelerometers
Microphones**

ICP1248

A **B**
C **e**
O **g**
u **i**
S **n**
t **s**
i **w**
c **i**
S **t**
O **h**
A
C
O



**MATT™
Family**

Mic Attenuator

Handle Higher Sound Pressure Levels

ACO Pacific, Inc., 2604 Read Ave., Belmont, CA 94002

Tel: (650) 595-8588 FAX: (650) 591-2891 E-Mail: sales@acopacific.com

Web Site: www.acopacific.com

TM

PASSIVE ACOUSTIC MEASUREMENT OF DIVE VOCAL BEHAVIOR AND GROUP SIZE OF BLAINVILLE'S BEAKED WHALE (*MESOPLODON DENSIROSTRIS*) IN THE TONGUE OF THE OCEAN (TOTO)

Nancy DiMarzio¹, David Moretti¹, Jessica Ward¹, Ronald Morrissey¹, Susan Jarvis¹,
Anna Maria Izzi³, Mark Johnson², Peter Tyack² and Amanda Hansen²

1 - Naval Undersea Warfare Center Division, 1176 Howell St., Newport, R.I., USA

2 - Woods Hole Oceanographic Institution, Woods Hole, MA., USA

3 - University of Rhode Island, Narragansett, R.I., USA

ABSTRACT

The vocal behavior of Blainville's beaked whale (*Mesoplodon densirostris*) was measured using the bottom-mounted hydrophones of the Atlantic Undersea Test and Evaluation Center (AUTEK) in the Bahamas. The statistics for the vocal durations and gaps within these vocal periods were measured over multiple deep foraging dives. The sizes of foraging groups of *M. densirostris* were estimated from the dive vocalization durations by applying click rate and detection ratio statistics derived from Woods Hole Oceanographic Institution (WHOI) Digital recording Tags (DTags) to visually verified data collected on the AUTEK range.

SOMMAIRE

Le comportement vocal de la baleine à bec de Blainville (*Mesoplodon densirostris*) a été mesuré à l'aide d'hydrophones installés sur le fond marin, au Centre d'évaluation et de tests sous-marins de l'Atlantique (AUTEK, de l'anglais Atlantic Undersea Test and Evaluation Center) des Bahamas. Les statistiques relatives à la durée des signaux acoustiques et les périodes qui s'écoulent entre les périodes de « chant » ont été mesurées pour des plongées multiples en profondeur (en comportement de recherche de nourriture). La taille des groupes de *M. densirostris* en recherche de nourriture ont été estimés d'après la durée des périodes de chant en plongée, en appliquant le taux de « clics » et des statistiques sur le rapport de détection dérivés des enregistrements numériques du DTags de la WHOI (*Woods Hole Oceanographic Institution*), afin de visualiser les données recueillies au centre AUTEK.

1. INTRODUCTION

Tests of a marine mammal passive acoustic monitoring, detection, and localization system have been carried out at the U.S. Navy's Atlantic Undersea Test and Evaluation Center (AUTEK) training range in conjunction with the Bahamas Marine Mammal Research Organization (BMMRO) and the Woods Hole Oceanographic Institution (WHOI). BMMRO trained observers were directed to vocalizing animals on the range based on passive acoustic detections and localizations. They visually verified species and group size, and obtained photo-IDs of animals. WHOI placed DTags on 7 Blainville's beaked whales (*Mesoplodon densirostris*). The DTag records pitch, roll, heading, and depth, and records acoustic signals using stereo hydrophones. The tag provides detailed data on the tagged animal's movements and vocalizations.

The AUTEK range is located in a deep ocean canyon known as the Tongue of the Ocean (TOTO) off Andros Island in the Bahamas. The range consists of 82 bottom-mounted hydrophones that are deployed up to depths of 2000 meters. Sixty-eight of the hydrophones are arranged in offset rows on approximately 4 km baselines. The bandwidth of these hydrophones is 50 Hz to approximately

45 kHz. Fourteen additional hydrophones are arranged into two 7-hydrophone hexagonal arrays with a center hydrophone. These hydrophones are separated by a baseline of about 1.2 km and have a bandwidth from 8 to 50 kHz [6, 7].

Since 2004 *M. densirostris* has been the focus of this research, as it is the species generally sighted by observers at AUTEK, and is frequently detected acoustically on the range. This species has been associated with mass strandings linked to Navy sonar.

In a study from 1997-2002 in the Northern Bahamas, D. Claridge found the mean group size of *M. densirostris* to be 4.1, with group sizes ranging from 1 to 11 animals [2]. Pairs of both *M. densirostris* and *Z. cavirostris* tagged simultaneously with DTags have been reported to dive and subsequently approach the surface in close proximity [5, 10]. Visual observations of *M. densirostris* at AUTEK by BMMRO observers on the water and passive acoustic monitoring on-shore also indicate that these animals dive as a group and vocalize at depth. Personnel acoustically monitoring the AUTEK range hydrophones detect groups of *M. densirostris* across the range throughout the course of any given day. This species is known to dive to depths in excess of 1200 meters [1, 8]. These dives are separated on

average by about 2 hours and the dive duration may exceed 60 minutes [1]. Vocalizations are produced only at depth during foraging. Vocalizations generally occur below 200 m [3].

M. densirostris produce short (~250 μ s) frequency-modulated upsweeps from approximately 25 to 55 kHz, with an inter-click interval (ICI) of 0.2 to 0.4 s [3, 4]. The beam pattern of another beaked whale species, Cuvier's beaked whale (*Ziphius cavirostris*), has been shown to be highly directional [10]. The same *M. densirostris* clicks are often detected concurrently on only 1 or 2 hydrophones, depending on the animal's bearing and direction of movement, making localization difficult. However, as individuals in the group move during foraging, the group's clicks are detected on a number of surrounding range hydrophones, with click detections often shifting back and forth between adjacent hydrophones. When a group of *M. densirostris* starts vocalizing, a few intermittent clicks are often heard before they start clicking fairly consistently on the surrounding hydrophones. Similarly, as foraging comes to an end, a few intermittent clicks can still be detected after consistent clicking has ceased.

Clicks are detected using a multi-stage FFT-based energy detector [7]. The sample rate for the data is 96 kHz. A 2048-point FFT with 50% overlap is used. It provides a 46.875 Hz frequency resolution and a 10.67 ms time resolution. An adaptive threshold is run in each bin of the FFT. If energy above threshold is detected, the bin is set to a "1" and a detection report is generated. *M. densirostris* clicks are then isolated based on the frequency distribution of the detected signal [6].

Raw acoustic, click detection, and localization data have been collected at the AUTECH range during these tests. For this study dive vocalization durations of *M. densirostris* were measured from the click detection data, and a method was developed to estimate *M. densirostris* group size given this dive vocalization duration.

Groups of vocalizing beaked whales were isolated and the durations of the vocal periods during the dives for the groups were measured. Gaps in the vocalizations during these vocal periods were also measured.

To estimate group size, a normalized detection ratio and mean click rate was derived from DTag dive data. The normalized detection ratio is the ratio of total number of clicks detected on the array to the product of the total number emitted by the tagged animal multiplied by known group size and the number of hydrophones in the array. These data were used to predict the number of animals in another group based on the total number of clicks detected for this group during its dive vocal period on a hydrophone array.

In October of 2006, a WHOI DTag was attached to a Blainville's beaked whale at the AUTECH range in the Bahamas. The tag was recovered after 17 hours. Five deep foraging dives were recorded. Statistics from three of these dives were used as a predictor of group size using verified data collected at AUTECH during an independent exercise conducted with BMMRO visual observers.

2. METHODS AND RESULTS

2.1 Vocal Behavior

Click detection data from a species verification test conducted at AUTECH in April, 2005 with BMMRO and from a test conducted with The Technical Cooperation Program (TTCP) at AUTECH in February, 2007 were analyzed to quantify the vocalization periods for groups of diving *M. densirostris*. Groups were isolated and baseline dive vocalization durations were measured.

A program that recorded the start and stop time of beaked whale detections on each hydrophone was run on archived detection data. The program recorded a start time when a beaked whale click was detected on a hydrophone. A 60-second timer was then initialized for this hydrophone. If a detection did not occur within the 60-second window, the timer was stopped, and the time of the last detected click marked as the stop time. Each time a new click was detected, the timer was reinitialized. The program output the start time, stop time, and total click count for each hydrophone that had detections. The total click count is the number of clicks detected between a single start and stop time on a single hydrophone.

To isolate animal groups and measure dive vocalization durations, the start and stop data from groups of adjacent hydrophones were overlaid on a single plot. Generally, 7-phone arrays centered on the presumed group location were examined. AUTECH hydrophones are laid in offset rows with a baseline of ~4 km to form hexagonal arrays with a center phone. Based on tag analysis, beaked whale vocalizations are detected up to a range of 6.5 km. In order to help isolate vocalizing groups, total click counts of less than 5 clicks were discarded. The vocalization duration was derived by measuring the start and stop time of detections on the set of hydrophones believed to represent a single beaked whale group. Durations for the April, 2005 dataset were measured from the onset of consistent clicking on the analysis hydrophone array to the cessation of this consistent clicking. Durations for the February, 2007 dataset were measured from the first to the last detected click. A characteristic plot of detection start and stop time versus total click count for 3 hydrophones over 3 dives is given in *Figure 1*. In the figure, the vocalization durations range from 20 to 21 minutes. Dives with vocalizations detected only on hydrophones on the edge of the range were not considered as part of the analysis. Vocalization periods longer than 52 minutes, twice the mean dive vocal phase of 26 minutes measured for *M. densirostris* from DTag data [9], were also not considered. Difficulty in measuring dive vocal periods from the range hydrophones arises when two or more groups in close vicinity ensound common sensors.

Sixty-four dives from April, 2005 and thirty-three dives from February, 2007 believed to be associated with single beaked whale groups were isolated and the vocalization durations measured. Dive vocalization statistics are shown in Table 1 and Figure 2.

Gap statistics were measured from the February, 2007 dataset. Gaps in the vocalizations refer to periods of time when fewer than five clicks from the group were detected

simultaneously on *any* given hydrophone associated with the group.

These gap statistics provided baseline data for typical beaked whale detections on the AUTEK range for the 2007 Behavioral Response Study (BRS). During this test a tagged Blainville's beaked whale was exposed to anthropogenic sound during a deep foraging dive. The animal's vocalizations were monitored using the AUTEK sensors during the exposure. Changes in typical vocal behavior were used as a real-time cue to the test conductor to secure the source.

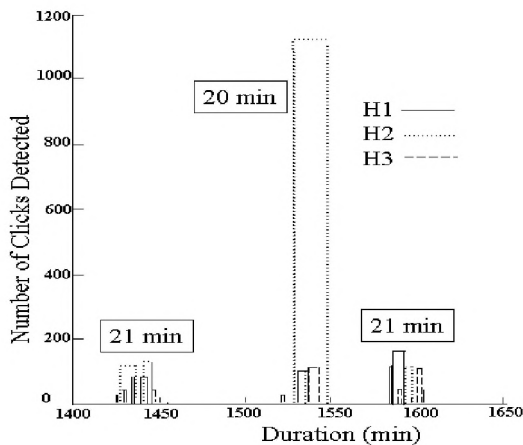


Figure 1: *M. densirostris* vocalization start and stop times vs. total number of clicks detected within the period for 3 dives on 3 hydrophones.

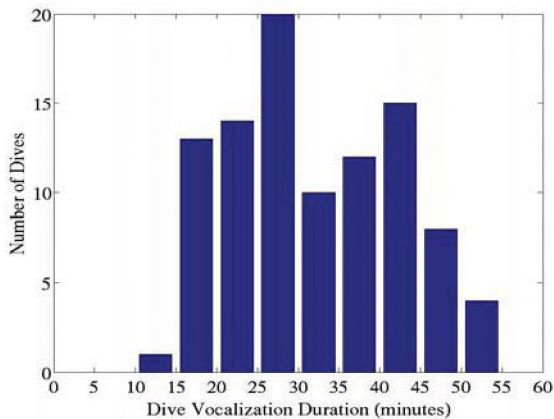


Figure 2: Vocalization duration vs. number of groups for 97 dives from the two datasets analyzed.

	April, 2005	February, 2007	Combined
# Dives	64	33	97
Mean	31.4	31.8	31.5
Median	28.5	32.6	30.3
Mode	26.0	37.0	26.0
Std Dev	9.9	10.5	10.6
Min	15.0	14.9	14.9
Max	51.0	50.4	51.0

Table 1 Dive vocalization statistics

Thirty-three dives were examined from 01/30/07-02/04/07, twelve of which had no gaps. A total of 63 gaps were measured; of these only five were more than 6 min long. 10% of the gaps started between 40 and 60% of the way through the corresponding dive vocalization duration. The gap statistics are indicated in Table 2.

	Mean	Median	Mode	Std Dev	Min	Max
# Gaps	1.9	1.0	0.0	2.3	0.0	8.0
Gap (min)	2.5	1.3	1.3	2.3	0.0	12.7

Table 2: Gap statistics

2.2 Click Counting

A method of click counting for the range hydrophones was developed to estimate the number of *M. densirostris* present in a group. This method uses the dive vocalization duration for each group, the total number of clicks detected on surrounding range hydrophones, and statistics (mean click rate and normalized detection ratio) derived from DTag data to estimate group size. DTag data were used in conjunction with nearby range hydrophones to estimate the mean click rate for an individual and the *normalized detection ratio* of the hydrophone sub-array of interest. The click rate is the number of clicks emitted by an individual per second during the vocal periods within deep foraging dives. The normalized detection ratio is the ratio of total number of clicks detected on the array to the product of the total number emitted by the tagged animal multiplied by known group size and the number of hydrophones in the array. The mean click rate and normalized detection ratio were then applied to three separate dives to estimate group size. These dives were randomly chosen from visually verified *M. densirostris* sightings with known group size. The goal of the analysis was to estimate the number of animals in the group from clicks detected on the hydrophone array in the vicinity of the sighted animals.

DTag Statistics

To estimate a mean beaked whale click rate, three dives were examined from a DTag deployed on an *M. densirostris* on 23 October 2006 by WHOI at the AUTEK range in the Bahamas. For each dive the total number of clicks detected on the tag from the tagged whale was divided by the tagged whale's vocalization period during the dive to estimate the click rate. These three click rates were then averaged to produce the mean click rate (CR) of 2.75 clicks/sec used for this analysis (Table 3).

Dive	Click Rate (CR)
1	2.66
2	2.87
3	2.72
Mean	2.75

Table 3: Mean click rate calculated from DTag tagged whale vocalizations.

To determine a normalized detection ratio (NDR) for the range hydrophones, a hydrophone sub-array of 7 to 10 hydrophones was first defined in the vicinity of the tagged animal. Detection data from range hydrophones surrounding the tagged animal were examined for the dive vocalization period. The hydrophone with the maximum number of clicks detected during this period was chosen as the center hydrophone for the sub-array. At the time of the analysis the exact position of the tagged animal was not known and the phone with the most detections was chosen as a first order approximation. A circle with a radius of 4.8 km was drawn around this center phone, and all hydrophones falling within the circle defined the hydrophone sub-array used in the analysis. An example of the sub-array designation is shown *Figure 3*. The 4.8 km radius circle was chosen to be consistent with the *M. densirostris* group localization density estimation method presented in [6].

The total number of clicks (C_{ha}) detected on all hydrophones in the sub-array during the dive vocalization period was summed. The total number of clicks produced by the tagged animal during its dive vocalization period and recorded on the DTag was defined as C_t . The tagged animal was sighted in a group of four (two mother-juvenile pairs), though two other animals were sighted at a distance from the group. Using a group size of four, the normalized detection ratio (NDR) for each dive was defined as the total number of clicks detected on the hydrophone sub-array during the tagged animal's dive vocalization period (C_{ha}) divided by the total number of clicks emitted by the tagged animal during this period (C_t) multiplied by the group size (GS) and the number of hydrophones in the sub-array (H_{ha}):

$$NDR = C_{ha} / (C_t * GS * H_{ha})$$

The normalized detection ratios for the three dives were then averaged to produce the mean normalized detection ratio of 0.031 used in this analysis (*Table 4*).

Dive	H_{ha}	C_{ha}	C_t	NDR
1	7	4626	4903	0.034
2	10	5528	4556	0.030
3	10	8639	7564	0.029
Mean NDR				0.031

Table 4: Detection ratio calculated from the DTAG.

Application

To test the efficacy of this click-counting approach for estimating the number of animals in a group, the mean click rate, CR, and normalized detection ratio, NDR, were then applied to data from three separate verified *M. densirostris* sightings. These verified data were collected during focal follows at AUTECH with trained surface observers from the Bahamas Marine Mammal Research Organization (BMMRO). *M. densirostris* groups were detected acoustically on the range. BMMRO observers were directed to a group's location and notified when vocalizations ceased. Within about 20 to 25 minutes the BMMRO

observers visually sighted and counted the number of animals in the group. When the animals went on a terminal dive the observers notified the personnel monitoring the acoustics on-shore, and the group's vocalizations were again detected acoustically within about 5 minutes. The sightings occurred on September 27, 2005 in three different parts of the range, and at different times during the day (*Table 5*, *Figure 3*).

The predicted number of animals and the predicted number of clicks detected on the hydrophone array were then compared to the number of animals sighted and the actual number of clicks detected on the array.

Closest Hyd	Date	Local Time	Group Size
H72	9/27/2005	10:30	2
H76	9/27/2005	11:06	5
H57	9/27/2005	15:25	4

Table 5: BMMRO verified sightings of *M. densirostris* groups at AUTECH on 09/27/2005.

For each sighting, click detection reports from the nearby hydrophones were examined to determine the dive vocalization duration (DVD) for the group, and the hydrophone with the maximum number of clicks detected during this period. This hydrophone was used as the center hydrophone, and a hydrophone array was again defined as all hydrophones that fell within a 4.8 km-radius circle centered on this hydrophone. The total number of clicks, C_{ha} , detected on the array during the dive vocalization period was determined. The number of animals predicted to be in the group, NA_p , was then calculated as the total number of clicks detected on the array divided by the product of the normalized detection ratio (NDR), the mean click rate (CR), the dive vocalization duration (DVD), and the number of hydrophones in the array (H_{ha}):

$$NA_p = C_{ha} / (NDR * CR * DVD * H_{ha})$$

For a known number of animals, NA, the predicted number of clicks detected on the array (C_{hap}) is given by:

$$C_{hap} = NA * NDR * CR * DVD * H_{ha}$$

Table 6 and *Table 7* show the hydrophone arrays used and the associated group dive vocalization durations for the three verified *M. densirostris* sightings. For the sighting at hydrophone 72, the 4.8 km circle defined around 72 only included five hydrophones, since hydrophone 72 is located on the edge of the range. In addition, it appeared that another group was simultaneously vocalizing just north of this array, producing clicks detected by hydrophone 64. To avoid conflict with an adjacent group, hydrophone 64 was removed from the analysis.

The results for the predicted number of animals versus the actual number of animals sighted, and the predicted number of clicks detected on the hydrophone array versus the actual number detected, are shown in *Table 8*.

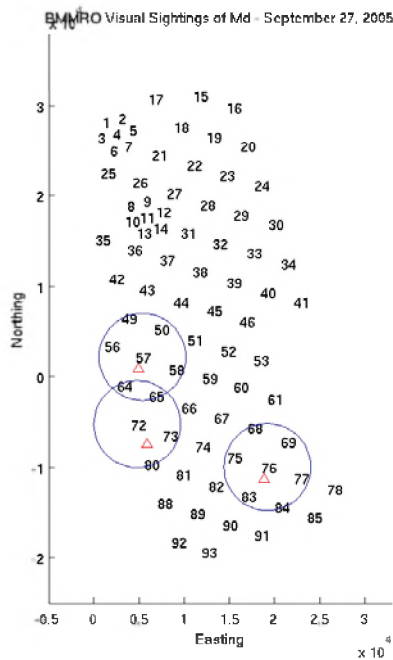


Figure 3: Three visual sightings by BMMRO of *M. densirostris* groups on 09/27/2005 are indicated by triangles. The corresponding 4.8km-radius circles used for the click counting are shown.

Closest Phone	Phones Used	Center Phone	H _{ha}	C _{ha}
H72	65, 72, 73, 80	H72	4	1343
H76	68, 69, 75, 76, 77, 83, 84	H76	7	5439
H57	49, 50, 56, 57, 58, 64, 65	H57	7	4884

Table 6: Analysis hydrophones used for the 3 verified *M. densirostris* sightings, with center hydrophone and total number of clicks detected on the array.

Sighting	Start Time	End Time	DVD (h:m:s)
H72	9:34:40	10:09:30	0:34:50
H76	10:14:51	10:46:15	0:31:24
H57	14:27:19	15:01:20	0:34:01

Table 7: Local start time, end time and dive vocalization durations in hours:minutes:seconds for the 3 verified *M. densirostris* sightings.

	Sighting Hydrophone Location		
	H76	H72	H57
NA _p	4.86	1.89	4.03
NA	5	2	4
C _{had}	5593.28	1418.25	4847.51
C _{ha}	5439	1343	4884

Table 8: Predicted number of animals and clicks detected versus the actual number sighted and detected for the 3 verified *M. densirostris* sightings.

3. DISCUSSION

Vocalizing groups of *M. densirostris* in the TOTO were isolated by measuring and comparing the start and stop times of vocalizations on adjacent, widely distributed (4-km baseline) bottom-mounted hydrophones. The dive vocalization durations for multiple groups of animals were measured. These vocalization periods occur during deep foraging dives [3, 4]. The mean vocalization periods for two separate datasets and gaps in vocal periods for the second dataset were measured.

A generalized method for estimating the number of animals vocalizing in an *M. densirostris* group during deep foraging dives was developed using baseline tag data as follows:

1. Isolate a vocalizing *M. densirostris* group.
2. Estimate the group's dive vocalization duration (DVD) and the hydrophones ensounded by the group.
3. Assign the hydrophone with the most clicks during the DVD as the center phone of the analysis hydrophone sub-array. This sub-array is defined as all hydrophones that fall within 4.8 km of the center phone.
4. Adjust the DVD if necessary using the hydrophones in this analysis sub-array.
5. Sum all the clicks detected during the DVD on the hydrophones in the analysis sub-array.
6. Divide the total number of *M. densirostris* clicks detected on the sub-array during the DVD by the product of the DVD multiplied by the normalized detection ratio, the mean click rate, and the number of hydrophones in the sub-array to estimate the number of animals in the group.
7. Compare the results to the actual number of animals sighted.

This method makes 4 major assumptions:

1. *M. densirostris* vocalizes at a constant rate.
2. The detection ratio is constant.
3. The distribution of hydrophones used to derive the estimating parameters is the same as for the test cases.
4. All detected vocalizations used for the estimate are produced by a single group.

The mean click rate and normalized detection ratio were obtained from one animal tagged on the AUTECH range. These estimates must be verified with data from additional tagged animals, and their variability quantified. Six tags were successfully placed on *M. densirostris* during the Behavioral Response Study at AUTECH 13 August – 27 September, 2007. These tag data will be analyzed and compared to those presented here. The calculation of variance within these data will allow an associated estimate of group size uncertainty. In addition, data from other locations must be quantified and compared, as click rate and normalized detection ratio may be site dependent.

The calculated detection ratio depends directly on the click probability of detection, which depends on environmental conditions. The detection performance is a

function of the receiver performance and the input Signal to Noise Ratio (SNR). The SNR at the receiver is inversely proportional to the sea-state. As the sea-state rises, ocean noise increases, and the SNR at the receiver decreases, which in turn decreases the probability of detection.

M. densirostris spend little time at the surface and are inconspicuous while at the surface [2]. Consequently, sighting these animals is difficult in anything but low sea-states. The estimating parameters (normalized detection ratio and mean click rate) were derived from DTag data. Both the tag data and the verified sighting data were collected in low sea-state (0-1) conditions. To generalize this methodology across varied environmental conditions, the effect of SNR on the detection ratio must be quantified for higher sea-states.

The normalized detection ratio was calculated by totaling the number of clicks detected on hydrophones within a measured area. The number of clicks detected is a function of both the number of hydrophones and their distribution within this area. Thus the detection ratio is dependent upon the given hydrophone distribution, and must be recalculated for new hydrophone configurations. Vocalizations associated with a group of beaked whales are detected on a set of hydrophones. It is assumed that the detected vocalizations used for this analysis were produced by single groups. However, there are cases in which adjacent groups may ensoundify common hydrophones. In one case presented in this paper, detections from a hydrophone were rejected as they were attributed to two adjacent groups. This required manual data analysis. Improved methods for isolating groups should be investigated.

4. CONCLUSIONS

The vocal periods of foraging dives, or dive vocalization durations, for beaked whale groups were measured using passive acoustic detections from the AUTEK range hydrophones. Gaps in the vocal periods were also quantified.

The group size estimation method presented was applied to three dives from visually verified groups of animals with good results. In each case, the method successfully predicted the number of animals in the vocalizing group. The mean error over the three examples was 3%. However, the estimating parameters were derived from a tagged animal in the TOTO, and the method was applied to animals in the same location under similar environmental conditions. Extension of the method must incorporate additional measurement parameters as discussed above.

A passive acoustic density estimation method for *M. densirostris* called group localization was presented in [6]. This method identifies groups of *M. densirostris* on the AUTEK range and uses a mean group size to estimate density. Click counting will be used in place of a mean group size in the group localization method to improve the estimate of *M. densirostris* on the AUTEK range.

It may be possible to estimate animal density directly from click counts using cue-counting methods similar to those used in traditional distance sampling. These methods are currently under investigation.

5. ACKNOWLEDGEMENTS

We would like to acknowledge our sponsor, Dr. Frank Stone at N45, and Jim Eckman at the Office of Naval Research. We would also like to give special thanks to Diane Claridge, Charlotte Dunn, Olivia Patterson, Meagan Dunphy-Daley, and all the members of the Bahamas Marine Mammal Research Organization, along with Todd Pusser and Leigh Hickmott, for their expertise in visual observation of beaked whales and other marine mammals, and their long hours spent on the water. We would also like to acknowledge the NUWC Division Newport Independent Laboratory Innovative Research program manager, Richard Philips, for providing funding to study density estimation methods, and the AUTEK range, especially Marc Ciminello, Tom Szlyk, Jose Arteiro, and Tod Michealis, for their support. We would also like to thank Alex Bocconcelli at WHOI for his continued support during multiple tagging exercises.

6. REFERENCES

- ¹Baird, R. W., D. L. Webster, D. J. McSweeney, A. D. Ligon, and G. S. Schorr, "Diving behavior and ecology of Cuvier's (*Ziphius cavirostris*) and Blainville's beaked whales (*Mesoplodon densirostris*) in Hawaii", Report by Cascadia Research Collective to Southwest Fisheries Science Center, National Marine Fisheries Service under Order No. AB133F-04-RQ-0928, 2005
- ²Claridge, D.E., "Fine-scale distribution and habitat selection of beaked whales," MSc Thesis, University of Aberdeen, Scotland, UK, 2006.
- ³Johnson, M., Madsen, P.T., Zimmer, W.M.X., Aguilar de Soto, N., and Tyack, P.L. **2004**. "Beaked whales echolocate on prey," Proc. R. Soc. London, Ser. B **271**, S383-S386.
- ⁴Johnson, M., Madsen, P.T., Zimmer, W.M.X.Z., Aguilar Soto, N., and Tyack, P.L.T. **2006**. "Foraging Blainville's beaked whales (*Mesoplodon densirostris*) produce distinct click types matched to different phases of echolocation," The Journal of Experimental Biology, 209, 5038-5050.
- ⁵Johnson, M. and Tyack, P. "Measuring the behavior and response to sound of beaked whales using recording tags," National Oceanographic Partnership Program Report: Award Number: OCE-0427577, 2005.
- ⁶Moretti, D., N. DiMarzio, R. Morrissey, J. Ward, S. Jarvis, "Estimating the density of Blainville's beaked whale (*Mesoplodon densirostris*) in the Tongue of the Ocean (TOTO) using passive acoustics," Oceans'06 MTS/IEEE-Boston, Boston, MA. September 18-21, 2006.
- ⁷Morrissey, R.P., Ward, J., DiMarzio, N., Jarvis, S., and Moretti, D.J. **2006**. "Passive acoustic detection and localization of sperm whales (*Physeter macrocephalus*) in the Tongue of the Ocean," Applied Acoustics 67(11-12): 1091-1105.

⁸Tyack, Peter L., Mark Johnson, Natacha Aguilar Soto, Albert Sturlese and Peter T. Madsen. **2006**. "Extreme diving of beaked whales," *The Journal of Experimental Biology*, 209, 4238-4253.

⁹Tyack, P. L., M. P. Johnson, W. M. X. Zimmer, P. T. Madsen, M. A. de Soto. "Acoustic behavior of beaked whales, with implications for acoustic monitoring," *Oceans 2006*, pp 1-6, September 2006.

¹⁰Zimmer, W.M.X., Johnson, M.P., Madsen, P.T., and Tyack, P.L. **2005**. "Echolocation clicks of free-ranging Cuvier's beaked whales (*Ziphius cavirostris*)," *J. Acoust. Soc. Am.* **117**(6), 3919-3927.

**West Caldwell
Calibration
Laboratories, Inc.**
uncompromised calibration
Web site: www.wccl.ca E-mail: info@wccl.ca

Head Office: 1575 State Route 96, Victor, NY 14564
Phone: 585-586-3900 Fax: 585-586-4327
Branch Office: 31 Ready Court, Brampton, ON L6Y 4T4
Phone: 905-595-1107 Fax: 905-595-1108

A SINGLE SOURCE LABORATORY

for Calibration and Repair of Sound, Vibration, and Electronic Test Instrumentation

SPECIALIZING IN:

- Accelerometers
- Microphones
- Sound Level Meters
- Field Calibrators
- Audiometric Equipment
- Vibration Meters
- Frequency Analyzers
- Vibration Test Equipment

OUR AUTOMATED FACILITY ASSURES YOU OF:

Calibrations Traceable to N.I.S.T.

Certification: ISO 9001:2000

Accreditation: ISO/IEC 17025:2005

Compliance: ISO 10012-1, MIL-STD-45662A, ANSI/NCSS 2540-1-1994

Superior Workmanship

Complete Test Documentation

Quick Turnaround time:

- 48 Hour Calibration Service Available for an Additional Fee
- 5-10 Days Standard Turnaround

OTHER SERVICES INCLUDE:

- Custom System Integration



Authorized Calibration and Repair Center for

- Rion
- Ono-Sokki
- Scantek Inc.

We service equipment manufactured by:

- ACO Pacific*
- Brüel & Kjær*
- CEL*
- Dytran*
- Endevco*
- Fluke
- G.R.A.S.*
- Hewlett-Packard
- Larson Davis*
- Metrosonics*
- Norsonic*
- Norwegian Electric*
- PCB*
- Rion*
- Simpson
- Syminex*
- Quest
- and others

FREE INITIAL OR NEXT CALIBRATIONS COMPLIMENTS FROM WCCL

Your cost of the instrument will be manufacturers list price.

* We will be pleased to order any instrument for you from the manufacturers marked with an "**"

Modular Platform

Type 2250's combination of software modules and innovative hardware makes the analyzer a dedicated solution for high-precision measurement tasks, in environmental, occupational and industrial noise application areas.

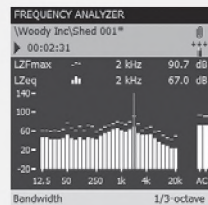
New software modules can easily be added, thus giving you the option of adding more functionality as your measurement requirements change. This way the platform ensures that your investment is securely protected now and in the future.

Currently available applications include:

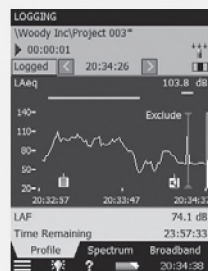
- **Frequency Analysis Software** – for real-time analysis of 1/1- and 1/3-octave bands
- **Logging Software** – log broadband data and spectra at intervals from 1s to 24h
- **Enhanced Logging Software** – for continuous monitoring and logging of periodic reports
- **Sound Recording Option** – record measurements signals to identify and document sound sources

Type 2250 has been designed in co-operation with users to be easy, safe and clever.

For more information please contact your local sales representative or go to www.type2250.com



Frequency Analysis Software



Logging Software



Enhanced Logging Software



Sound Recording Option

HEADQUARTERS: DK-2850 Nærum · Denmark · Telephone: +4545800500
Fax: +4545801405 · www.bksv.com · info@bksv.com

USA: 2815 Colonnades Court · Norcross, GA 30071
Toll free (800) 332-2040 · www.BKHome.com · bkinfo@bksv.com

Hand-held Analyzer *Type 2250*

Brüel & Kjær

MODELING THE EFFECT OF BOAT TRAFFIC ON THE FLUCTUATION OF HUMPBACK WHALE SINGING ACTIVITY IN THE ABROLHOS NATIONAL MARINE PARK, BRAZIL

Renata S. Sousa-Lima and Christopher W. Clark

Bioacoustics Research Program, Laboratory of Ornithology, Cornell University,
159 Sapsucker Woods Road, Ithaca, NY 14850, USA

ABSTRACT

Since the moratorium on whaling, the Brazilian government and local Non-Governmental Organizations (NGOs) have adopted and encouraged a more sustainable use of whales as tourist attractions. Nevertheless, concerns about boat traffic impacts on whale population health have arisen, especially in protected areas such as marine parks. The Abrolhos Marine National Park is the seasonal habitat for the breeding population of humpback whales in the Western South Atlantic. We acoustically monitored 7% of the park area during 26 days using marine autonomous recording units and evaluated the responses of whales to boat traffic by measuring changes in male singing activity. The recorded humpback whale songs were analyzed to locate and count individual singers. We modeled the fluctuation in the number of singers over time in response to: number of acoustic boat events, tide height, lunar phase, hour of the day, the quadratic function of hour of day, day of the season, and presence of light. Generalized linear models were used to fit the singer count data into a Poisson distribution and log link. We found an important negative effect of boat traffic on singing activity. There is evidence that the interaction between phases of the moon and the quadratic function of hour of day also affect singing behavior. Adaptive management should aim at reducing the number of noise events per boat, which can improve the whale watching experience and reduce the impact on male singing behavior.

SOMMAIRE

Depuis le moratoire sur la pêche à la baleine, et dans une optique de tourisme durable, une utilisation plus raisonnée des baleines en tant qu'attraction pour touristes a été adoptée et encouragée par le gouvernement brésilien et les O.N.G. locales. Mais des problèmes concernant les impacts du trafic de bateaux sur la santé des populations de baleines ont toutefois surgi, particulièrement dans des secteurs protégés comme les parcs marins. Le parc national marin d'Abrolhos abrite la population reproductrice des baleines humpback du sud-ouest de l'océan atlantique. Afin d'évaluer les réponses des baleines au trafic de bateaux, nous avons réalisé des enregistrements sonores sur 7% du parc pendant 26 jours en utilisant des unités autonomes marines d'enregistrement. Ces enregistrements nous ont permis de mesurer les changements dans l'activité de chant des mâles. Les sons émis par les baleines humpback ont été détectés et les différents chanteurs ont été localisés et comptés. Nous avons modélisé les fluctuations du nombre de chanteurs au cours du temps en fonction du nombre d'événements acoustiques émis par les bateaux, de l'amplitude des marées, de la phase lunaire, de l'heure, de la fonction quadratique de l'heure, de la date, et de la présence de lumière. Des modèles linéaires généralisés ont été utilisés pour adapter le nombre de chanteurs à une distribution de Poisson et un lien log. Nous avons trouvé un effet négatif important du trafic de bateaux sur l'activité de chant. L'interaction entre la phase lunaire et la fonction quadratique de l'heure semblent également affecter le comportement de chant. Une gestion adaptative devrait viser à réduire le nombre d'événements de bruit émis par les bateaux, qui améliorerait l'expérience d'observation des baleines et réduire l'impact sur le comportement de chant des mâles.

1. INTRODUCTION

There has been a dramatic increase in the appreciation of whales as living beings, with the shift away from exploitation (whaling) to ecotourism (whale-watching). However, whale-watching has been somewhat controversial. There can be costs associated with whale-watching to individuals and populations and this activity has coexisted with whaling in many cultures [1-2]. In Brazil, since the

moratorium on whaling was initiated, whale watching has been considered a more acceptable way to coexist with these large marine mammals. The potential negative effect of whale watching on marine mammals is of special concern in breeding areas, where animals congregate and where boat noise has been shown to be disruptive to vocal behavior related to reproduction [3-4]. In their wintering grounds, male humpback whales produce a conspicuous, long, and patterned sequence of sounds denominated as "song" [5]. Various hypotheses for the function of these complex male

vocal displays have been offered, and the prevailing one is that songs are socially important, and related to reproduction [6-12]. Boat disturbances of singing behavior on breeding grounds may thus affect individual mating success, and have even more far-reaching effects on the long-term viability of populations.

The Abrolhos National Marine Park is the most important breeding site for the southwestern Atlantic population of humpback whales [13]. Local whale-watching activity increased until 1998, and since then, the annual number of park visitors has been 4,000-5,000 people. The boat-based trips to Abrolhos provide substantial income to the small coastal towns from which the tour boats depart [14]. Increasing boat traffic from these coastal towns may put the Abrolhos humpback whale population at risk. Fortunately, as a designated Marine Park, there is the opportunity for better regulation of the potential human disturbances on the whales, while still allowing economically important ecotourism to continue.

Monitoring programs designed to improve management decisions should focus on the fundamentals of the humpback's biology: its population size, distribution, behavior, and mating system. In a marine environment, to disregard acoustic communication between individuals is to ignore their primary sensory modality which enables social interactions, and thus, the mechanisms that strongly influence reproductive success and the mating system. There is limited information about how whale vocal behavior varies through time. Studies on humpback whale singing activity have focused on geographic and temporal changes in song patterns [6, 15-24]. Only a few authors [25-28] have addressed temporal variability in the singing activity of individuals to evaluate effects of disturbances. Studies have shown an effect of noise from human activities on humpback whale singing activity [28-30] and in one case the factor that most influenced ambient noise was the total number of vessels passing within the area per unit time [31].

Here we employed advanced technologies in passive acoustics to: 1) explain the natural fluctuation in humpback singing over a 26-day period and 2) determine how the noise generated by boat traffic affects the variation in singing activity in the Abrolhos National Marine Park, Brazil. Our approach was to create multiple models incorporating variables that we hypothesized might affect singing behavior. A list of the hypotheses is presented. These hypotheses are not mutually exclusive; some of the variables may affect the others, so interaction terms between effects are included in the models. The prediction of each hypothesis is given in parenthesis and the specific variable is in bold.

H1) singing is negatively affected by boat noise (**Boats** has a negative coefficient);

H2) singing is a function of the density of whales and increases as the season progresses, peaks in September and decreases until the end of the season [32] (**Day** is not an important variable and the quadratic function, **Day**², has a negative coefficient but is not included in the models because data collection was concluded in August);

H3) singing increases linearly as the season progresses (**Day** has a positive coefficient);

H4) singing decreases as a function of density (**Day** has a negative coefficient, **Day**² has a positive coefficient);

H5) singing is intensified at night (**Light** has a negative coefficient);

H6) singing is affected by moon cycle, and will decrease during phases with moonlight (**Moon** has a negative coefficient when "Full" and positive when "New");

H7) singing increases with tide height (**Tide** has a positive effect), and;

H8) singing is a function of time of the day (**Hour**² has a positive effect; **Hour** has a negative effect).

2. METHODS

2.1. Data Collection

The Abrolhos Bank is located off the coast of Brazil between 16°40'-19°30'S covering an area of approximately 30,000km² [33]. The Abrolhos Marine National Park includes the Abrolhos archipelago in the northeast portion of the bank [34]. The local humpback whale population has been estimated at around 3,000 individuals [35], representing almost 15% of the total population of humpbacks thought to occur in the Southern Hemisphere [36]. Approximately 7% of the park area was acoustically monitored using an array of pop-ups (marine autonomous recorders developed by the Bioacoustics Research Program of the Cornell Lab of Ornithology - details at www.birds.cornell.edu/brp). Each pop-up carries an onboard clock that makes it possible to perform sound source localization and tracking of signals recorded by an array of synchronized pop-ups. Our array consisted of 4 pop-ups deployed northwest of the Abrolhos archipelago. The 4 pop-ups were programmed to record continuously from 22 July to 16 August, 2003 at a sampling rate of 2kHz.

2.2. Data Processing

The 4-channel sound files were submitted to detection and location algorithms. The detections were identified using 30 to 70 different templates of humpback whale sounds (song units) extracted from the same 24-hour recording using a custom software analysis program, XBAT (xbat.org). This detection procedure was repeated for each day of recording in order to avoid reduced detection probability due to changes in song units that are known to occur through time [18]. All XBAT detections were located using a custom location tool (Cortopassi & Fristrup, unpublished), and the resulting locations of each detected sound were checked by an experienced analyst using a browsing time window of 10 to 30s. This protocol insured that false detections and locations were eliminated and that missed humpback whale sounds were detected and located individually. Detections of boat acoustic events were done manually by an operator listening to the files and drawing a box over any continuous bout of engine sound produced by a single or multiple boats in one channel while browsing each 24-hour sound file with a window of 100-300s. The "Energy Distribution

Measurement” tool [37] was used to obtain the “center-time” of each boat event (*i.e.*, the time at which the median amplitude of the boat’s noise occurred on a given channel).

2.3. Data Sampling and Analyses

We extracted two variables from the sound recordings: 1) the response variable: “Singers” and 2) one of the predictor variables: “Boats”. The number of singers was first counted following each singing male continuously through time. Individual singing bouts (*i.e.*, time spent singing continuously by a single male, $N = 136$) varied from 30 minutes to 20.5 hours, with median duration of 90 minutes. The absolute number of singers was then counted separately in each consecutive 30 minute period as a continuous count of singers. A time series analysis was performed in version 6.2 of the S-PLUS statistical software package (S-PLUS 2003) to determine the time lag between independent samples. The autocorrelation was negligible after six 30-min periods. Thus, counts were done with a lag of 3 hours or more to avoid autocorrelation and to sample all 30-min periods of a day. The number of singers in a 1-minute count is proportional to the number of singers in continuous counts ($N = 47$; Regression through the origin: 1-Minute Count = $0.7850467 \cdot \text{Continuous counts}$; $R^2 = 0.8963$). Given this result, the number of singers within the 5th minute of each 30 minutes sampled ($N = 141$) was used as the singer count to reduce analysis time.

The variable “Boats” was used as a measure of boat noise and is the number of boat events over each 30-min period. A boat event was counted when its center time was within the sampled 30-min period. Whales typically startle when exposed to unexpected, loud, suddenly louder or different sounds, such as a nearby engine starting up. The same sounds may not elicit a reaction if the sound is continuous and predictable, such as engine noise from a distant, approaching boat traveling at a constant speed [4, 38]. To address this observed response to noise events, cases in which the same boat’s engine was turned off and then back on with a silent interval between them were counted as two boat events. The effect of a boat’s source level and distance from the study area were also accounted for by counting every boat event that appeared on each channel of the array, even if coming from the same boat. Therefore, a higher weight was given to the boats that were louder or closer to the array and the nearby whales. The models included other measurable predictor variables that might affect humpback whale singing behavior according to the literature (Table 1). All predictor values were standardized. We excluded interactions between predictors that did not make sense and between correlated variables. Generalized linear models [39] were used to fit the singer count data into a Poisson distribution with log link. All statistical procedures were carried out in version 9.1 of the SAS statistical software package (SAS Institute 2002-2003). A

set of 70 models was explored trying to balance between under- and over-fitted models, with an effort to avoid over-fitting of the relatively small sample size. We used the Akaike information criteria [40], corrected for small sample sizes, AICc [41-42], to choose the best model in the set, *i.e.*, the model that minimizes the information loss about the system, given the data. We also used multimodel inference (model averaging) to reduce bias of the estimates [43]. The relative likelihood of model i versus model j is termed the “evidence ratio”, and a ratio of their Akaike weights (w_i) was used to compare several models and make inferences about the importance of the different predictors [44].

3. RESULTS

The singer counts ($N = 141$) varied from 0-9 (Mean = 2.62, Standard Deviation (SD) = 1.54), while boat counts ranged from 0-19 events (Mean = 2.34, SD = 4.02, Median = 0, Interquartile Range (IQR) = 3). The selected best model (Singers = Boats + Hour + Hour² + Moon + Hour²*Moon) is not convincingly the single best. If the Akaike differences (ΔAICc) are ranked from smaller to larger, the evidence ratio of the best model over each subsequent model decreases gradually, until the models become less plausible to be the best ($\Delta \text{AICc} > 10$). We then selected the models that had some support ($\Delta \text{AICc} < 10$) and included these 39 models in a 99% confidence set to recalculate the Akaike weights (Table 2) [44]. The confidence intervals for the predictors (Table 3) were estimated using the averaged model ($S = B + H + H2 + M + T + L + D + H \cdot M + H \cdot B + H \cdot T + H2 \cdot B + H2 \cdot M + H2 \cdot L + H2 \cdot T + M \cdot T$).

Table 1: Descriptions of model independent variables.

Variable	Description
Boats (B)	Integer value of the number of boat acoustic events within each sampled 30-minute period.
Day (D)	Continuous variable counted from the first day of the local humpback whale season.
Hour (H)	Continuous variable calculated based on the first hour of the first day of the humpback whale season, in half hour increments.
Hour² (H2)	Square value of Hour based on the quadratic fit of the averaged count by time of day.
Moon (M)	Phase categories (4 levels) based on NOAA Astronomical online data.
Light (L)	Binary variable based on rise and set times of the sun in Abrolhos (U.S. Naval Observatory online database).
Tide (T)	Height of tide at the end of the half hour period averaged from the hourly local values.

Table 2: Model selection results for the 39 most plausible models ($\Delta AICc < 10$). AICc values were scaled (adding the number 8) to avoid negative numbers. K = number of parameters. The best model is highlighted in bold in the table below.

Model	#	LnLikelihood	K	AICc	$\Delta AICc$	w_i	w_{21}/w_i
S = B H H2 M H2*M	21	14.4255	10	0.8412	0.0000	0.1805	1.00
S = B H H2 M T H2*M	57	14.9879	11	2.0708	1.2295	0.0976	1.85
S = B H H2 M L H2*M	68	14.6747	11	2.6971	1.8558	0.0714	2.53
S = B H H2 M H2*M H2*B	67	14.5059	11	3.0347	2.1935	0.0603	2.99
S = B H H2 M H*B H2*M	35	14.4644	11	3.1178	2.2765	0.0578	3.12
S = B H H2 M D H2*M	65	14.4276	11	3.1913	2.3501	0.0557	3.24
S = B H	42	5.4801	3	3.2149	2.3737	0.0551	3.28
S = B H H2 M L H2*M H2*L	70	15.5522	12	3.3331	2.4919	0.0519	3.48
S = B H H2 M H*M	61	13.0015	10	3.6892	2.8480	0.0434	4.15
S = B H H2 M T H2*M H2*T	69	15.0896	12	4.2583	3.4170	0.0327	5.52
S = B H H2 M T H2*M H2*B	55	15.0808	12	4.2759	3.4347	0.0324	5.57
S = B H H2 M L H2*M H2*B	51	14.7692	12	4.8991	4.0578	0.0237	7.61
S = B H H2 M T H2*M M*T	59	17.1765	14	4.9803	4.1391	0.0228	7.92
S = B H H2	31	5.6030	4	5.0881	4.2469	0.0216	8.36
S = B H H*B	43	5.4884	4	5.3172	4.4760	0.0193	9.37
S = B H H2 L	32	6.3518	5	5.7408	4.8995	0.0156	11.59
S = B M	44	6.3328	5	5.7788	4.9375	0.0153	11.81
S = B H M	41	7.3690	6	5.8888	5.0476	0.0145	12.48
S = B H H2 T	33	6.0710	5	6.3024	5.4611	0.0118	15.34
S = B H H2 M T L H2*M H2*B	53	15.2777	13	6.3108	5.4695	0.0117	15.41
S = B H H2 M H*M H2*M	66	15.2262	13	6.4138	5.5725	0.0111	16.22
S = B H H2 M	30	8.1838	7	6.4745	5.6332	0.0108	16.72
S = B M L	38	6.9107	6	6.8054	5.9641	0.0091	19.73
S = B H H2 M T H2*M H2*B M*T	54	17.4739	15	6.8923	6.0510	0.0088	20.60
S = B M T	62	6.7471	6	7.1327	6.2914	0.0078	23.24
S = B H H2 D	11	5.6376	5	7.1693	6.3281	0.0076	23.67
S = B H H2 M T	34	8.7642	8	7.5624	6.7212	0.0063	28.81
S = B H H2 M L	29	8.6338	8	7.8233	6.9820	0.0055	32.82
S = B H H2 L D	10	6.3757	6	7.8755	7.0342	0.0054	33.69
S = B H M H*B	40	7.3795	7	8.0831	7.2418	0.0048	37.37
S = B H H2 T H2*T	60	6.2251	6	8.1767	7.3354	0.0046	39.16
S = B H H2 M H2*B	39	8.2422	8	8.6065	7.7653	0.0037	48.55
S = B H H2 M H*B	36	8.2248	8	8.6414	7.8002	0.0037	49.41
S = B H H2 M T L H*B H*T H2*M H2*L	24	16.4777	15	8.8846	8.0433	0.0032	55.79
S = B H H2 M T L H2*M H2*B M*T	52	17.6475	16	9.0920	8.2508	0.0029	61.89
S = B H H2 M T L	28	9.1323	9	9.1094	8.2682	0.0029	62.43
S = B M T M*T	63	9.0888	9	9.1965	8.3552	0.0028	65.21
S = B H H2 M T H2*B	56	8.8289	9	9.7163	8.8751	0.0021	84.57
S = B H H2 M T H*B H*M M*T	64	15.9764	15	9.8872	9.0459	0.0020	92.11

Table 3: Parameter 95% confidence intervals (CIs = coefficient estimate \pm 1.96* Standard Error (SE)) for the predictors of the averaged model.

Parameter	Lower Limit	Upper Limit
INTERCEPT	-2.4403	1.3721
DAY	-0.0283	0.1901
HOUR	-1.2365	10.5929
HOUR ²	-7.5842	1.1652
BOATS	-0.3743	-0.0954
LIGHT		
1	-0.0701	0.1121
0	0	0
MOON		
Full	-3.9447	0.8133
Last quarter	-22.5943	2.9131
New	-0.6259	0.9612
First quarter	0	0
HOUR ² *MOON		
Full	-0.7800	2.673
Last quarter	-30.8003	4.7036
New	-4.6914	1.2647
First quarter	0	0
TIDE	-0.0446	0.0620
HOUR*BOATS	-0.0177	0.0206
HOUR*TIDE	-0.0007	0.0005
HOUR ² *LIGHT		
1	-0.0405	0.0250
0	0	0
TIDE*MOON		
Full	-0.0229	0.0366
Last quarter	-0.0425	0.0707
New	-0.0155	0.0215
First quarter	0	0
HOUR ² *BOATS		
Full	-0.0312	0.0428
HOUR ² *TIDE		
Full	-0.0061	-0.0045
HOUR*MOON		
Full	-1.2013	1.7672
Last quarter	-2.9588	1.7631
New	-1.8024	1.0836
First quarter	0	0

“Boats” (B) has a negative effect and is undoubtedly the most important predictor of variation in number of singers (CI does not include zero). All the models that are plausible to be the best in the set include B as a predictor. Additionally, model 21 is 504.47 times more likely to be the best model in the set than a model which differs from model 21 only by the lack of B (not included in Table 2). Therefore, there is strong evidence to conclude that singing is negatively affected by boat noise.

“Hour” and “Hour²” are important predictors for a good model. Although their coefficient CIs may include zero, the majority of the plausible models include both predictors. Nevertheless, evidence ratios between models that exclude and include H and H2 ($w_{44}/w_{30} = 1.42$; $w_{38}/w_{29} = 1.65$) show that models that include both these predictors are less likely to be best. Therefore, there is not enough evidence to support H8: humpback whale singing behavior is affected by time.

The prediction was that singing activity would decrease

during midday but the estimated coefficient of H2 tends to negative values, contrary to the prediction. Only when we investigated the importance of the interaction between Hour² and Moon (H2*M) that the importance of H and H2 became clear. The inclusion of H2*M makes model 21 16.72 times more likely to be the best if compared to model 30. There is strong evidence that a change in the effect of “Moon” changes the effect “Hour²” on singing behavior, *i.e.*, some phases of the moon affect the temporal pattern of singing activity more than others. The main effects (H2 and M) must be kept in the model if the interaction is important. Also, H2 is a function of H, and the same rule applies. Similarly, “Moon” needs to be in the model. The change from first to last quarter has a strong negative effect on singing. Light level is unlikely to play a role given that both phases have the same percentage moon illumination. Evidence ratios ($w_{42}/w_{41} = 3.8$; $w_{31}/w_{30} = 2$) show that models that do not include “Moon” are slightly more likely to be best. Therefore the support for H6 is weak and the importance of “Moon” might also be due to the importance of the H2*M interaction. “Light” and “Day” are less important predictors. Their coefficient CIs include zero, and evidence ratios indicate that there is considerably less support for hypotheses 2-5.

The inclusion of “Tide” improves the fit of the model, and the evidence ratio indicates that model 57 is also likely to be the best in the set ($w_{21}/w_{57} = 1.85$). Nevertheless, the coefficient CI for this predictor includes zero and the estimated magnitude of its effect is very small. Inasmuch, there is little evidence to support H7, given the data. The interaction between “Hour²” and “Tide” (H2*T) seems to have an important negative effect on singing based on its CI (Table 3) but the inclusion of this predictor in the model is not as important as the inclusion of H2*M ($w_{21}/w_{60} = 39.16$), and the fit of the model that includes H2*T instead of H2*M is not very good. It is plausible that changes in tide height change the temporal pattern of singing activity, but the effect is small, and not necessarily important to explain most data variation. All the other predictors (H*M, H*B, H*T, H2*B, H2*L, and M*T) have similarly small effects and are likely unimportant variables (all coefficient CIs include zero).

4. DISCUSSION AND CONCLUSIONS

An increase in the number of boat acoustic events negatively affects whale singing activity. Although masking makes song more difficult to detect, in our analyses the counts were made on a single minute during the 30-min period, only when boat noise was not enough to mask whale signals. The mechanism of this negative effect could, then, be: 1) male humpback whales are displaced and move outside of the location range, 2) males quit singing, or 3) a combination of 1 and 2. Clark & Altman [45] showed a decrease in the detection probability of fin whale sounds during transmissions of LFA U.S. NAVY sonar due to the same 2 possible phenomena. Variation in sound propagation can result in different radii of boat noise influence [46]. In

fact, short-term avoidance responses to ships and boats were observed at distances ranging from less than 30 m to more than 4 km for different studies [3]. Additionally, area avoidance by whales exposed to noise can last for 20 minutes to several days [3]. Long-term (almost two decades) avoidance of areas during periods of increased commercial shipping (and associated dredging activities) has been suggested for gray whales [47]. Permanent avoidance has also been hypothesized for gray whales in San Diego Bay [48], although the direct link to the whales' displacement is controversial [3]. It has been shown that whale density is inversely related to number of boats in an area [49], and our results showed a similar trend for the absolute number of vocally active male humpback whales.

It has been proposed that a higher level of singing activity at night may indicate that the male vocal display might be favored as a mating tactic in the absence of light. The assumption is that light and vision are important for competitive group formation, so that males engage in fighting as a primary mating tactic during the day, as opposite to solo singing at night [26, Cholewiak et al., unpublished]. In contrast, we found that "Light" was unimportant in explaining the fluctuations in singing behavior once we controlled for the other covariates. There was little evidence for an influence of "Tide", and if there is such effect, it is small. After we controlled for the effect of boats, the only temporal effect detected was the one linked to the changes in moon phase, *i.e.*, the temporal trend in singing activity is altered by the phase of the moon. The negative coefficient estimate of "Hour²" might then be an artifact of the small sample size for each half hour period, which ranged from 1 to 6 samples per period. This might have prevented us from detecting a real temporal trend found by others [26, Cholewiak et al., unpublished]. The most likely explanation for the decrease in singing activity during midday observed in the raw data is that it is a reflection of the effect of "Boats", which has a negative coefficient and increases during midday. Overall singing activity decreases as boat noise increases and the remaining temporal trend found is probably related to a cyclical biorhythm, influenced by the moon's phase, and maybe tidal cycle rather than related to light.

The sustainability of the whale-watching tourism industry depends on: 1) maintaining visitor numbers close to the carrying capacity of the whale watching fleet; 2) the local and regional fluctuations in the economy, and most importantly; 3) the maintenance of the resources on which the tourism relies [2]. If whales are being disturbed resulting in them moving out of the area, then current levels of whale watching activity might not be sustainable. Actions to make this human activity less distressing to whales should be implemented. Acoustic isolation of engines, scientifically-validated approach protocols, and reinforcement of regulations of numbers and speeds of boats in areas used by marine mammals are sensible measures that should be applied. Adaptive management [50] should aim at reducing the number of noise events per boat, which can both

improve the whale watching experience and reduce the impact on singing behavior for a given number of boats. It is also important to address the need for enforcement of existing management guidelines, which clearly depends on political will and better prioritization of governmental resources.

5. ACKNOWLEDGEMENTS

We thank: Stephen Morreale, Milo Richmond, John Hermanson, and two anonymous reviewers for their constructive suggestions; Guilherme Lessa and Maria E. Morete for providing data on covariates; Alexandre Paro, Carlo D'Angelo, Edson Patricio and the Tomara crew for field assistance; Andrew Schwarz, Christopher Ko, Justin Tsai, Kunal Jain, Michelle Mathios, Miguel Mayol, Susan Earle, and Roman Lesko for assistance in data processing; and Wesley Hochachka, Russ Loyd, Robert Strawderman, and David Anderson for statistical advice. Funding and support were provided by the Brazilian Government (CAPES PhD scholarship to RSSL), Cornell University's Graduate School, International Students and Scholars Office, Laboratory of Ornithology and Field of Zoology; the New York Fish & Wildlife Cooperative Unit; the Animal Behavior Society's Cetacean Behavior and Conservation Award; the Instituto Baleia Jubarte / Petrobras; and The Canon National Parks Science Scholars Program. We thank Mark Wilson and Aurélie Coulon for translating the abstract into French.

6. REFERENCES

1. Hoyt, E. 2001. *Whale watching 2001: worldwide tourism numbers, expenditures and expanding socioeconomic benefits*. IFAW, Yarmouth Port.
2. Higham, J.E.S. & Lusseau, D. 2007. Urgent Need for Empirical Research into Whaling and Whale Watching, *Conserv. Biol.* 21:554-558.
3. Richardson, W.J., Greene Jr., C.R., Malme, C.I. & Thomson, D.H. 1995. *Marine Mammals and Noise*. Academic Press, San Diego, 576 pp.
4. Sousa-Lima, R.S., Morete, M.E., Fortes, R.C., Freitas, A.C. & Engel, M.H. 2002. Impact of boats on the vocal behavior of humpback whales off Brazil. *J. Acoust. Soc. Am.* 112:2430-2431.
5. Payne, R.S. & McVay, S. 1971. Songs of humpback whales. *Science* 173: 585-597.
6. Winn, H.E. & Winn, L.K. 1978. The song of the humpback whale, *Megaptera novaeangliae*, in the West Indies. *Mar. Biol.* 47: 97-114.
7. Tyack, P. 1981. Interactions between singing Hawaiian humpback whales and conspecifics nearby. *Behav. Ecol. Sociobiol.* 8:105-116.
8. Baker, C.S. & Herman, L.M. 1984. Seasonal contrasts in the

- social behavior of the humpback whale. *Cetus* 5:14-16.
9. Helweg, D.A., Frankel, A.S., Mobley, J.R. & Herman, L.M. 1992. Humpback whale song: our current understanding. In: *Marine mammal sensory systems* (Thomas, J. et al., ed.). Plenum Press, New York, p. 459-483.
 10. Frankel, A.S., Clark, C.W., Herman, L.M. & Gabriele, C. 1995. Spatial distribution, habitat utilization, and social interactions of humpback whales, *Megaptera novaeangliae*, off Hawaii determined using acoustic and visual techniques. *Can. J. Zool.* 73:1134-1146.
 11. Darling, J.D. & Bérubé, M. 2001. Interactions of singing humpback whales with other males. *Mar. Mammal Sci.* 17:570-584.
 12. Darling, J.D., Jones, M. E. & Nicklin, C.P. 2006. Humpback whale songs: Do they organize males during the breeding season? *Behaviour* 143:1051-1101.
 13. Engel, M.H. 1996. Comportamento reprodutivo da baleia jubarte (*Megaptera novaeangliae*) em Abrolhos. *Anais de Etologia* 14: 275-284
 14. Palazzo, Jr., J.T., M. Kammers, and I. Linhares. 1994. Whalewatching sites in Brazil: a summary of available information. IWC/46/WW Working paper, 46th IWC, 8 pp.
 15. Payne, R. 1978. Behavior and vocalizations of humpback whales (*Megaptera* sp.). In: *Report on a workshop on problems related to humpback whales (Megaptera novaeangliae) in Hawaii* (Ed. by K. S. Norris & R. R. Reeves), pp.56-78. U. S. Dep. Commer. NTIS PB-280 794.
 16. Winn, H.E., Thompson, T.J., Cummings, W.C., Hain, J., Hudnall, J., Hays, H. & Steiner, W.W. 1981. Song of the humpback whale: population comparisons. *Behav. Ecol. Sociobiol.* 8:41-46.
 17. Payne, R. & Guinee, L.N. 1983. Humpback whale (*Megaptera novaeangliae*) songs as an indicator of "stocks". In: *Communication and behavior of whales: AAAS Selected Symposium 76* (Ed. by R. Payne), pp. 333-358. Boulder, CO: Westview Press.
 18. Payne, K., Tyack, P. & Payne, R. 1983. Progressive changes in the song of humpback whales songs (*Megaptera novaeangliae*): A detailed analysis of two seasons in Hawaii. In: *Communication and behavior of whales: AAAS Selected Symposium 76* (Ed. by R. Payne), pp. 9-57. Boulder, CO: Westview Press.
 19. Payne, K. & Payne, R. 1985. Large Scale Changes over 19 Years in Songs of Humpback Whales in Bermuda. *Z. Tierpsychol.*, 68: 89-114.
 20. Matilla, D.K. Guinee, L.N. & Mayo, C.A. 1987. Humpback whale songs on a North Atlantic feeding ground. *J. Mamm.* 68:880-883.
 21. Noad, M.J., Cato, D.H., Bryden, M.M., Jenner, M.N. & Jenner, C S. 2000. Cultural revolution in whale songs. *Nature* 408:537.
 22. Cerchio, S. Jacobsen, J.K. & Norris, T.F. 2001. Temporal and geographical variation in songs of humpback whales, *Megaptera novaeangliae*: synchronous change in Hawaiian and Mexican breeding assemblages. *Anim. Behav.* 62:313-329.
 23. Darling, J.D. & Sousa-Lima, R.S. 2005. Songs indicate interaction between humpback whale (*Megaptera novaeangliae*) populations in the Western and Eastern South Atlantic Ocean. *Mar. Mamm. Sci.* 21:557-566.
 24. Eriksen, N., Millar, L.A., Tougaard, J., & Helweg, D.A. 2005. Cultural change in the songs of humpback whales (*Megaptera novaeangliae*) from Tonga. *Behaviour* 142: 305-328.
 25. Helweg, D.A. & Herman, L.M. 1994. Diurnal patterns of behavior and group membership of humpback whales (*Megaptera novaeangliae*) wintering in Hawaiian waters. *Ethology* 98:298-311.
 26. Au, W.W.I., Mobley, J., Burgess, W.C. & Lammers, M.O. 2000. Seasonal and diurnal trends of chorusing humpback whales wintering in water off Western Maui. *Mar. Mamm. Sci.* 16:530-544.
 27. Charif, R., Clapham, P.J., Gagnon, W., Loveday, P. & Clark, C.W. 2001. Acoustic detections of singing humpback whales in the waters of the British Isles. *Mar. Mammal Sci.* 17:751-768.
 28. Frstrup, K.M, Hatch, L.T. & Clark, C.W. 2003. Variation in humpback whale (*Megaptera novaeangliae*) song length in relation to low-frequency sound broadcasts. *J. Acoust. Soc. Am.* 113:3411-3424.
 29. Norris, T.F. 1995. Effects of boat noise on the singing behavior of humpback whales (*Megaptera novaeangliae*). Master Thesis. Department of Moos Landing Marine Laboratories, San Jose State University, 69 pp.
 30. Miller, P.J.O., Biassoni, N., Samuels, A., & Tyack, P.L. 2000. Whale songs lengthen in response to sonar. *Nature* 405: 903.
 31. Haviland-Howell, G., Frankel, A.S., Powell, C.M., Bocconcelli, A., Herman, R.L. & Sayigh, L.S. 2007. Recreational boating traffic: A chronic source of anthropogenic noise in the Wilmington, North Carolina Intracoastal Waterway. *J. Acoust. Soc. Am.* 122:151-160.
 32. Morete, M.E., Pace III, R.M., Martins, C.C.A., Freitas, A.C. & Engel, M.H. 2003. Indexing seasonal abundance of humpback whales around Abrolhos archipelago, Bahia, Brazil. *LAJAM* 2:21-28.
 33. Fainstein, R. & Summerhayes, C.P. 1982. Structure and origin of marginal banks off Eastern Brazil. *Mar. Geol.* 46:199-215.
 34. IBAMA/FUNATURA 1991. *Plano de Manejo: Parque Nacional Marinho dos Abrolhos. Brasilia, Brazil.*
 35. Freitas, A.C., Kinas, P.G., Martins, C.C.A. & Engel, M.H. 2004. Abundance of humpback whales on the Abrolhos Bank wintering ground, Brazil. *J. Cetacean Res. Manage.* 6:225-230.
 36. Klinowska, M. 1991. *Dolphins, Porpoises and whales of the world: The IUCN Red Data Book.* Gland, Switzerland: IUCN.

37. Cortopassi, K. A. 2006. <http://www.birds.cornell.edu/brp/research/algorithm/automated-and-robust-measurement-of-signal-features>.
38. Watkins, W.A. 1986. Whale reactions to human activities in Cape Cod waters. *Mar. Mamm. Sci.* 2:251-262.
39. McCullagh, P. & Nelder, J.A. 1989. *Generalized Linear Models*. 2nd Ed. Chapman & Hall, New York, NY.
40. Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. In: *Second International Symposium on Information Theory*. (Eds B. N. Petrov and F. Csaki.) pp. 267–281. Akademiai Kiado: Budapest.
41. Sugiura, N. 1978. Further analysis of the data by Akaike's information criterion and the finite corrections. *Comm. Stat., Theory and Methods* A7:13-26.
42. Hurvich, C.M. & Tsai, C-L. 1989. Regression and time series model selection in small samples. *Biometrika* 76:297-307.
43. Buckland, S.T., Burnham, K.P. & Augustin, N.H. 1997. Model selection: an integral part of inference. *Biometrics* 53:603–618.
44. Burnham, K.P. & Anderson, D.R. 1998. *Model Selection and Inference: a Practical Information-Theoretic Approach*. Springer-Verlag: New York, 488 pp.
45. Clark, C.W. & Altman, N.S. 2006. Acoustic Detections of Blue Whale (*Balaenoptera musculus*) and Fin Whale (*Balaenoptera physalus*) Sounds During a SURTASS LFA Exercise. *IEEE J. Oceanic Engineering* 31:120-128.
46. Watkins, W.A. & Goebel, C.A. 1984. Sonar observations explain behaviors noted during boat maneuvers for radio tagging of humpback whales (*Megaptera novaeangliae*) in the Glacier Bay area. *Cetology* 48:1-8.
47. Bryant, P.J., Lafferty, C.M. & Lafferty, S.K. 1984. Reoccupation of Laguna Guerrero Negro, Baja California, Mexico, by gray whales. P. 375-387 In: Jones, M.L., Swartz, S.L. & Leatherwood, S. (Eds.). *The gray whale, Eschrichtius robustus*. Academic Press, Orlando, FL. 600 pp.
48. Rice, D.W. & Wolman, A.A. 1971. *The life history and ecology of the gray whale (Eschrichtius robustus)*. *Am. Soc. Mammal., Spec. Publ.* 3, 142 pp.
49. Baker, C.S. & Herman, L.M. 1989. Behavioral responses of summering humpback whales to vessel traffic: experimental and opportunistic observations, *Final Rep. No. NPS-NR-TRS-89-01*. United States Department of the Interior, National Park Service, Anchorage, Alaska.
50. Blumstein, D.T. 2007. Darwinian Decision Making: Putting the Adaptive into Adaptive Management. *Conserv. Bio.* 21:552-553.



Photo Credit: Bahamas Marine Mammal Research Organisation



Photo Credit: Bahamas Marine Mammal Research Organisation

EXCERPTS FROM “ACOUSTICS IN THE NEWS”, IN ECHOS, ASA

Twenty-two varieties of beaked whales roam the seas, diving as deep as a mile to feed on bottom-dwelling squid and small fish on the dark ocean floor. According to a story in the October 15 issue of the *Washington Post*, the realization that sonar can disorient or frighten whales sufficiently to leave them beached and dying has spurred protests and lawsuits. The Navy first denied but now acknowledges the problem, but it has resisted efforts to limit testing of their sonar, saying it is essential to national security. The Navy has now funded a \$6 million project to learn more about beaked whales and their response to sonar and loud ocean noises. The goal is to learn more about beaked whales by attaching sophisticated motion detectors to record the timing, depth and angles of their dives and ascents to see how the animals react when exposed to sounds approaching the intensity of sonar signals. Beaked whales can dive for periods as long as 85 minutes.

Test sections of asphalt rubber in the Seattle area are drawing favorable comments, according to a story in the December 17 issue of *The Seattle Times*. Recent tests show older asphalt registers about 105 decibels when measured with a microphone on a rear wheel of a vehicle about 2 inches above the pavement. Brand-new conventional asphalt registers about 100 decibels, while new rubberized asphalt tends to be about 95 to 96 decibels. Pound for pound, asphalt rubber and polymer asphalt are more expensive than conventional asphalt, but since they're placed at half the thickness, they end up costing about the same. However, the life span of asphalt rubber tends to be several years shorter. Since 1988, the Arizona Department of Transportation has used asphalt rubber in more than 3,000 miles of pavement overlays. Arizona now recycles 70 percent of its used tires back into the highways, eating up about 1,500 tires per lane mile of highway.

Quiet hotels were the subject of two articles in *The New York Times*. Although luxury hotels have often made efforts to “soundproof” their rooms, an article in the October 21 issue describes efforts by AmericInn, a mid-range hotel chain, to reduce room noise by using masonry blocks filled with sound-absorbing foam, in addition to drywall that is 5/8-inch thick instead of 1/2-inch. It also installs gaskets and door sweeps to minimize hallway noise and obtain a Sound Transmission Class test of 50 or higher. The Fairmont Vancouver Airport hotel recently created a “quiet zone” on its sixth floor for daytime sleepers. Loews Hotels have been offering guests free sound-masking machines that emit white noise for light-sleeping guests. An article in the October 2 issue cites other examples of construction with double-glazed windows and insulated walls. Older luxury hotels often were built “like the Maginot Line,” with enormous thick walls but when hotels add plumbing or wiring to such a structure they have a temporary noise problem. Some hotels, especially in the luxury market, deliberately encourage the kind of bustle and excitement in lobbies and bars that can lead to noise seeping into guest rooms. One hotel was recently built with 8-inch thick walls between rooms.

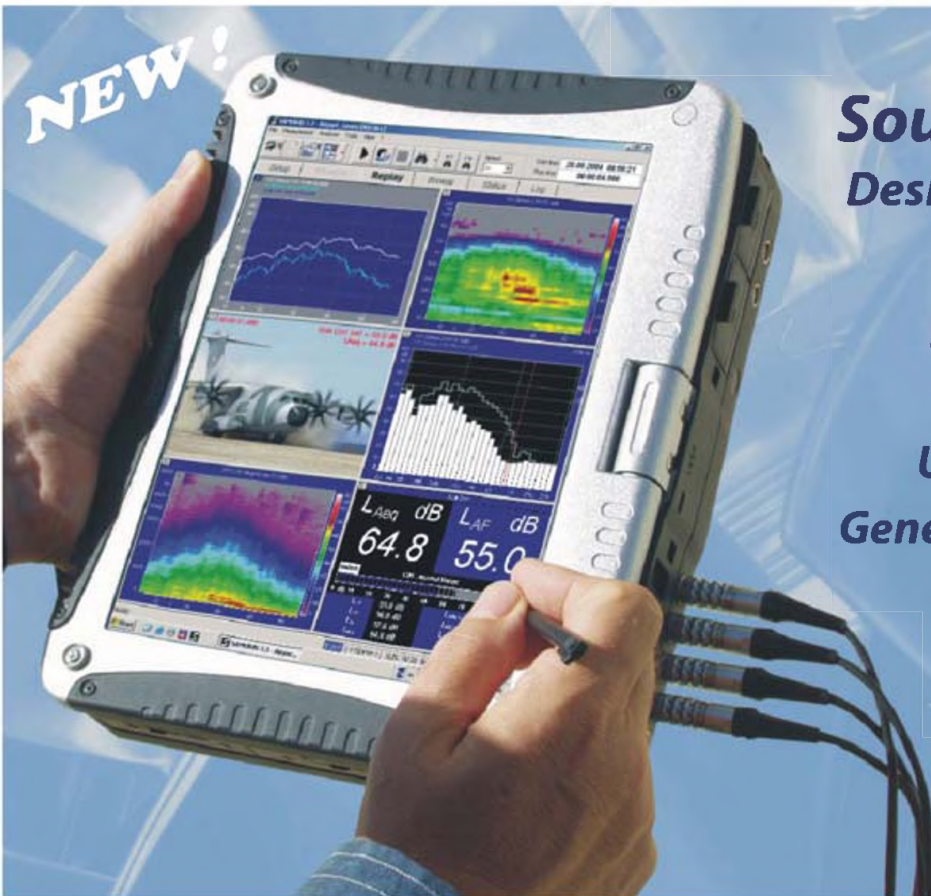
A federal judge limited the Navy's ability to use mid-frequency sonar on a training range off the Southern California coast, according to a story in the January 4 issue of the *Washington Post*. The court ruled that the loud sounds would harm whales and other marine mammals if not tightly controlled. The order banned the use of sonar within 12 nautical miles of the coast and expanded from 1100 yards to 2200 yards the “shut down” zone in which sonar must be turned off whenever a marine mammal is spotted. The judge also forbade sonar use in the Catalina Basin, an area with many marine mammals. The decision is a blow to the Navy, which has argued that it needs the flexibility to train its sonar operators without undue restrictions.

NEW!

Soundbook™

Designed for You:

- Innovative ✓
- IEC conform ✓
- Inexpensive ✓
- User friendly ✓
- General purpose ✓
- Tough (MIL) ✓
- Reliable ✓



SOUNDBOOK

Multichannel SLM IEC61672-1, IEC60804 & IEC60651 Type 1

SAMURAI Basic Software includes Sound Recorder, Frequency Analyzer, Reverberation time measurement

2/4/8/channels with 40kHz bandwidth, 2x tacho, 5 AUX and 2/4 analog outputs

Various software options for Acoustics and Vibration

Remote Control, Network Integration & wireless synchronization of several devices possible

Alternative packages for ME'scope(direct device), si++workbench, SINUS MATLAB Toolbox

PTB Type Approval

SINUS

G.R.A.S.
SOUND & VIBRATION

LDS
DACTRON

01dB-Stell
IMV Technologies group

BYTRAN
BYTRAN SYSTEMS, INC.

MetroLaser, Inc. *

Vibrant Technology, Inc.

Integrated Solutions from World Leaders

- Precision Measurement Microphones
- Intensity Probes
- Outdoor Microphones
- Hydrophones
- Ear Simulation Devices
- Speech Simulation Devices
- Calibrators
- Array Microphones
- Sound Quality
- Sound Intensity
- Sound Power
- Room Acoustics
- Noise Monitoring
- Dynamic Signal Analyzers
- Multi Channel Dynamic Analyzer/Recorders
- Electro Dynamic Shaker Systems
- Advanced Sound & Vibration Level Meters
- Doppler Laser Optical Transducers (Laser Vibrometers)



New KEMAR Manikin



Ottawa

613-598-0026

info@noveldynamics.com

NOVEL DYNAMICS INC.
Dynamic Test and Analysis Systems



Toronto

519-853-4495

Ametelka@cogeco.ca

OBITUARY / NÉCROLOGIE



PROF. JOHN E. K. FOREMAN (1922 - 2007)

Compiled by Elizabeth Wells (Prof. Foreman's daughter)

Professor John E. K. Foreman passed away in his 85th year on Aug 7, 2007. He taught at the University of Toronto, Cornell University in USA and Cambridge University in England and joined the University of Western Ontario in 1956, starting up the Mechanical Engineering department. He was Chairman of the Mechanical Group for 12 years and head of the Sound and Vibrations Laboratory from 1972 until his retirement in 1987, when he was awarded Professor Emeritus. John was instrumental in equipping a Sound and Vibrations Mobile Laboratory Van which was used for many research projects. John loved to teach and emphasized the importance of simply doing your best. He was very proud of his students' efforts in research e.g. developing a battery powered car in the 1970's. He wrote a textbook, entitled "Sound Analysis and Noise Control" for use in classes and as a reference book in the acoustical field. His research projects included bio-medical research into fetal monitoring, research for the Ontario Ministry of Transport on high voltage transmission lines, and testing attitudinal responses to these noises. He also acted as a consultant for building acoustics (e.g. Alumni Hall), and industrial and general community noise. John was a member of the Canadian Acoustical Association for many years, contributing several articles to the publication. He was married for 60 years, and leaves behind his widow, Janet Foreman, 4 children and 4 grandchildren. He made a

difference in his life and will be missed by many whose lives he touched.

Prof. Ramani Ramakrishnan adds:

Prof. Foreman has been involved with the Canadian Acoustical Association from its inception. He chaired the CAA meetings in London (1970) and in Ottawa (1971). He also served as the President of the association from 1971 to 1972. Prof. Foreman's guest article was the main feature of the first issue of the CAA newsletter, edited by Dr. Tony Embleton. The first issue was released in 1973 during the 12th meeting of CAA held in Ottawa. He was also part of the group that coined the name for the association, "Canadian Acoustical Association/Association canadienne de l'acoustique." On a personal note, I found John to be charming and eager in sharing of his knowledge with others. In that sense, he was a true academic. He was still continuing his research during the last few years of his life, in spite of his failing health. He wrote two papers dealing with the impact of noise and vibration on plant growth and seed germination. A friendly correspondance ensued between John Foreman and me, as the Editor-in-Chief of the Canadian Acoustics Journal during the publication of his articles. I was happy to have known him during his last years.

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>).

Next deadline: Shaw, Bell, Fessenden, Eckel and Héту Prizes: 30 April 2008
Prochaine échéance: Prix Shaw, Bell, Fessenden, Eckel et Héту: 30 Avril 2008

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 HEARING • PRIX ÉTUDIANT ALEXANDRE GRAHAM BELL EN COMMUNICATION VERBALE ET AUDITION

\$800 for a graduate student enrolled at a Canadian institution and conducting research in the field of speech communication or hearing • \$800 à un(e) étudiant(e) inscrit(e) au 2e ou 3e cycle universitaire dans une institution canadienne et menant un projet de recherche en communication verbale ou en audition.

FESSENDEN GRADUATE STUDENT PRIZE IN UNDERWATER ACOUSTICS • PRIX ÉTUDIANT FESSENDEN EN ACOUSTIQUE SOUS-MARINE

\$500 for a graduate student enrolled at a Canadian 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 universitaire dans une institution 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 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 universitaire dans une institution canadienne et menant un projet de recherche relié à l'avancement des pratiques en contrôle du bruit.

RAYMOND HÉTU UNDERGRADUATE PRIZE IN ACOUSTICS • PRIX ÉTUDIANT RAYMOND HÉTU EN ACOUSTIQUE

One book in acoustics of a maximum value of \$150 and a one-year subscription to *Canadian Acoustics* for an undergraduate student enrolled at a Canadian institution and having completed, during the year of application, a project in any field of acoustics or vibration. • Un livre sur l'acoustique d'un montant maximal de 150 \$ et un abonnement d'un an à la revue *Acoustique Canadienne* à un(e) étudiant(e) inscrit(e) dans un programme au 1er cycle universitaire dans une institution 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.

NEWS / INFORMATIONS

CONFERENCES

If you have any news to share with us, send them by mail or fax to the News Editor (see address on the inside cover), or via electronic mail to stevenb@aciacoustical.com

2008

08-10 March. 34th Meeting of the German Association for Acoustics. Dresden, Germany. Web: <http://2008daga-tagung.de>

17-19 March. Spring Meeting of the Acoustical Society of Japan. Narashino, Japan. Web: www.asj.gr.jp/index-en.html

March 30 - April 01. SAE-Brazil Noise and Vibration Conference-NVH. Florianopolis, SC, Brazil. Web: www.saebrasil.org.br/eventos/secao_parana_sc/nvh2008/site/

March 31 - April 04. International Conference on Acoustics, Speech, and Signal Processing (IEEE ICASSP 2008). Las Vegas, Nevada, USA. Web: www.icassp2008.org

08-11 April. Oceans '08. Kobe, Japan. Web: www.oceans08mtsieekobe-technoocean08.org/index.cfm

10-11 April. Institute of Acoustics (UK) Spring Conference. Reading, UK. Web: www.ioa.org.uk/viewupcoming.asp

17-18 April. Swiss Acoustical Society Spring Meeting. Bellinzona (Tessin), Switzerland. Web: www.sga-ssa.ch

12-15 May. 10th Spring School on Acousto-optics and Applications. Sopot, Poland. Web: <http://univ.gda.pl/~school>

29 June - 04 July: Joint Meeting of European Acoustical Association, Acoustical Society of America, and Acoustical Society of France. Paris, France. Web: www.sfa.asso.fr/en/index.htm

06-10 July. 15th International Congress on Sound and Vibration. Daejeon, Korea. Web: www.icsv15.org

7-10 July: 18th International Symposium on Nonlinear Acoustics (ISNA18). Stockholm, Sweden. E-mail: benflo@mech.kth.se

21-25 July. 9th International Congress on Noise as a Public Health Problem. Mashantucket, CT, USA. Web: www.icben.org

27-30 July. Noise-Con 2008. Dearborn, MI, USA.

27-31 July. 10th Mechanics of Hearing Workshop. Keele University, UK. Web: www.mechanicsofhearing.com

28 July - 1 August: 9th International Congress on Noise as a Public Health Problem. Mashantucket, Pequot Tribal Nation, (CT, USA). Web: www.icben.org

25-29 August. 10th International Conference on Music Perception and Cognition. Sapporo, Japan. Web: <http://icmpc10.typepad.jp>

08-12 September: International Symposium on Underwater Reverberation and Clutter. Lerici, Italy. Web: <http://isurc2008.org>

10-12 September. Autumn Meeting of the Acoustical Society of Japan. Fukuoka, Japan. Web: www.asj.gr.jp/index-en.html

15-17 September: International Conference on Noise and Vibration Engineering (ISMA2008). Leuven, Belgium. Web: www.isma-isaac.be

CONFÉRENCES

Si vous avez des nouvelles à nous communiquer, envoyez-les par courrier ou fax (coordonnées incluses à l'envers de la page couverture), ou par courriel à stevenb@aciacoustical.com

2008

08-10 mars. 34th Meeting of the German Association for Acoustics. Dresden, Germany. Web: <http://2008daga-tagung.de>

17-19 mars. Spring Meeting of the Acoustical Society of Japan. Narashino, Japan. Web: www.asj.gr.jp/index-en.html

mars 30 - avril 01. SAE-Brazil Noise and Vibration Conference-NVH. Florianopolis, SC, Brazil. Web: www.saebrasil.org.br/eventos/secao_parana_sc/nvh2008/site/

mars 31 - avril 04. International Conference on Acoustics, Speech, and Signal Processing (IEEE ICASSP 2008). Las Vegas, Nevada, USA. Web: www.icassp2008.org

08-11 avril. Oceans '08. Kobe, Japan. Web: www.oceans08mtsieekobe-technoocean08.org/index.cfm

10-11 avril. Institute of Acoustics (UK) Spring Conference. Reading, UK. Web: www.ioa.org.uk/viewupcoming.asp

17-18 avril. Swiss Acoustical Society Spring Meeting. Bellinzona (Tessin), Switzerland. Web: www.sga-ssa.ch

12-15 mai. 10th Spring School on Acousto-optics and Applications. Sopot, Poland. Web: <http://univ.gda.pl/~school>

29 juin - 04 juillet: Joint Meeting d'Europe Acoustical Association, Acoustical Society d'America, et Acoustical Society du France. Paris, France. Web: www.sfa.asso.fr/en/index.htm

06-10 juillet. 15th International Congress on Sound and Vibration. Daejeon, Korea. Web: www.icsv15.org

7-10 juillet: 18th International Symposium sur Nonlinear Acoustics (ISNA18). Stockholm, Sweden. E-mail: benflo@mech.kth.se

21-25 juillet. 9th International Congress on Noise as a Public Health Problem. Mashantucket, CT, USA. Web: www.icben.org

27-30 juin. Noise-Con 2008. Dearborn, MI, USA.

27-31 juillet. 10th Mechanics of Hearing Workshop. Keele University, UK. Web: www.mechanicsofhearing.com

28 juillet - 1 août: 9th International Congress sur Noise as a Public Health Problem. Mashantucket, Pequot Tribal Nation, (CT, USA). Web: www.icben.org

25-29 août: 10th International Conference on Music Perception and Cognition. Sapporo, Japan. Web: <http://icmpc10.typepad.jp>

08-12 septembre: International Symposium on Underwater Reverberation and Clutter. Lerici, Italy. Web: <http://isurc2008.org>

10-12 septembre. Autumn Meeting of the Acoustical Society of Japan. Fukuoka, Japan. Web: www.asj.gr.jp/index-en.html

15-17 septembre: International Conference on Noise and Vibration Engineering (ISMA2008). Leuven, Belgium. Web: www.isma-isaac.be

16-18 September: Underwater Noise Measurement. Southampton, UK. Web: www.ioa.org.uk/viewupcoming.asp

22-26 September: Interspeech 2008 - 10th ICSLP, Brisbane, Australia. Web: www.interspeech2008.org

23-25 September: Underwater Noise Measurement. Southampton, U.K. Web: www.ioa.org.uk/viewupcoming.asp

21-23 October. 13th Conference on Low Frequency Noise and Vibration. Tokyo, Japan. Web: www.lowfrequency2008.org

21-24 October. Acustica 2008. Coimbra, Portugal. Web: www.spacustica.pt

26-29 October: Internoise 2008, Shanghai, China. Web: www.internoise2008.org

01-05 November. IEEE International Ultrasonic Symposium. Beijing, China. Web: www.ieee-uffa.org/ulmain.asp?page=symposia

10-14 November. 156th Meeting of the Acoustical Society of America, Miami, Florida, USA. Web: www.asa.aip.org

24-26 November. Australian Acoustical Society National Conference. Victoria, Australia. Web: www.acoustics.asn.au

2009

19-24 April. International Conference on Acoustics, Speech, and Signal Processing. Taipei, R.O.C. Web: icassp09.com

18-22 May. 157th Meeting of the Acoustical Society of America, Portland, Oregon, USA. Web: www.asa.aip.org

23-26 August: Internoise 2009, Ottawa, Canada.

23-27 August: International Confress on Acoustics 2010. Sydney, Australia. Web: www.acoustics.asn.au

06-10 September: Interspeech 2009. Brighton, UK. Web: www.interspeech2009.org

2010

19-24 March. International Conference on Acoustics, Speech, and Signal Processing. Dallas, TX, USA. Web: icassp2010.org

23-27 August: International Confress on Acoustics 2010. Sydney, Australia. Web: www.acoustics.asn.au

26-30 September: Interspeech 2010. Makuhari, Japan. Web: www.interspeech2010.org

16-18 septembre: Underwater Noise Measurement. Southampton, UK. Web: www.ioa.org.uk/viewupcoming.asp

22-26 septembre: Interspeech 2008 - 10th ICSLP, Brisbane, Australia. Web: www.interspeech2008.org

23-25 septembre. Underwater Noise Measurement. Southampton, U.K. Web: www.ioa.org.uk/viewupcoming.asp

21-23 octobre. 13th Conference on Low Frequency Noise and Vibration. Tokyo, Japan. Web: www.lowfrequency2008.org

21-24 octobre. Acustica 2008. Coimbra, Portugal. Web: www.spacustica.pt

26-29 Octobre: Internoise 2008, Shanghai, China. Web: www.internoise2008.org

01-05 novembre. IEEE International Ultrasonic Symposium. Beijing, China. Web: www.ieee-uffa.org/ulmain.asp?page=symposia

10-14 novembre. 156th Meeting of the Acoustical Society of America, Miami, Florida, USA. Web: www.asa.aip.org

24-26 novembre. Australian Acoustical Society National Conference. Victoria, Australia. Web: www.acoustics.asn.au

2009

19-24 avril. International Conference on Acoustics, Speech, and Signal Processing. Taipei, R.O.C. Web: icassp09.com

18-22 mai. 157th Meeting of the Acoustical Society of America, Portland, Oregon, USA. Web: www.asa.aip.org

23-26 août: Internoise 2009, Ottawa, Canada.

23-27 août: International Confress sur Acoustics 2010. Sydney, Australia. Web: www.acoustics.asn.au

06-10 septembre: Interspeech 2009. Brighton, UK. Web: www.interspeech2009.org

2010

19-24 mars. International Conference on Acoustics, Speech, and Signal Processing. Dallas, TX, USA. Web: icassp2010.org

23-27 août: International Confress sur Acoustics 2010. Sydney, Australia. Web: www.acoustics.asn.au

26-30 septembre: Interspeech 2010. Makuhari, Japan. Web: www.interspeech2010.org

NEWS

We want to hear from you! If you have any news items related to the Canadian Acoustical Association, please send them. Job promotions, recognition of service, interesting projects, recent research, etc. are what make this section interesting.

EXCERPTS FROM “WE HEAR THAT...”, IN ECHOS, ASA

Gerhard Sessler, Professor at the Darmstadt University of Technology in Germany, received the Technology Award of the Eduard Rhein Foundation for “outstanding and internationally acknowledged achievements in numerous areas of technical acoustics.”

Laymon Miller is the sixth recipient of the 2007 C. Paul Boner award from the National Council of Acoustical Consultants (NCAC). The award is presented to a member of the acoustical consulting community who embodies the qualities of the late Paul Boner—teacher, scientist, administrator, technician—and who has made outstanding contributions to the science of acoustics. Miller is a Fellow of ASA and a member of the Institute of Noise Control Engineering (INCE).

Lisa Zurk, Director of the Northwest Electromagnetics and Acoustics Research Laboratory at Portland State University, received a Presidential Early Career Award for Scientists and Engineers at the White House. Zurk was nominated by the National Science Foundation from which she previously received a five-year, \$400,000 NSF Career Award. Her teaching and research are in electromagnetics, acoustics, and computational methods.

The National Academy of Engineering presented the 2007 Arthur M. Bueche Award to **Jordan Baruch**, President, Jordan Baruch Associates “for the promotion of the innovation and management of science and technology nationally and internationally, thereby enhancing the economy of the U.S. and developing nations.” Baruch is a Fellow of ASA.

Michael Howe, Professor of Theoretical Mechanics at Boston University, was awarded the 2007 Rayleigh Medal for his outstanding contributions to research, mainly in aeroacoustics stretching over almost four decades. The presentation was made by Colin English, President of the Institute of Acoustics at the Institute’s recent Autumn Conference on Advances in Noise and Vibration Engineering at Oxford. The Rayleigh Medal is awarded without regard to age to persons of undoubted renown for outstanding contributions to acoustics. It is normally presented to a UK acoustician in even numbered years and an overseas acoustician in odd numbered years. Howe is an ASA Fellow.

The Canadian Association of Physicists has honoured University of Windsor physics professor **Roman Maev** with the 2007 CAP medal of Outstanding Achievement in Industrial and Applied Physics for his work in the field of acoustic microscopy. Maev holds the Natural Sciences and Engineering Research Council/DaimlerChrysler/University of Windsor Industrial Research Chair in Applied Solid State Physics and Material Characterization.

Bishwajit Chakraborty, a senior scientist in the Geological Oceanography Division at National Institute of Oceanography, Dona Paula, Goa, is a recipient of the National Mineral Award -2006 for his significant contributions in the field of earth sciences and related fields under the National Mineral Award Scheme of the Ministry of Mines, Government of India.

Michael Canney is the new chair of the Student Council. Michael is a doctoral student in Bioengineering at the University of Washington, where his research is focused on therapeutic ultrasound for noninvasive surgery. He has served on the ASA Student Council since the spring of 2005 as the representative for the Biomedical Ultrasound / Bioresponse to Vibration.

Emily Tobey, Nelle C. Johnston Chair in Communication Disorders in the School of Behavioral and Brain Sciences at the University of Texas at Dallas, has been named a Sigma Xi Distinguished Lecturer for 2008-2009. Tobey, a Fellow of ASA and the American Speech-Language and Hearing Association, was named the University of Texas at Dallas Polykarp Kusch Lecturer, the highest honor an individual faculty member can receive from the University.

EXCERPTS FROM “SCANNING THE JOURNALS”, IN ECHOS, ASA

Hair cells of the inner ear transduce mechanical vibrations arising from sound waves into electrochemical signals. At the apical surface each hair cell contains a bundle of stereocilia. According to a letter in the 6 September issue of *Nature*, two cadherins interact to form tip links that connect the stereocilia and are thought to gate the mechano-electrical transduction channel. Cadherin 23 (CDH23) and protocadherin 15 (PCDH15) localize to the upper and lower part of the tip links, respectively. Ions that affect the tip-link integrity and a mutation in PCDH15 that causes a recessive form of deafness disrupt interaction between CDH23 and PCDH15. A remarkable photo by the authors of this report appears on p. 112 of the October issue of *Physics Today*. In this photo, CDH23, tagged with a green fluorescent antibody, is found at the tail end of the tip link.

Physicists have discovered a simple way to “store” light pulses in a material by converting them into **sound waves** according to an article in the 14 December issue of *Science*. The optical data pulses are converted into long-lived acoustic excitations in an optical fiber by means of stimulated Brillouin scattering. These stored pulses can be retrieved later, after a time interval limited by the lifetime of the acoustic excitation. In the experiment reported, smooth 2-nanosecond-long pulses were stored for up to 12 ns with

good readout efficiency. This method can potentially be implemented at any wavelength where the fiber is transparent and can be incorporated into existing telecommunication networks because it operates using commercially available components at room temperature.

“Stringing the fiddle: **the inner ear’s two-part invention**” is the title of an article in the October issue of *Nature Neuroscience* which discusses the roles of CDH23 and PCDH15 in transduction by the stereocilia tip links. Physiological experiments illuminate channel function in a dazzling array of detail. Deflection of a hair cell’s stereocilia bundle by ~100 nm opens ~100 transduction channels, located at the tips. Each channel is in series with a “molecular gating spring” of ~1 millinewton per meter stiffness that can stretch by >100 nm. Each tip link, about 170 nm long, appears as a twisted pair of strands 5 nm in diameter.

Would-be conductors now have a system that allows them to change the **tempo and dynamics** of a virtual orchestra with the wave of a hand, according to a note in the 20 October issue of *New Scientist*. Audio and video recordings of a real orchestra are used to create the virtual orchestra. The would-be conductor wears an ewatch, a device the size of a wristwatch that contains accelerometers and tilt sensors. The ewatch records the user’s hand movements and sends them to a computer, where software translates the actions into dynamics and tempo commands and feeds them to the virtual orchestra.

People blinded early in life often develop **better hearing** than sighted people. According to a note in the 20 October issue of *New Scientist*. They do this by taking over the parts of the visual system that are easiest to adapt, the medial occipital. In sighted people the medial occipital plays a crucial role in registering visual signals by setting the thresholds at which they are noticed by the brain. Researchers played a series of sounds to blind subjects, each preceded by a cue warning the brain to pay attention. Brain scans showed that as subjects heard the cue their medial occipitals became more active, indicating that the brain uses the same region that alerts them to visual signals to prime them to listen for sounds.

“A Sound Use for Heat” is the title of an article on **thermoacoustics** in the November/December issue of *American Scientist*. The article focuses largely on the work of Orest Symko and his colleagues at the University of Utah (see *ECHOES*, Summer 2006 issue). Their work has been directed especially toward converting waste heat from computers, electronics, power plants, and automobiles into electricity. To accomplish this, the heat is first used to generate sound, which is then converted into electricity by means of piezoelectric transducers. The thermoelectric converters, which have no moving parts, can work with a temperature difference as low as 25 degrees Celsius, although larger temperature gradients increase efficiency.

The WGBH Educational Foundation, in association with Sigma Xi, has launched a website, sciencecafes.org, to promote the growing **Science Café** movement in the U.S., according to a story in the November-December issue of *American Scientist*. From a handful of gatherings a few years ago to more than 50 around the country today, the café format has proven that people of all ages and backgrounds enjoy talking about the latest developments in science. The largest at the time of my editorial “Café Acoustique?” in the Fall 2006 issue of *ECHOES* was the Denver Café Scientifique which draws about 150 people. Meanwhile, has anyone tried out the Café Acoustique idea, or even made a presentation on acoustics at a Café Scientifique?

The history and appearance of renal (kidney) tissue in rabbits after **histotripsy** is the subject of a paper in the October issue of the *Journal of Endourology*. Histotripsy, defined as “noninvasive, nonthermal, mechanical (cavitation) tissue ablation,” is an experimental type of bloodless surgery. The left kidneys of 29 rabbits were treated with 750-kHz bursts of ultrasound from an 18-element phased-array transducer. After 60 days only a small fibrous scar persisted adjacent to a wedge of tubular dilation and fibrosis underlying a surface-contour defect.

“There’s more to **yodeling** than meets the ear,” according to an article in the December 22/29 issue of *New Scientist*. Kerry Christensen, who performed a mini-concert at the ASA meeting in Salt Lake City, can lay claim to being one of the world’s most versatile yodelers. His repertoire of 1500 tunes includes Cajun yodeling as well as a number in which performs a series of rapid chromatic runs up and down the musical scale. Human voices have two distinct ranges that singers call the “head voice” and the “chest voice.” There is a distinct gap between these two ranges, which is noticed most in inexperienced singers. Opera singers are experts at smoothing over this break, while yodelers accentuate it, says Ingo Titze of the University of Iowa.

In August, the World Health Organization (WHO) released preliminary estimates of the number of Europeans killed or disabled by **exposure to noise**, according to a story in 22/29 December issue of *New Scientist*. For example, chronic and excessive traffic noise is implicated in the deaths of 3 per cent of people in Europe with ischaemic heart disease. Noise kills in much the same way as chronic stress does, by causing an accumulation of stress hormones, inflammation and changes in body chemistry that eventually lead to problems such as impaired blood circulation and heart attacks. Next year the WHO will finalize its estimates of the amazing effects of noise and will also provide guidelines on exposure levels that are likely to cause harm.

A brain transcription called FOXP2 is necessary in order for zebra finches to **learn to sing**, according to a paper in the Public Library open access journal *PLoS Biology* 5, e321 (2007). After a 3-month tutoring period with an adult bird, the songs of the birds with reduced FOXP2 were missing syllables and contained inappropriately repeated segments. Without sufficient FOXP2 normal developmental motor learning could not take place.



--- SECOND ANNOUNCEMENT ---

ACOUSTICS WEEK IN CANADA

Vancouver, 6 - 8 October 2008

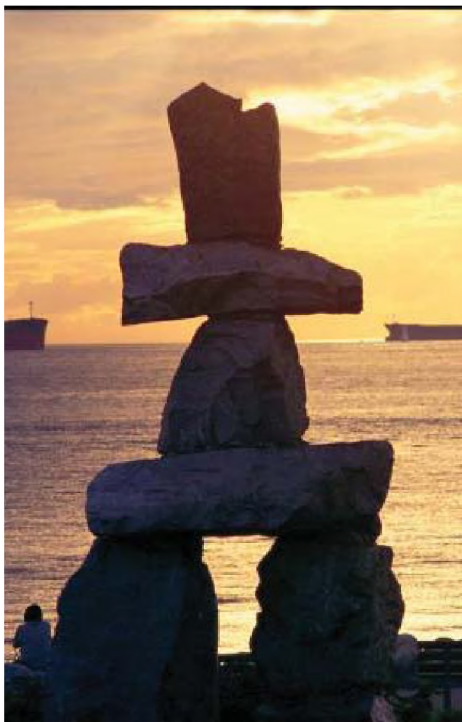
Acoustics Week in Canada 2008, the annual conference of the Canadian Acoustical Association, will be held in Vancouver, British Columbia from 6 to 8 October 2008. This is the premier Canadian acoustical event of the year, and is being held in beautiful, vibrant Vancouver, making it an event that you do not want to miss. The conference will include three days of plenary lectures, technical sessions on a wide range of areas of acoustics, the CAA Annual General Meeting, an equipment exhibition, and the conference banquet and other social events.

Venue and Accommodation – The conference will be held at the Coast Plaza Hotel & Suites [http://www.coasthotels.com/hotels/canada/bc/vancouver/coast_plaza/overview], in the dynamic West End of downtown Vancouver, steps from the beach at English Bay, walking distance to beautiful Stanley Park and trendy Robson Street, near Granville Island and Chinatown. Participants registering with the hotel before 5 September 2008 will receive the reduced room rate of \$129/night (single or double). Stay at the conference hotel to be near all activities and your colleagues, and to help make the conference a financial success, to the benefit of all CAA members.

Plenary Lectures – Plenary lectures will be presented by Prof. Stan Dosso, School of Earth and Ocean Sciences, University of Victoria, Prof. Barry Truax, Department of Communication, Simon Fraser University and Prof. John Esling, Department of Linguistics, University of Victoria.

Special Sessions – Special sessions consisting of invited and contributed papers are currently being organized on the following topics:

- Architectural and Classroom Acoustics
- Auditory Scene Analysis
- Sound Absorbing Materials
- Biomedical Acoustics
- Speech Production and Speech Disorders
- First Nations Languages Acoustics
- Speech Perception
- Acoustical Consulting—Challenges and Opportunities
- Second Language Acquisition Acoustics
- Occupational Noise Standards
- Psychological Acoustics
- Vibroacoustics



If you would like to propose and/or organize a special session in your technical area, please contact the Conference Chair or Technical Co-Chair as soon as possible.

Equipment Exhibition – The conference will include a one-day exhibition of acoustical equipment and products on Tuesday 7 October 2008. If you are an equipment supplier interested in participating in the exhibition, please contact the Exhibition Coordinator as soon as possible.



Social Events – The conference will begin on Monday morning with an opening ceremony and welcome by Elder Larry Grant, Musqueam Indian Band (<http://www.musqueam.bc.ca/Home.html>). On Monday evening, a reception will be held for all delegates, followed by a visit to Christ Church Cathedral in downtown Vancouver, where acoustical consultant Michael Noble, BKL will discuss renovations to improve the Cathedral’s acoustical environment, after which delegates will experience the acoustics during an organ recital.

Courses / Seminars – If you would like to propose to offer a course / seminar in association with Acoustics Week in Canada, please contact the Conference Chair. Assistance can be provided in accommodating such a course / seminar, but it must be financially independent of the conference.

Student Participation – The participation of students is strongly encouraged. Travel subsidies and reduced registration fees will be available. A hotel room-sharing program will be available to reduce costs. Student presenters are eligible to win prizes for the best presentations at the conference.

Paper Submission – Following are the deadlines for submission of abstracts, and of two-page summaries for publication in the proceedings issue of *Canadian Acoustics*: submission of abstracts: 13 June 2008; submission of two-page summaries: 10 July 2008.



Registration – details of registration fees and the registration form will be made available on the conference website. Early registration at a reduced fee is available until 5 September 2008.

Local Organizing Committee

- *Conference Chair: Murray Hodgson* [murray.hodgson@ubc.ca]
- *Technical Co-Chair: Kimary Shahin* [kns3@sfu.ca]
- *Venue: Linda Rammage* [linda.rammage@vch.ca]
- *Treasurer: Mark Cheng* [mark_cheng@yvr.ca]
- *Equipment Exhibition: Mark Bliss* [bliss@bkl.ca]
- *Audio/Visual: Christine Harrison*
[christine.harrison@worksafebc.com]
- *Student Issues, Translation: Hind Sbini* [sbihi@interchg.ubc.ca]
- *Administrator: Bernadette Duffy* [bduffy@interchg.ubc.ca]

Conference Website at <http://www.caa-aca.ca/>

--- SECONDE ANNONCE --- SEMAINE CANADIENNE D'ACOUSTIQUE



Vancouver, 6 - 8 Octobre 2008

La conférence annuelle de l'Association Canadienne d'Acoustique (ACA) se tiendra à Vancouver en Colombie-Britannique du 6 au 8 octobre 2008. Il s'agit du plus important événement canadien de l'acoustique de l'année, et aura lieu à Vancouver, une des plus pittoresques et vibrantes villes canadiennes. Trois jours de sessions plénières, ainsi que des sessions techniques parallèles seront présentées, couvrant un large éventail du domaine de l'acoustique. La conférence comprendra aussi la réunion annuelle générale de l'ACA, l'exposition de divers équipements acoustiques, un banquet et autres événements sociaux.

Lieu du congrès et hébergement – La conférence se tiendra au Coast Plaza Hotel & Suites [http://www.coasthotels.com/hotels/canada/bc/vancouver/coast_plaza/overview], dans le quartier dynamique West End du centre-ville de Vancouver, à quelque pas de la plage de la baie des Anglais (English Bay), à proximité du fameux parc Stanley et de la chic rue Robson, et proche du marché populaire de l'île Granville, et du Chinatown. Les délégués qui réserveront leur chambre avant le 5 septembre 2008 bénéficieront d'un tarif préférentiel de \$129/nuit (occupation simple ou double). Choisissez cet hôtel pour participer pleinement au congrès, à proximité de toutes les activités et de vos collègues, et pour assurer le succès de la conférence pour le bénéfice de tous les membres de l'ACA.

Sessions plénières – Les professeurs Stan Dosso, Earth and Ocean Sciences, Université de Victoria, Barry Truax, Département de Communication, Université Simon Fraser, et John Esling, Département de Linguistique, Université de Victoria, assureront les présentations des sessions plénières.

Sessions spéciales – Des sessions présentées par des conférenciers invités ou par des communications soumises par les délégués sont actuellement organisées autour de divers sujets, tels que:



- Acoustique architecturale et de salles de classe
- Matériaux absorbants
- Acoustique biomedical
- Troubles du débit
- Acoustique des langues des Premières Nations
- Perception du langage
- Consultation en acoustique – défis et perspectives
- Acoustique de l'acquisition d'une seconde langue
- Standards du bruit en milieu de travail
- Psychoacoustique
- Vibroacoustique

Si vous désirez suggérer un sujet de session spéciale et/ou organiser une de ces sessions, veuillez communiquer avec le président du congrès ou le directeur scientifique.

Exposition technique – Le mardi 7 octobre 2008 sera consacré à l'exhibition d'instruments et autres produits de l'acoustique. Si vous êtes un fournisseur d'équipement intéressé de participer, veuillez contacter la personne en charge de la coordination de l'exhibition.



Activités – La conférence débutera le lundi matin avec une cérémonie d’ouverture avec un discours de bienvenue par Elder Larry Grant, Musqueam Indian Band (<http://www.musqueam.bc.ca/Home.html>) Lundi soir, une réception est prévue, suivie par une visite de la Cathédrale Christ Church au centre-ville où Michael Noble, consultant chez la firme BKL, présentera les rénovations récentes qui ont été entreprises pour améliorer l’environnement acoustique de la Cathédrale. Les délégués pourront par la suite assister à un récital d’orgue.

Cours / Séminaires – Si vous désirez présenter un cours/séminaire en association avec la semaine canadienne d’acoustique, veuillez contacter le président du comité d’organisation. Sous condition d’une indépendance financière, l’accommodation d’un cours/séminaire pourra être appuyée.

Participation étudiante – La participation des étudiants au congrès est vivement encouragée. Des aides financières pour le déplacement et une réduction pour l’inscription seront mises à disposition. Un programme pour faciliter le partage des chambres sera mis sur pied pour réduire les dépenses. Les étudiants présentant leurs travaux seront éligibles pour les prix des meilleures présentations au congrès.



Soumission des présentations – Les dates limites pour soumission sont le 13 juin 2008 pour les résumés et le 10 juillet 2008 pour les sommaires de deux pages aux actes.

Inscription – Les détails ainsi que le formulaire d’inscription seront mis en ligne sur le site Web de la conférence. Une réduction sera effective pour toute inscription avant le 5 septembre 2008.

Comité d’organisation

- *Président: Murray Hodgson [murray.hodgson@ubc.ca]*
- *Directeur scientifique: Kimary Shahin [kns3@sfu.ca]*
- *Accommodations: Linda Rammage [linda.rammage@vch.ca]*
- *Tresorier: Mark Cheng [mark_cheng@yvr.ca]*
- *Exposition: Mark Bliss [bliss@bkl.ca]*
- *Audio-visuel: Christine Harrison [christine.harrison@worksafebc.com]*
- *Student Issues, Translation: Hind Sbihi [sbihi@interchg.ubc.ca]*
- *Administrateur: Bernadette Duffy [bduffy@interchg.ubc.ca]*

Site Web de la conférence à <http://www.caa-aca.ca/>

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 technical 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 photos 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.



Application for Membership

CAA membership is open to all individuals who have an interest in acoustics. Annual dues total \$65.00 for individual members and \$25.00 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 August 31 will be applied to the current year and include that year's back issues of *Canadian Acoustics*, if available. New membership applications received after August 31 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 \$65.00. Many companies and institutions prefer to be a Sustaining Subscriber, paying \$300.00 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 August 31 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 | 5. Psychological / Physiological Acoustic | 9. Underwater Acoustics |
| 2. Engineering Acoustics / Noise Control | 6. Shock and Vibration | 10. Signal Processing / Numerical Methods |
| 3. Physical Acoustics / Ultrasound | 7. Hearing Sciences | 11. Other |
| 4. Musical Acoustics / Electro-acoustics | 8. Speech Sciences | |

For student membership, please also provide:

_____ (University) _____ (Faculty Member) _____ (Signature of Faculty Member) _____ (Date)

I have enclosed the indicated payment for:

- CAA Membership \$ 65.00
 CAA Student Membership \$ 25.00

Payment by: Cheque
 Money Order
 VISA credit card (Only VISA accepted)

- Institutional Subscription \$ 65.00
 plus mailing surcharge outside Canada:
 \$8 to USA, \$15 other International

For payment by VISA credit card:

Card number _____

Name of cardholder _____

- Sustaining Subscriber \$ 300.00
 includes subscription (4 issues /year)
 to *Canadian Acoustics*.

Expiry date _____

 (Signature) (Date)

Mail application and attached payment to:

D. Quirt, Secretary, Canadian Acoustical Association, PO Box 74068, Ottawa, Ontario, K1M 2H9, Canada



Formulaire d'adhésion

L'adhésion à l'ACA est ouverte à tous ceux qui s'intéressent à l'acoustique. La cotisation annuelle est de 65.00\$ pour les membres individuels, et de 25.00\$ pour les étudiants. Tous les membres reçoivent *l'Acoustique Canadienne*, la revue de l'association. Les nouveaux abonnements reçus avant le 31 août 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 31 août 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 65.00\$. Des compagnies et établissements préfèrent souvent la cotisation de membre bienfaiteur, de 300.00\$ 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 31 août 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 31 août s'appliquent à l'année suivante.

Pour obtenir des formulaires électroniques, visitez le site Web: caa-aca.ca

Pour correspondance administrative 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 | 5. Physio / Psycho-acoustique | 9. Acoustique sous-marine |
| 2. Génie acoustique / Contrôle du bruit | 6. Chocs et vibrations | 10. Traitement des signaux / Méthodes numériques |
| 3. Acoustique physique / Ultrasons | 7. Audition | 11. Autre |
| 4. Acoustique musicale / Electro-acoustique | 8. Parole | |

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 65.00 \$
 Membre étudiant(e) 25.00 \$
 Abonnement institutionnel 65.00 \$
Surtaxe d'envoi à l'extérieur du Canada :
 8 \$ vers les États-Unis
 15 \$ tout autre envoi international
 Abonnement de soutien 300.00 \$
(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 :

The Canadian Acoustical Association l'Association Canadienne d'Acoustique



PRESIDENT PRÉSIDENT

Christian Giguère
Université d'Ottawa
Ottawa, Ontario
V8W 3P6
(613) 562-5800 x4649
cgiguere@uottawa.ca

PAST PRESIDENT PRÉSIDENT SORTANT

Stan Dosso
University of Victoria
Victoria, British Columbia
V8W 3P6
(250) 472-4341
sdosso@uvic.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
411 Confederation Parkway, Unit 19
Concord, Ontario
L4K 0A8
(905) 660-2444
dalila@jadeacoustics.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

WORLD WIDE WEB HOME PAGE: <http://www.caa-aca.ca>

Dave Stredulinsky
(902) 426-3100

ASSISTANT EDITOR RÉDACTEUR ADJOINT

Ralph Baddour
Department of Medical Biophysics
University of Toronto
rbaddour@uhnres.utoronto.ca

DIRECTORS DIRECTEURS

Alberto Behar
(416) 265-1816
behar@sympatico.ca

Vijay Parsa
(519) 661-2111 Ex. 88947
parsa@nca.uwo.ca

Jérémie Voix
(514) 932-2674
voix@caa-aca.ca

Nicole Collison
(902) 426-3100, Ext. 394
nicole.collison@drdc-rddc-gc.ca

Richard Peppin
(410) 290-7726
peppinr@scantekinc.com

Clair Wakefield
(250) 370-9302
nonoise@shaw.ca

Tim Kelsall
(905) 403-3932
tkelsall@hatch.ca

Frank Russo
(416) 979-5000 ext. 2647
russo@caa-aca.ca

SUSTAINING SUBSCRIBERS / ABONNES DE SOUTIEN

The Canadian Acoustical Association gratefully acknowledges the financial assistance of the Sustaining Subscribers listed below. Their annual donations (of \$300.00 or more) enable the journal to be distributed to all at a reasonable cost.

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 \$300.00 et plus) rendent le journal accessible à tous nos membres.

ACI Acoustical Consultants Inc.

Mr. Steven Bilawchuk - (780) 414-6373
stevenb@aciacoustical.com - Edmonton, AB

Aercoustics Engineering Ltd

Mr. John O'Keefe - (416) 249-3361
aercoustics@aercoustics.com - Rexdale, ON

Dalimar Instruments Inc.

Mr. Daniel Larose - (514) 424-0033
daniel@dalimar.ca - Vaudreuil-Dorion, QC

Eckel Industries of Canada Ltd.

Mr. Bruce Allan - (613) 543-2967
eckel@eckel.ca - Morrisburg, ON

Hatch Associates Ltd.

Mr. Tim Kelsall - (905) 403-3932
tkelsall@hatch.ca - Mississauga, ON

Integral DX Engineering Ltd.

Mr. Greg Clunis - (613) 761-1565
greg@integraldxengineering.ca - Ottawa, ON

Jade Acoustics Inc.

Ms. Dalila Giusti - (905) 660-2444
dalila@jadeacoustics.com - Concord, ON

MJM Conseillers en Acoustique Inc.

MJM Acoustical Consultants Inc.
M. Michel Morin - (514) 737-9811
mmorin@mjm.qc.ca - Montréal, QC

OZA Inspections Ltd.

Mr. David Williams - (800) 664-8263x25
oza@ozagroup.com - Grimsby, ON

Pinchin Environmental Ltd.

(905) 363-0678; FAX: (905) 363-0681
info@pinchin.com - Mississauga, ON

Scantek Inc.

Mr. Richard J. Peppin, (410)-290-7726
peppinr@scantekinc.com - Columbia, MD

Soft dB Inc.

M. André L'Espérance - (418) 686-0993
contact@softdb.com - Sillery, QC

State of the Art Acoustik Inc.

Dr. C. Fortier - (613) 745-2003,
sota@sota.ca - Ottawa, ON

Valcoustics Canada Ltd.

Dr. Al Lightstone - (905) 764-5223
solutions@valcoustics.com
Richmond Hill, ON

Water & Earth Science Associates (WESA)

Dejan Zivkovic, M.Sc. - (905) 639-5789 x151
dzivkovic@wesa.ca - Burlington, ON

ACO Pacific Inc.

Mr. Noland Lewis - (650) 595-8588
acopac@acopacific.com - Belmont, CA

Bruel & Kjaer North America Inc.

Mr. Andrew Khoury - (514) 695-8225
andrew.khoury@bksv.com - Pointe-Claire, QC

Dodge-Regupol

Mr. Paul Downey - (416) 440-1094
pcd@regupol.com - Toronto, ON

Enviro Noise Control Corp.

Alex V. Tardecilla - (403) 279-2764
alex@enctech.net - Calgary, AB

HGC Engineering Ltd.

Mr. Bill Gastmeier - (905) 826-4044
info@hgcengineering.com - Mississauga, ON

J.E. Coulter Associates Ltd.

Mr. John Coulter - (416) 502-8598
jcoulter@on.aibn.com - Toronto, ON

JASCO Research Ltd.

Mr. Scott Carr - (902) 405-3336
scott@jasco.com - Halifax, NS

Novel Dynamics Test Inc.

Mr. Andy Metelka - (519) 853-4495
ametelka@cogeco.ca - Acton, ON

Peutz & Associés

M. Marc Asselineau +33 1 45230500
m.asselineau@peutz.fr
Paris, FRANCE

Pyrok Inc.

(914) 777-7770; FAX: (914) 777-7103
info@pyrokinc.com - Mamaroneck, NY

SILEX Innovations Inc.

Mr. Mehmood Ahmed - (905) 612-4000
mehmooda@silex.com, Mississauga, ON

SounDivide Inc.

C.W. Ray Bakker - (877) 816-5435
ray.bakker@SounDivide.com - Calgary, AB

Swallow Acoustic Consultants Ltd.

Mr. John Swallow - (905) 271-7888
jswallow@jsal.ca - Mississauga, ON

Vibro-Acoustics

Mr. Tim Charlton - (800) 565-8401
tcharlton@vibro-acoustics.com
Scarborough, ON

West Caldwell Calibration Labs

Mr. Stanley Christopher - (905) 595-1107
info@wccl.com - Brampton

Acoustec Inc.

Dr. J.G. Migneron - (418) 834-1414
courrier@acoustec.qc.ca - St-Nicolas, QC

CDMca Ltd.

Jonas Sackanskis - (905) 265-7401
- Woodbridge, ON

Earth Tech Canada Inc.

Miroslav Ubovic - (905) 886-7022-x2215
noisevibration@earthtech.ca - Markham, ON

H.L. Blachford Ltd.

Mr. Dalton Prince - (905) 823-3200
amsales@blachford.ca - Mississauga, ON

Hydro-Quebec

M. Blaise Gosselin - (514) 840-3000x5134
gosselin.blaise@hydro.qc.ca - Montréal, QC

J.L.Richards & Assoc. Ltd.

Mr. Terry Vivyurka, P.Eng. - (613) 728-3571
mail@jlrichards.ca - Ottawa, ON

Mc SQUARED System Design Group

Mr. Wade McGregor - (604) 986-8181
info@mcsquared.com - North Vancouver, BC

Owens-Corning Canada Inc.

Mr. Salvatore Ciarlo - (800) 988-5269
salvatore.ciarlo@owenscorning.com -
St.Leonard, QC

Michel Picard

(514) 343-7617; FAX: (514) 343-2115
michel.picard@umontreal.ca - Brossard, QC

RWDI AIR Inc.

Peter VanDelden - (519) 823-1311
peter.vandelden@rwdi.com - Guelph, ON

SNC/Lavalin Environment Inc.

M. Jean-Luc Allard - (514) 651-6710
jeanluc.allard@snclavalin.com - Longueuil, QC

Spaarg Engineering Ltd.

Dr. Robert Gaspar - (519) 972-0677
gasparr@kelcom.igs.net - Windsor, ON

Tacet Engineering Ltd.

Dr. M.P. Sacks - (416) 782-0298
mal.sacks@tacet.ca - Toronto, ON

Wakefield Acoustics Ltd.

Mr. Clair Wakefield - (250) 370-9302
clair@wakefieldacoustics.com - Victoria, BC

Wilrep Ltd.

Mr. Don Wilkinson - (905) 625-8944
info@wilrep.com - Mississauga, ON