Available online at www.sciencedirect.com
ScienceDirect
Applied Acoustics 67 (2006) 1091–1105
applied acoustics
www.elsevier.com/locate/apacoust

Passive acoustic detection and localization of sperm whales (*Physeter macrocephalus*) in the tongue of the ocean

R.P. Morrissey *, J. Ward, N. DiMarzio, S. Jarvis, D.J. Moretti

Naval Undersea Warfare Center, Code 71, Bldg 1351, Newport, RI 02841, United States

Received 11 February 2006; received in revised form 14 April 2006; accepted 21 May 2006
Available online 25 July 2006

Abstract

A set of algorithms for real-time detection and localization of vocalizing marine mammals has been developed as part of the Marine Mammal Monitoring on Navy Ranges (M3R) program. These algorithms work on a broad variety of vocalizations including sperm whale clicks. The detection algorithm is a two stage process utilizing a binary thresholded FFT as the first stage. The second stage examines the FFT output to determine whether a click is present in a given FFT window. Detected clicks are split out of the data stream and sent to a data association algorithm called a scanning sieve. Time differences of arrival (TDOAs) are calculated which are then fed into 2D and 3D hyperbolic localization algorithms. Software written to implement the algorithms was used to process a data set consisting of sperm whale vocalizations provided as part of the *2nd International Workshop on Detection and Localization of Marine Mammals using Passive Acoustics*. Real-time detection and localization results from the data set are provided, along with a detailed description of the algorithms.

Keywords: Marine mammal; Detection; Localization; Sperm whale

* Corresponding author. Tel.: +1 401 832 5600; fax: +1 401 832 4441.
E-mail address: morrisseyrp@npt.nuwc.navy.mil (R.P. Morrissey).

0003-682X/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apacoust.2006.05.014

1. Introduction

Detection and localization of vocalizing marine mammals has been accomplished using a variety of methods, including spectrogram analysis [1–6], matched filters [7] and neural networks [2]. Some of the earliest work that applied detection and localization techniques to study marine mammals, and in particular sperm whales, was conducted by Watkins and Schevill [8–10]. Localization is typically done either via hyperbolic positioning methods [11,12,24] or by using acoustic propagation models [13–15].

This paper presents a method for using wide-baseline acoustic arrays for passive marine mammal tracking developed as part of the Marine Mammal Monitoring on Navy Ranges (M3R) project. Spectrogram analysis is utilized for click detection, coupled with a pattern matching algorithm for aligning click sequences among multiple hydrophones and determining time differences of arrivals (TDOAs). A hyperbolic positioning algorithm is then used to compute the animal's location. No attempts are made to automatically separate marine mammal vocalizations from other noise sources.

Sperm whales produce broadband impulsive clicks with a well-defined click train, as shown in Fig. 1. The clicks range in frequency from less than 100 Hz to at least 32 kHz [5,20,16], and in duration from 2 to 24 ms [20]. A variety of click train structures have been identified: regular or usual clicks, creaks, slow clicks and codas. It is thought that usual clicks and creaks are used for echolocation, whereas codas and slow clicks may have a communicative function [5,17–21].

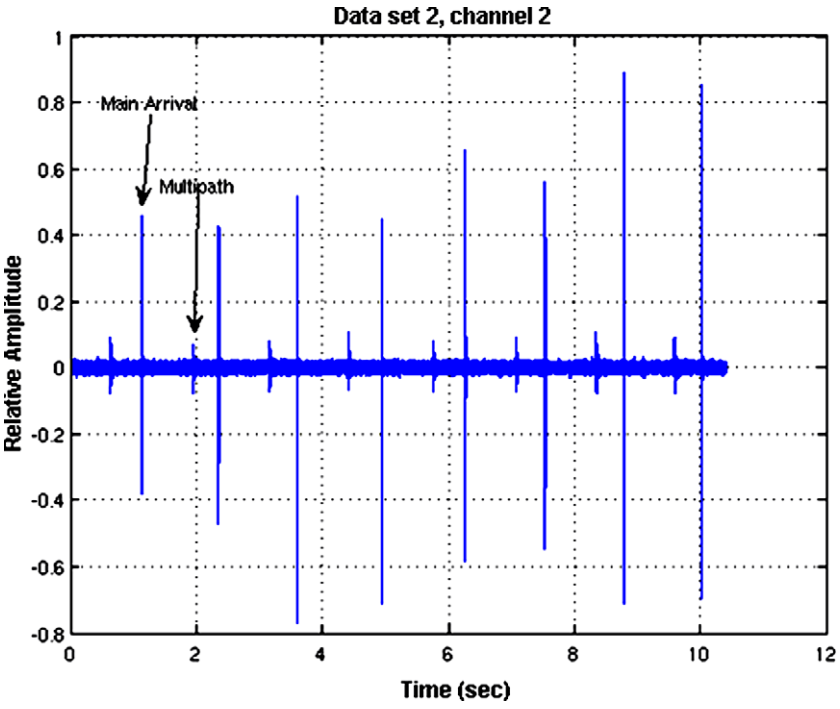

Fig. 1. Sperm whale click train.

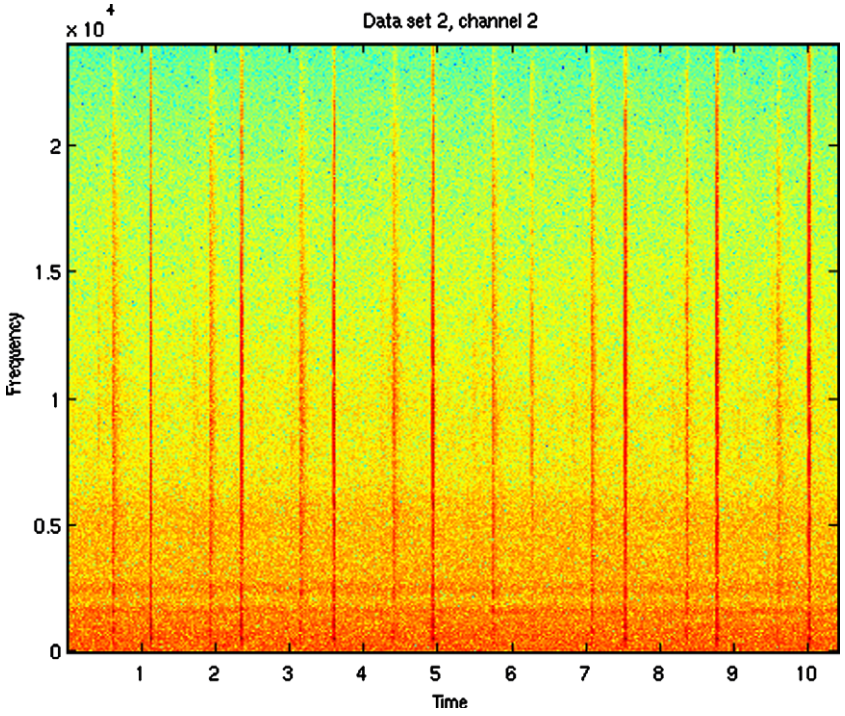

Fig. 2. Spectrogram of sperm whale click train. Amplitude ranges from green (lowest) to red (highest). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 2 shows a spectrogram of the data from Fig. 1. Clicks from a single sperm whale, plus reverberation, are evident. These clicks are characteristic of usual clicks, which are highly directional with typical inter-click intervals (ICIs) of 0.5–2 s [19]. They are broadband and extend from below 500 Hz to nearly the 24 kHz limit of the recording equipment used to capture the data.

Two data sets of sperm whale vocalizations were made available as part of the November 2005 workshop on detection and localization of marine mammals using passive acoustics [22]. The data sets were recorded on two separate sets of hydrophones. The hydrophones were bottom-mounted at a depth of about 2000 m, and spaced approximately two nautical miles apart. Required information such as the hydrophone locations and sound velocity profile were included with the data set. The first set of hydrophones contained data from multiple sperm whales, while the second set of hydrophones contained data from a single sperm whale (see Fig. 3.)

The data were provided in file form with a sample rate of 48 kHz and a resolution of 16 bits.

2. Technique

Marine mammal click localization consists of detection, data association, and position determination. Detection is broken into two stages. The first stage is an FFT based energy

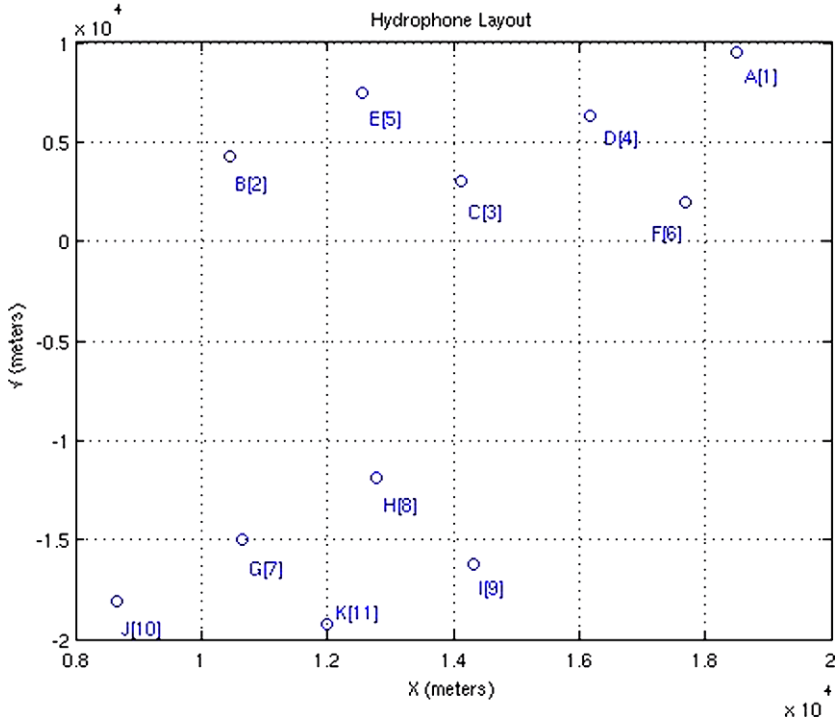

Fig. 3. Data set hydrophone layout.

detector that produces a binary frequency map as its output. The second stage determines whether a click is present by examining the frequency map. Detected clicks are time-aligned to produce a click pattern, or "click map", for each hydrophone. An association algorithm correlates click maps between several hydrophones and a master in a pair-wise fashion to produce time differences of arrivals (TDOAs). Finally, a hyperbolic positioning algorithm is applied to determine animal location.

2.1. Detector

Detection consists of a frequency domain energy detector coupled with a click detector. Each hydrophone is independently processed. The first stage is implemented using N-point fast Fourier transforms (FFTs) with a rectangular window and variable overlap (typically 50%). For this analysis, an FFT size of 512 points was chosen. Each frequency bin of the FFT output is compared to a time varying threshold, $D(f,t)$. The threshold is set to be m dB above the (time) average power within frequency bin f. For this data set, m was empirically chosen as -23 dB above the noise, and the time averaging window was 21.3 ms. The threshold (m) is typically chosen as low as possible to ride just over the noise. For each hydrophone, the output of the detector, $Q_i(f,t)$, is a binary valued frequency map derived from the FFT, which contains a 1 in each frequency bin that exceeded $D(f,t)$ and a 0 everywhere else. Frequency maps are only produced when at least one bin is above threshold. If no frequency bins are above threshold, no frequency map is produced. Consecutive frequency maps are used to form a binary spectrogram that indicates, in real-time, the pres-

ence of whale vocalizations and provides information on their frequency content. Fig. 4 shows the output of the detector for the click train from Fig. 1. The frequency map for each time step is plotted along the vertical axis. Cascading multiple frequency maps forms a binary spectrogram.

For each reported frequency map, clicks are detected by comparing the number of bins set in the map with a threshold, typically 10. If the number of bins set exceeds this threshold, it is determined that a click is present. Fig. 5 shows the binary spectrogram with detected clicks highlighted in red.

A detected click is removed from the data stream and passed to the click association algorithm.

2.2. Click association

Inter-click interval patterns have been found to be an effective means of both differentiating between individual whales and associating patterns of detections among hydrophones [23]. A series of clicks from a single animal is received on the surrounding hydrophones. Different (spatially separated) hydrophones may pick up different portions of the emitted click structure, but the inter-click interval will be nearly the same on all hydrophones. We assume animals exhibit their own unique pattern of clicks, and can thus be separated by cross-correlating a series of clicks across multiple hydrophones.

In the first step of the M3R click association algorithm the detected clicks are formed into a "click map" or comb filter for each hydrophone as shown in Fig. 6 below.

The click map is 10 s wide divided into time slots derived from the FFT size, decimated sample rate, and overlap. For this analysis, the data were downsampled to 24 kHz, resulting in a time slot of 10.7 ms (50% overlap * 512 FFT points/24,000 samples/s). The click map for a hydrophone contains a '1' in the time slot for each detected click with a '0' everywhere else. Detected clicks may belong to one or more animals. Conceptually, the next step is to cross correlate the click maps from several hydrophones with a master

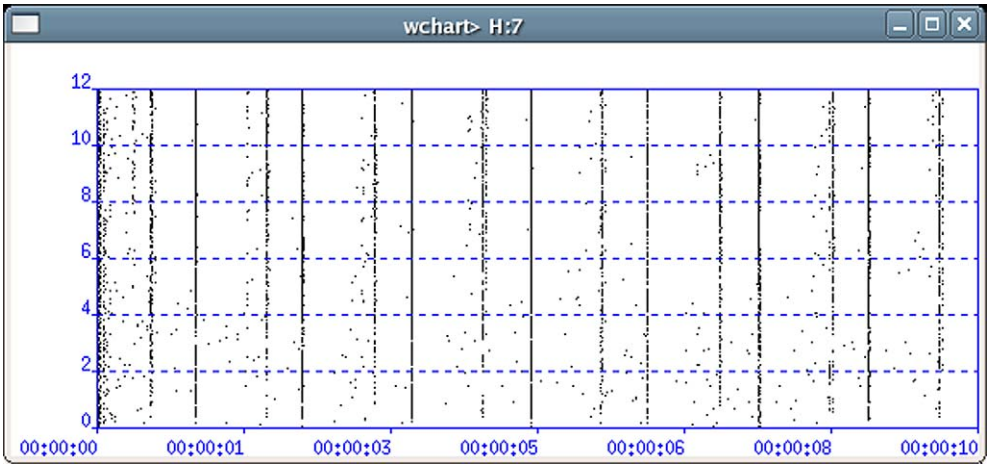

Fig. 4. Binary-valued detector output for time series in Fig. 1, 512 point FFT, 50% overlap, sample rate = 24 kHz, threshold $m = -23$ dB.

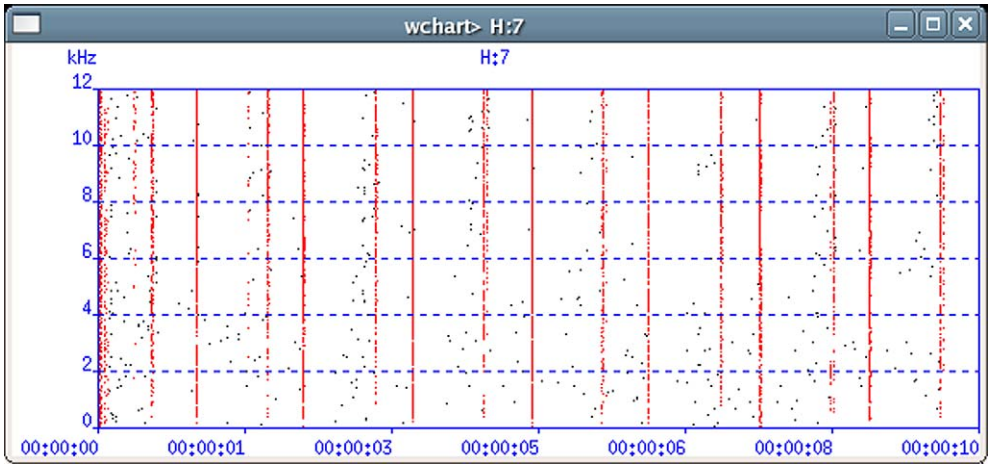

Fig. 5. Binary spectrogram shown in Fig. 4 with click detections indicated by red and rejected frequency maps indicated in black. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

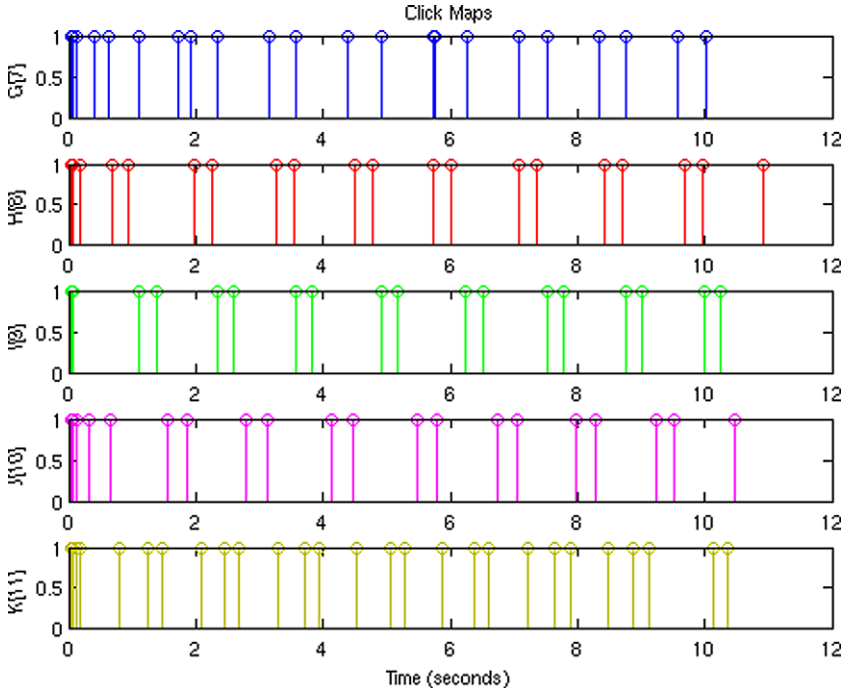

Fig. 6. Click maps for first 10 s from hydrophones in data set 2 (see Fig. 3).

hydrophone to align individual click patterns, and to find the TDOA between each hydrophone and the master. A click pattern unique to an individual will match the same pattern time-delayed on a different hydrophone. The time delay is proportional to the difference in

distance from the animal to the respective hydrophones. The M3R click association algorithm uses the notion of a "scanning sieve" to match click maps between hydrophones. Hydrophones are grouped into arrays which typically consist of five to seven hydrophones. For this data set, arrays are defined with four slave hydrophones surrounding a center master hydrophone. The click map on the master channel is used as a template by the scanning sieve. The master template is compared against the click map for each hydrophone in the array. The scanning sieve starts on a click detection on the master template. The selected slave click map is shifted across the master template one click at a time. The resultant correlation value at each shift of the slave click map represents the number of matches between the master template and the slave. The delay corresponding to the maximum correlation value between the master and a specified slave represents the difference in time of arrival (TDOA) between the two sensors. The output of the scanning sieve process are sets of TDOAs between the master and each hydrophone in the array. The master template is then shifted to the next click and the entire process is repeated.

Fig. 7 shows TDOAs resulting from the scanning sieve plotted versus time for four hydrophone pairs with hydrophone G[7] as the master. Each hydrophone pair is shown in a different color. Both direct and indirect path are evident for each hydrophone pair.

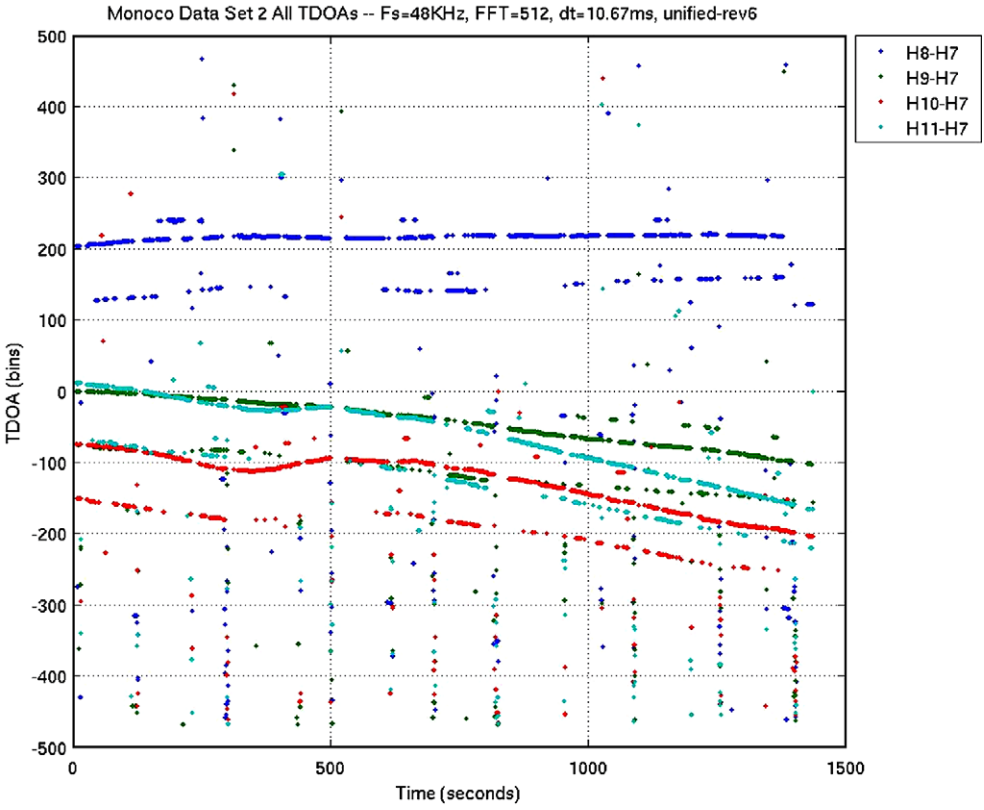

Fig. 7. Preliminary output of the M3R click association algorithm, showing the estimated TDOAs between the scanning sieve template, hydrophone G[7], and the four additional hydrophones for data set 2.

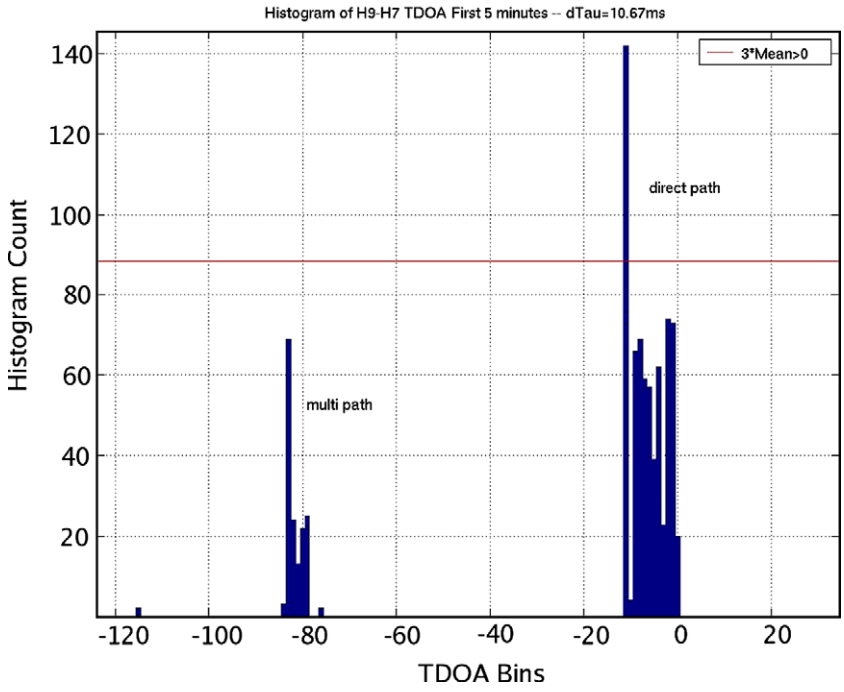

Fig. 8. Histogram of the TDOAs, with threshold lines for significant populations drawn.

To de-noise the TDOAs, the TDOA data from each hydrophone pair (master, slave) are histogrammed. Histograms are typically computed over a window of approximately 5 min. Significant TDOA groups are determined by thresholding the histograms. For this data set, the threshold is set to a value three times the mean, which has been empirically found to be effective on previous data sets. TDOA groups that pass threshold are sent to a multilateration localization algorithm that calculates a 2D or 3D position. More than one significant TDOA group may be present in the histogram. Each significant TDOA group is associated with a single animal, a co-located group of animals, or a reverb from another TDOA group. Histogramming the data from hydrophone pair G[7] and I[9], indicated in green in Fig. 7, yields the plot shown in Fig. 8.

Two peaks are evident in the histogram, corresponding to the direct and indirect path. The width of the direct and multipath clusters are proportional to the distance the animal has moved over the time the histogram is computed.

2.3. Localization

Positions are computed from the output of the click association sieve using a hyperbolic multilateration positioning algorithm developed by Vincent [24]. A representative sound speed profile is required. TDOAs are required between at least three hydrophones and the master in order to compute a 3D position. If less than four hydrophones are available, a 2D position is computed by fixing the depth at 0. For marine mammal tracking purposes, an arbitrarily shaped hydrophone array may be used. For the dense arrays consist-

ing of many hydrophones typical of Navy ranges, the array is usually divided into hexagon shaped hydrophone sub-arrays. These sub-arrays are independent and may be processed in parallel. Multiple processing nodes may be used to allow the entire hydrophone field to be processed in real-time.

3. Results

A replay tool was used to take the individual data files provided in the conference data sets and replay them through the detection system as if the data was being acquired in real time. To roughly match the 26.1 kHz sampling rate the system was originally tuned for, the data were decimated by two before input to the detector. The detector was configured to run a 512 point FFT with 50% overlap. This introduces a frequency folding into the output of the FFT, but does not noticeably affect the detected click arrival times.

The data were run in a completely automated fashion for both data sets.

Data set 1 consisted of six hydrophones with the geometry shown in Fig. 3 (hydrophones A–F). The data set contained multiple vocalizing animals. Four hydrophone arrays, each consisting of five hydrophones, were used: [C] D E B F, [D] A C E F, [E] B C D F, [F] A C D E, where the brackets indicate the master hydrophone in the given array.

Processing the raw acoustic data yielded 84 distinct positions within five nmi of the closest array master, 20 of which were 3D, and the rest 2D. These positions are shown in Figs. 9 and 10.

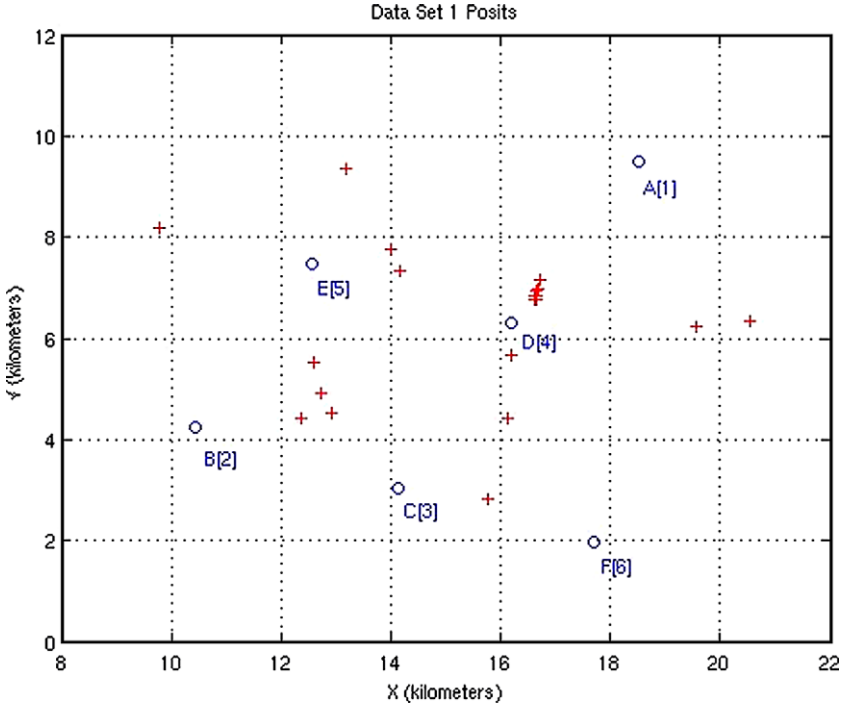

Fig. 9. Data set 1 *XY* (automated).

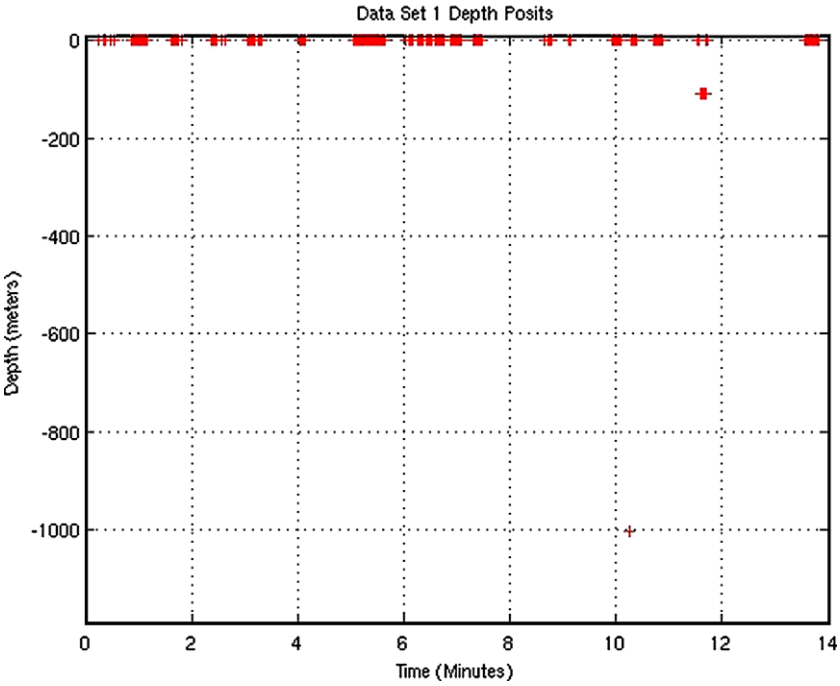

Fig. 10. Data set 1 depth (automated).

Data set 2 consisted of five hydrophones with the geometry shown in Fig. 3 (hydrophones G–K). The data set contained a single vocalizing animal. Four hydrophone arrays were used: [G] H I J K, [H] G I J K, [I] G H J K, [J] G H I K, consisting of the same five hydrophone with different array masters. The brackets again indicate the master hydrophone in the given array.

Automated processing of the raw acoustic data yielded 374 distinct positions within five nmi of the closest array master, of which 10 were 3D, and the rest 2D. A multipath track is clearly visible in the results. The positions are shown in Figs. 11 and 12.

Localization results from automated processing of both data sets were extremely poor, with a large number of 2D positions. Hand examination of the data indicated that many more 3D localizations should have been possible, particularly with the single animal in data set 2. To determine the reason for the poor localization performance, a hand analysis was conducted on the data from data set 2. The analysis indicated that when certain hydrophones were included in solution, the positioning algorithm would not converge. A section of data set 2, from 5 to 10 min, was chosen for further examination (see Table 1).

Three distinct groups of clicks from this section were observed on all hydrophones. The first clicks from groups one and three were arbitrarily chosen for comparison. The time of arrival on each hydrophone and the time difference between the first and second arrivals of the click were measured.

Based on a ray trace, the first arrival is assumed to be the direct path and the second arrival a surface reflection. In this case, the time difference between the direct path and

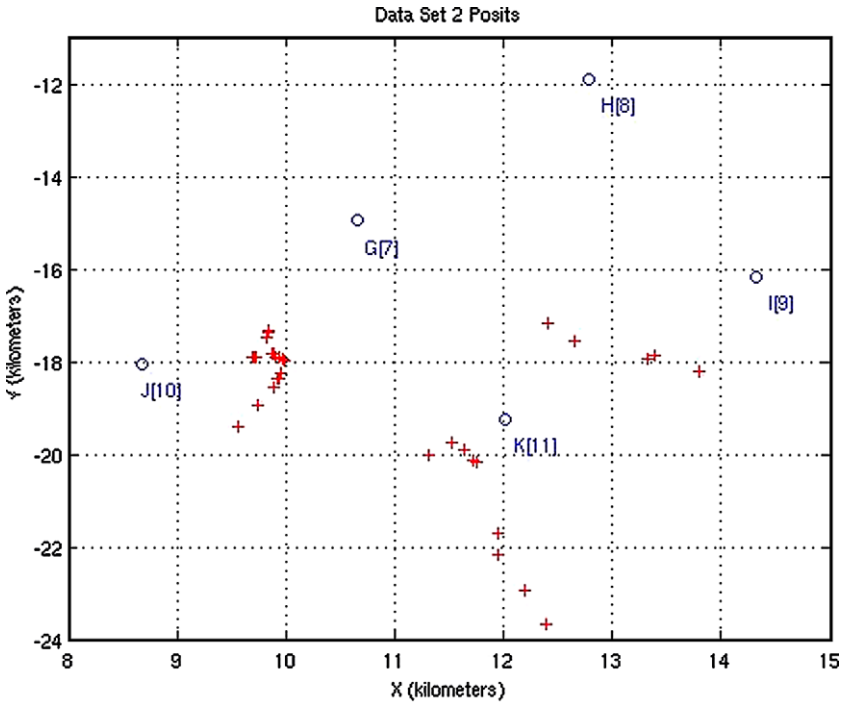

Fig. 11. Data set 2 *XY* (automated).

the surface reflection should be greatest on the hydrophone that is closest to the animal. However the arrival time on hydrophone G (the hydrophone with the largest multipath spread) is actually fourth out of five.

The data description from the conference web site indicates that hydrophones G and H were recorded on a different recorder than hydrophones I, J, and K. Assuming that the data cuts from the two recorders are out of sync would produce results in accordance with the observations. In addition, hydrophones G and H are roughly consistent in arrival times, as are I, J, and K.

In order to estimate an offset between the two recorders, a 2D track was computed on the animal using hydrophones I, J, and K. Click posit 60 was chosen to estimate a time offset as it was relatively isolated in time, and easy to identify on multiple hydrophones. The click was identified on hydrophone G, and the time of arrival and multipath spread were recorded. A ray trace was used to estimate the depth of the animal at 765 m, based on the multipath spread and the horizontal distance from the 2D posit to hydrophone G. Finally, the transit time of the animal click from the calculated position to hydrophone G was calculated using the ray trace. Using the computed time of emission (TOE) from the click posit, the calculated transit time, and the measured time of arrival of the click at the hydrophone, an offset of 2.249501 s was estimated between the two recorders. An independent estimate was then performed by playing back the synchronized recorders into a sound analysis program and precisely measuring the time offset between the recorders, yielding an offset of 2.3395 seconds, in good agreement with the estimated offset.

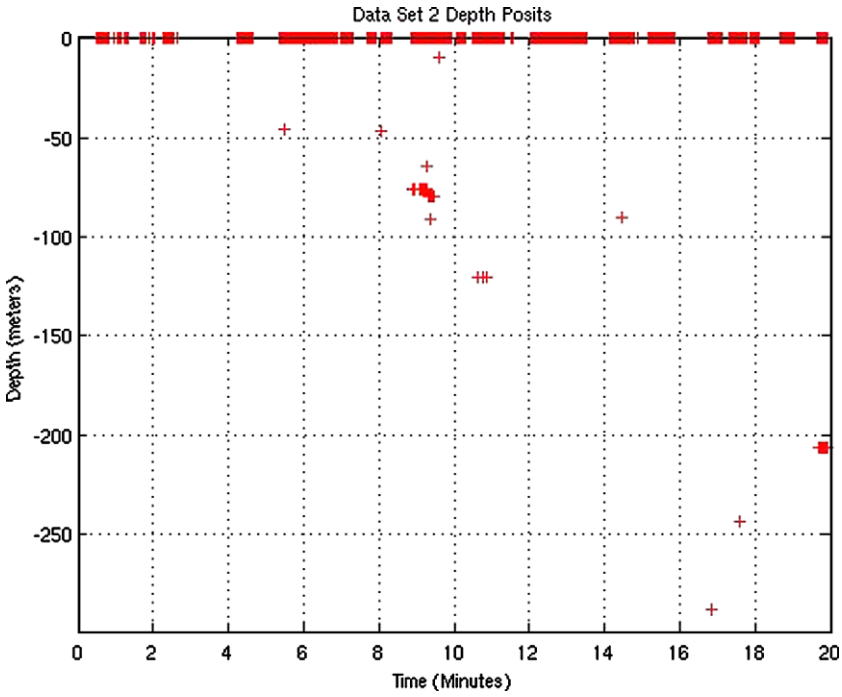

Fig. 12. Data set 2 depth (automated).

Table 1
Click arrival times

Hydrophone	Group 1		Group 3	
	TOA (MM:SS)	Delta-T (s)	TOA (MM:SS)	Delta-T (s)
J	00:20.029	0.342	03:49.523	0.327
K	00:20.958	0.262	03:50.265	0.265
I	00:21.062	0.250	03:50.274	0.270
G	00:21.212	0.744	03:50.528	0.784
H	00:23.521	0.243	03:52.827	0.251

The data set was corrected by subtracting 2.249501 s from the click arrival times on hydrophones G and H, and re-running the localization algorithms. The localization algorithm converged and produced the track shown in Figs. 13 and 14.

4. Summary

M3R has developed spectrogram-based algorithms for the passive detection and localization of marine mammals using widely spaced, bottom-mounted hydrophones characteristic of Navy undersea tracking ranges. They are applicable to any fixed or portable range with widely spaced sensors. These algorithms require repetitively vocalizing marine mammals with sufficient source levels to be detected on multiple hydrophones.

The detection algorithm combines a frequency domain energy detector with a thresholded click detector. The output of the energy detector is a binary frequency map derived

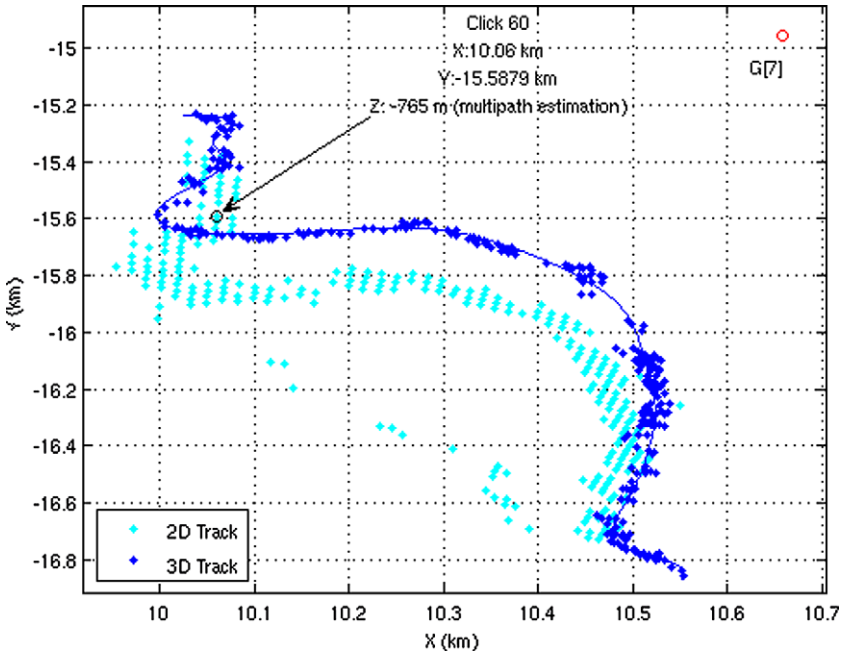

Fig. 13. 2D (fixed depth) and 3D tracks for data set 2. 2D track computed using hydrophones I, J, and K. 3D track has line running through it and was computed using all five hydrophones.

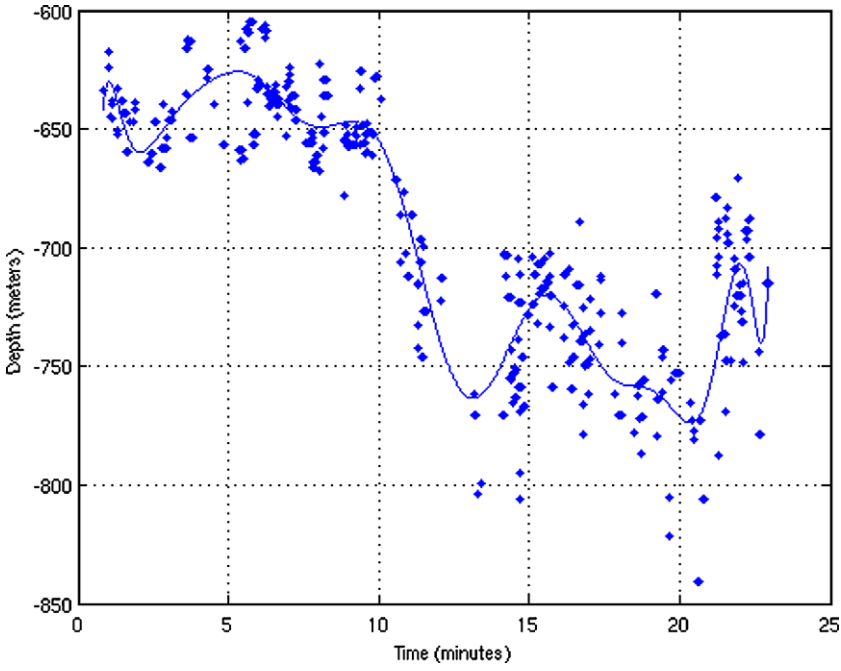

Fig. 14. Depth track for data set 2.

from the FFT. The presence of a click is determined by comparing the number of bins set in a single FFT window against a threshold. Detected click arrival times are passed to an association algorithm. The click associator cross-correlates patterns of click detections to determine time differences of arrivals (TDOAs). Positions are then determined using hyperbolic multilateration.

The algorithms were applied to two sperm whale (*Physeter macrocephalus*) datasets provided as part of a workshop on the detection and localization of marine mammals using passive acoustics, held in Monaco November 16–18, 2005. The first dataset consisted of multiple sperm whales vocalizing simultaneously. The second dataset was comprised of a single sperm whale with reverberation. Results from the application of the algorithms to both datasets were presented. The algorithms produced a combination of 2D and 3D positions for both datasets. A timing error in the second data set was discovered. This error was quantified and a corrected track produced.

Application of the software to the second conference data set produced a track on a single sperm whale after correcting for timing bias on two of the five hydrophones.

Acknowledgements

We would like to acknowledge our sponsors, Dr. Mardi Hastings and Dr. Robert Gisiner, at the Office of Naval Research. We would also like to thank the Atlantic Undersea Test and Evaluation Center (AUTEC) for access to their considerable infrastructure and ongoing support.

References

[1] Gillespie D. Detection and classification of right whale calls using an 'edge' detector operating on a smoothed spectrogram. Can Acoust 2004;32(2):39–47.
[2] Mellinger D. A comparison of methods for detecting right whale calls. Can Acoust 2004;32(2):39–47.
[3] Mellinger D. Automatic detection of regularly repeating vocalizations. J Acoust Soc Am 2005;118(3):1940.
[4] Mellinger DK, Clark CW. Bioacoustic transient detection by image convolution. J Acoust Soc Am 1993;93(4):2358.
[5] Mellinger DK, Stafford KM, Fox CG. Seasonal occurrence of sperm whale (*Physeter macrocephalus*) sounds in the Gulf of Alaska 1999–2001. Marine Mammal Sci 2004;20(1):48–62.
[6] Harland EJ, Armstrong MS. The real-time detection of the calls of cetacean species. Can Acoust 2004;32(2):39–47.
[7] Stafford KM, Fox CG, Clark DS. Long-range acoustic detection and localization of blue whale calls in the northeast Pacific Ocean. J Acoust Soc Am 1998;104(6):3616–25.
[8] Watkins WA, Schevill WE. Sound source location by arrival-times on a non-rigid three-dimensional hydrophone array. Deep-Sea Res 1972;19:691–706.
[9] Watkins WA. Biological sound-source locations by computer analysis of underwater array data. Deep-Sea Res 1976;23:175–80.
[10] Watkins WA, Schevill WE. Spatial distribution of *Physeter catodon* (sperm whales) underwater. Deep-Sea Res 1977;24:693–9.
[11] Freitag LE, Tyack PL. Passive acoustic localization of the Atlantic bottlenose dolphin using whistles and echolocation clicks. J Acoust Soc Am 1993;93(4):2197–205.
[12] Mitchell S, Bower J. Localization of animal calls via hyperbolic methods. J Acoust Soc Am 1995;97(5):3352–3.
[13] Tiemann CO, Porter MB, Frazer LN. Localization of marine mammals near Hawaii using an acoustic propagation model. J Acoust Soc Am 2004;115(6):2834–43.
[14] Spiesberger JL, Fristrup KM. Passive localization of calling animals and sensing of their acoustic environment using acoustic tomography. Am Natur 1990;135(1):107–53.

[15] Tiemann CO, Porter MB. A comparison of model-based and hyperbolic localization techniques as applied to marine mammal calls. J Acoust Soc Am 2003;114(4):2406.
[16] Watkins WA. In: Busnel Rene-Guy, Fish James F, editors. Acoustics and the behavior of sperm whales from animal sonar systems. Plenum Publishing Corporation; 1980.
[17] Wahlberg M, Frantzis A, Alexiadou P, Madsen PT, Mohl B. Click production during breathing in a sperm whale (*Physeter macrocephalus*). J Acoust Soc Am 2005;118(6):3404–7.
[18] Laplanche C, Adam O, Lopatka M, Motsch J-F. Male sperm whale acoustic behavior observed from multipaths at a single hydrophone. J Acoust Soc Am 2005;118(4):2677–87.
[19] Zimmer WMX, Tyack PL, Johnson MP, Madsen PT. Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis. J Acoust Soc Am 2005;117(3):1473–85.
[20] Backus RH, Schevill WE. Physeter clicks. In: Norris Kenneth S, editor. Whales dolphins and porpoises. Berkeley: Universirty of California; 1966. p. 510–28.
[21] Watkins WA, Schevill WE. Sperm whale codas. J Acoust Soc Am 1977;62(6):1485–90.
[22] http://www.accobams.org/events/isns_workshop/isns_workshop.html.
[23] Ward JA. 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].
[24] Vincent, H. Models, Algorithms, and measurements for underwater acoustic positioning, Ph.D. dissertation. University of Rhode Island, Kingston, RI; 2001.