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

Francine Desharnais[1], Alex E. Hay[2]
Workshop Organizers
[1]DRDC Atlantic, PO Box 1012, NS, Canada, B2Y 3Z7, francine.desharnais@drdc-rddc.gc.ca
[2]Department of Oceanography, Dalhousie University, Halifax, NS, Canada B3H 4J1

1. INTRODUCTION

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

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

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

2. BACKGROUND PAPERS

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

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

3. DETECTION & CLASSIFICATION

3.1 Techniques

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

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

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

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

Harland and Armstrong [9] presented a suite of

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

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

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

3.2 Detection results and algorithms review

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

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

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

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

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

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

3.3 General comments

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

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

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

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

4. LOCALIZATION

4.1 Techniques

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

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

Optimization techniques home in on a position by

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

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

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

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

4.2 Results

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

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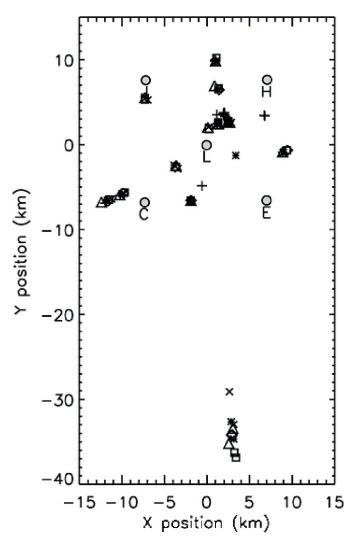

Fig. 1. 2002 results: Laurinolli and Hay (x); Simard et al., spectrum coincidence(); Simard et al., cross-correlation(*); Morrissey et al. (+); Simons et al. (◊); Desharnais et al. (Δ). Shaded circles indicate the OBH positions.

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

Table 1. Sound speed used by authors

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

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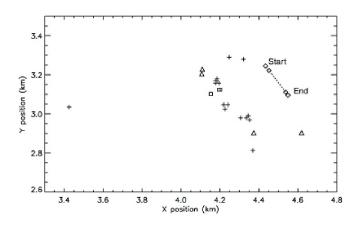
Fig. 2. 2000 results: Simard et al., spectrum coincidence (); Morrissey et al., (+); Desharnais et al., (△); location of the RHIB (◊).

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

4.3 Errors

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

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

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

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

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

5. DISCUSSION TOPICS

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

5.1 Automation

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

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

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

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

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

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

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

6. CONCLUSIONS

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

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

Recherche Scientifique (www.cebc.cnrs.fr). This workshop could be hosted in Monaco, October 2005.

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

ACKNOWLEDGEMENTS

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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