ACCURACY ANALYSES OF PASSIVE TRACKING OF SEVERAL CLICKING SPERM WHALES
A Case of Complex Sources Binding

Frédéric Caudal and Hervé Glotin
System & Information Sciences Laboratory (LSIS - UMR CNRS 6168), Université du Sud Toulon Var
BP 20132, 83957 La Garde Cedex, France

Keywords: Delay estimation, marine mammals, acoustic tracking, Cramér-Rao Lower Bound, acoustic propagation.

Abstract: This paper provides a real-time passive underwater acoustic method to track multiple emitting whales using four or more omni-directional widely-spaced bottom-mounted hydrophones and to evaluate the performance of the system via the Cramér-Rao Lower Bound (CRLB) and Monte Carlo simulations. After a 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 or linear sound speed profile. The complete algorithm is tested on real data from the NUWC and the AUTEC. The CRLB and Monte Carlo simulations are computed and compared with the tracking results. Our model is validated by similar results from the US Navy and Hawaii univ labs in the case of one whale, and by similar whales counting from the Columbia univ. ROSA lab in the case of multiple whales. At this time, our tracking method is the only one giving typical speed and depth estimations for multiple (5) emitting whales located at 1 to 5 km from the hydrophones.

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 institutes like the AUTEC[1] and the NUWC[2]. Motivation for processing MM signals stems from increasing interest in the behavior of endangered MM[3]. 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 in tracking an unknown number of sperm whales (Physeter catodon). Clicks are recorded on two datasets of 20 and 25 minutes on a 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. In Section 2 we briefly review studies of source separation methods and the main characteristics of MM signals and we propose a real-time algorithm for MM transient call localization. In Section 3, we recall briefly the CRLB (Kay, 1993) and the confidence ellipses theory. In Section 4 we show and compare results of tracks estimates with results from specialized teams and the performance system with the CRLB.

[1]Atlantic Undersea Test & Evaluation Center - Bahamas

[2]Naval Undersea Warfare Center of the US Navy

[3]This work is founded by the "Conseil régional Provence-Alpes-Côte d'Azur" France, and partly by Chrisar Software Inc.

2 PROBLEM FORMULATION

This papers deals with the 3D tracking of MM using a widely-spaced bottom-mounted array in deep water. It focuses on sperm whale clicks; detection and classification are not a concern. There were previous algorithms developed in the state of art (Giraudet and Glotin, 2006b; Giraudet and Glotin, 2006a; Nosal and Frazer, 2006; Morrissey et al., 2006) 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. 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

Caudal F. and Glotin H. (2008).
ACCURACY ANALYSES OF PASSIVE TRACKING OF SEVERAL CLICKING SPERM WHALES - A Case of Complex Sources Binding.
In *Proceedings of the International Conference on Signal Processing and Multimedia Applications*, pages 55-62
DOI: 10.5220/0001933700550062

hydrophones include multipath effects (and reverberations), which create secondary peaks in the Cross-Correlation (CC) function that the generalized CC methods cannot eliminate.

2.1 Material

Table 1: Hydrophones positions: D=Datasets, Dist=Distance to barycenter (m).

D	Hydros	Dist	X (m)	Y (m)	Z (m)
D1	H 1	5428	18501	9494	-1687
	H 2	4620	10447	4244	-1677
	H 3	2514	14119	3034	-1627
	H 4	1536	16179	6294	-1672
	H 5	3126	12557	7471	-1670
	H 6	4423	17691	1975	-1633
D2	H 7	1518	10658	-14953	-1530
	H 8	4314	12788	-11897	-1556
	H 9	2632	14318	-16189	-1553
	H 10	3619	8672	-18064	-1361
	H 11	3186	12007	-19238	-1522

The signals are records from the ocean floor near Andros Island - Bahamas (Tab.1), provided with celerity profiles and recorded in March 2002. Datasets are sampled at 48 kHz and contain MM clicks and whistles, 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 $c = 1500ms^{-1}$ or a linear profile with $c(z) = c_0 + gz$ where z is the depth, $c_0 = 1542ms^{-1}$ is the sound speed at the surface and $g = 0.051s^{-1}$ is the gradient. Sound source tracking is performed by continuous localization in 3D using Time Delays Of Arrival (T) estimation from four hydrophones.

2.2 Signal Filtering

A sperm whale click is a transient increase of signal energy lasting about 20 ms (Fig.1-a). Therefore, we use the Teager-Kaiser (TK) energy operator on the discrete data:

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

where n denotes the sample number. An important property of TK 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 $s(n)$ as:

$$s(n) = x(n) + u(n),$$

where $x(n)$ 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 TK to $s(n)$, $\Psi[s(n)]$ is:

$$\Psi[s(n)] \approx \Psi[x(n)] + w(n),$$

where $w(n)$ is a random gaussian process (Kandia and Stylianou, 2006). The output is dominated by the clicks energy. Then, we reduce the sampling frequency to 480Hz by the mean of 100 adjacent bins to reduce the variance of the noise and the data size. We apply the Mallat's algorithm (Mallat, 1989) with the Daubechies wavelet (order 3). We chose this wavelet for its great similarity to the shape of a decimated click (Giraudet and Glotin, 2006b; Giraudet and Glotin, 2006a). The signal is denoised with a universal thresholding (Donoho, 1995) defined as $D(u_k, \lambda) = sgn(u_k)max(0, |u_k| - \lambda)$, with u_k the wavelet coefficients, $\lambda = \sqrt{(}2log_e(Q))\sigma_N\sigma_{\tilde{N}}$, and Q the length of the signal resolution level to denoise. 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 without any parameter. Fig.1-d is the filtered signal on multiple (Fig.1-c) emitting MMs.

2.3 Rough $\widetilde{T}$ Estimation

First, T estimates are based on MM click realignment only. Every 10s, 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 $T(i, j) = t_j - t_i$. Its estimate $\widetilde{T}(i, j)$ is calculated by CC of 10s chunks (2s shifting) of the filtered signal for hydrophones i and j (Giraudet and Glotin, 2006b; Giraudet and Glotin, 2006a). We keep the 35 highest peaks on each CC to determine the corresponding $\widetilde{T}(i, j)$. The filtered signals give a very fast rough estimate of T (precision $\pm$ 2ms). Fig.(1.e) shows the CC with the raw signal and (1.f) with the filtered signal. Without filtering, CC generates spurious delays estimates and the tracks are not correct. The maximum $\widetilde{T}$ rank (Fig.2) in D1, pitching the source localization, are high among the 35 $\widetilde{T}$ kept in the CC which justifies this number.

2.4 $\widetilde{T}$ Selection and Localization with a Constant Profile

Each signal shows echoes for each click (Fig.1 b), maybe due to the reflection of the click train off the

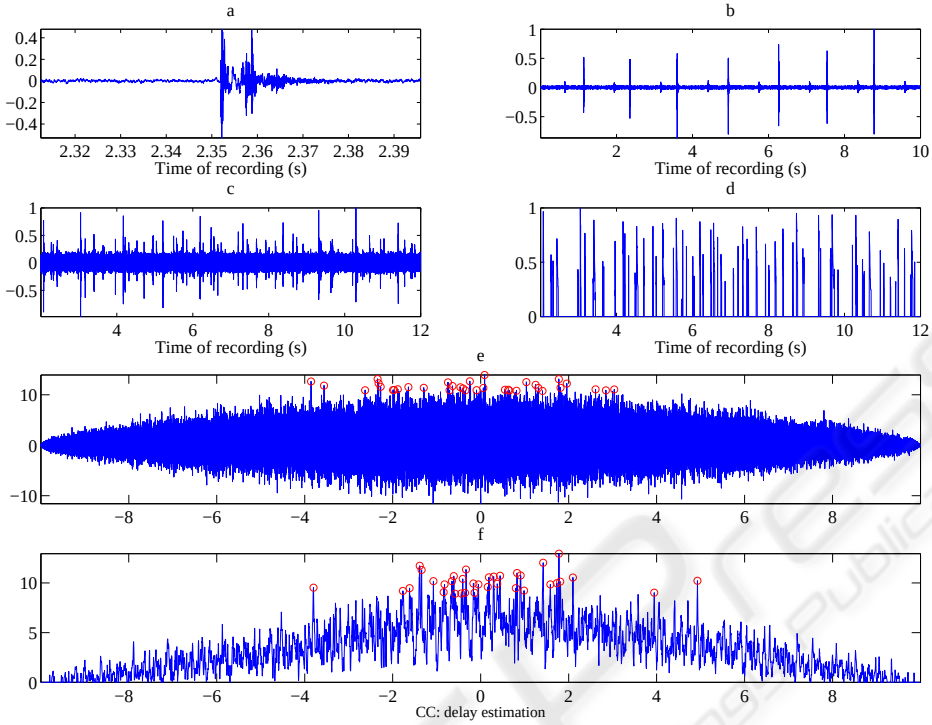

Figure 1: (a): Detail of a click. (b): raw signal (D2) of hydrophone 7 (H7) during the first 10s of recording, containing 7 clicks and their echoes. (c): raw signal (D1) of H3 during the first 10s of recording showing multiple emissions. (d): (c) after filtering. (e): CC between (c) and corresponding raw signal chunk of H1. (f): idem than (e) but with (d). Circles correspond to the 35 maximum peaks.

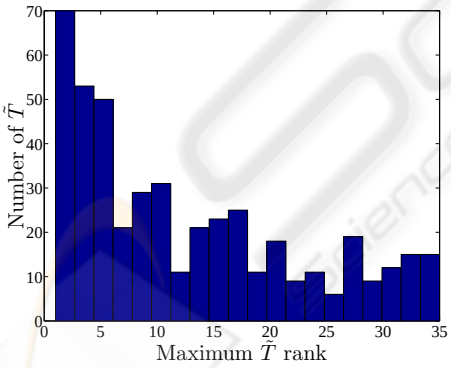

Figure 2: Maximum $\widetilde{T}$ rank histogram in the CC for each triplet.

ocean surface or bottom or different water layers. We use a method based on autocorrelation (Giraudet and Glotin, 2006b; Giraudet and Glotin, 2006a) to eliminate it. Then, thanks to the $\widetilde{T}$ transitivity system described in (Giraudet and Glotin, 2006b) we keep $\widetilde{T}$ triplets coming from the same source. Finally, thanks to the measured delays and an acoustic model based on a constant sound speed profile, the least squares cost function determines the MM positions using a multiple non linear regression with Gauss-Newton method (Levenberg-Marquardt) (Giraudet and Glotin, 2006b). The residuals are approximatively following a Chi-square distribution with $Nc-d$ degrees of freedom, noted X^2_{Nc-d}, Nc is the number of hydrophones couples considered and d the number of unknowns, here 3 (x,y,z). The position is accepted if the residual is inferior to a threshold x, That is calculated solving $P = prob(X^2_{Nc-d} > x)$ with $P = 0.01$ (we keep 99% of the estimates).

2.5 Source Localization with a Linear Sound Speed 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 (Fig.3). c_s is the sound speed at the source, θ_s is the launch angle of the ray at the source, measured relatively to the horizontal. From the geometry shown in Fig.3, the center of the circle, (xc,zc), along which the ray path is an arc, is (White et al., 2006):

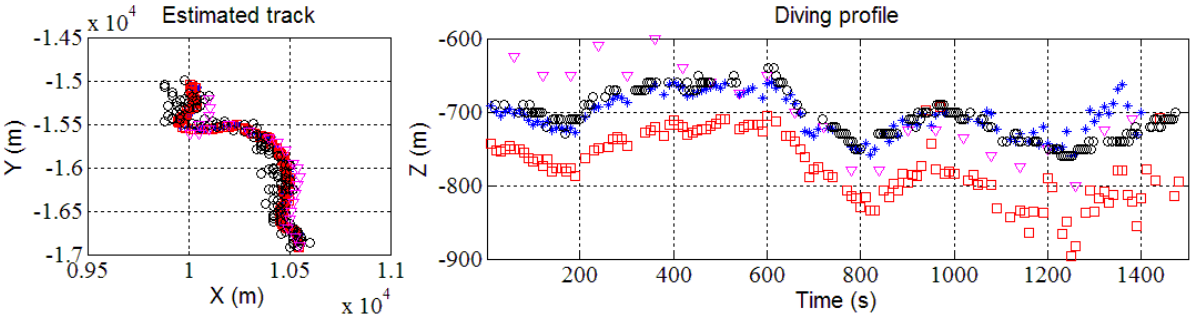

Figure 4: Plan view and diving profile of the MM in D2, our estimates with a linear (⋆) (Caudal and Glotin, 2008) or constant profile (□); and estimates from Morrissey's (Morrissey et al., 2006) (▽) and from Nosal's (Nosal and Frazer, 2006) methods (o).

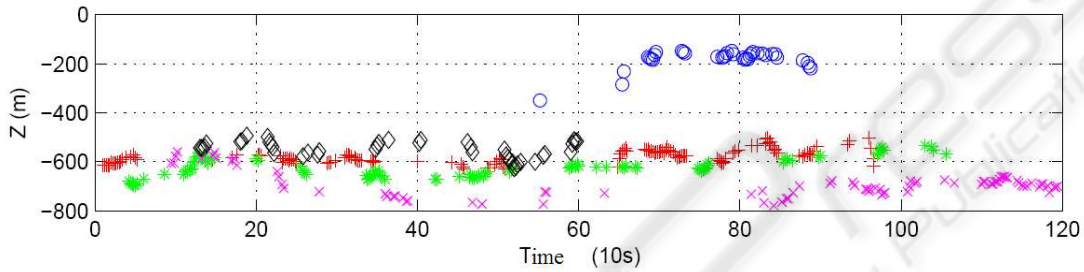

Figure 5: Averaged diving profile in D1. Each symbol corresponds to one whale in Fig.6 (Bottom to top (+), (o), (⋆), (x), (⋄)).

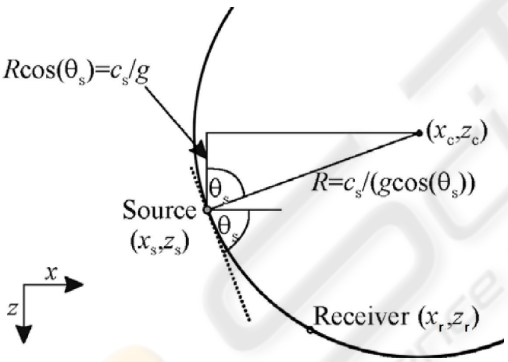

Figure 3: Geometry for a source and receiver in a linear profile (White et al., 2006).

$$x_c = \frac{x_s + x_r}{2} + \frac{(z_s - z_r)}{2(x_s - x_r)}(z_r - z_s + \frac{2c_s}{g}),$$
$$z_c = z_s - \frac{c_s}{g}. \quad (2)$$

For linear sound speed profile the course time τ of the ray is:

$$\tau = \frac{1}{g}\left\{\log\left(\frac{z_c - z_s}{z_c - z_r}\right) - \log\left(\frac{R + x_c - x_s}{R + x_c - x_r}\right)\right\}. \quad (3)$$

Using Eqs.(2)-(3) allows one to compute the propagation time from the source to any receiver and then the whale position.

3 THE CRLB AND CONFIDENCE REGIONS

3.1 The CRLB

The CRLB provides the maximum acurracy for the estimation of the position. Considering a constant sound speed profile, the function model of the TDOA is defined:

$$s(\theta) = \frac{1}{c_s}\left\{\sqrt{\sum_{k=1}^{3}(X_{i,k} - \theta_k)^2} - \sqrt{\sum_{k=1}^{3}(X_{j,k} - \theta_k)^2}\right\}.$$

where $X_{\{i,j\}}$ is the vector coordinate of hydrophone $\{i, j\}$, θ is the unknown parameters vector $[x\ y\ z]^T$ and c_s the celerity. Thus, considering the TDOA noise Gaussian and B its variance-covariance matrix, the Fisher Information matrix is:

$$I_\theta = \nabla_\theta s(\theta) B^{-1} \nabla_\theta^T s(\theta).$$

Then, the CRLB is $B_\theta = I_\theta^{-1}$.

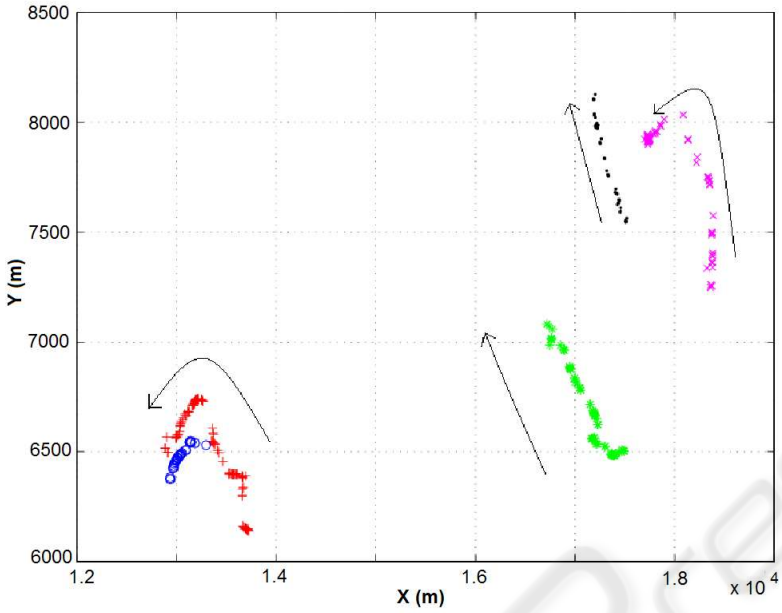

Figure 6: Plan view in D1. Each symbol corresponds to one of the 5 whales. The arrows stress the directions.

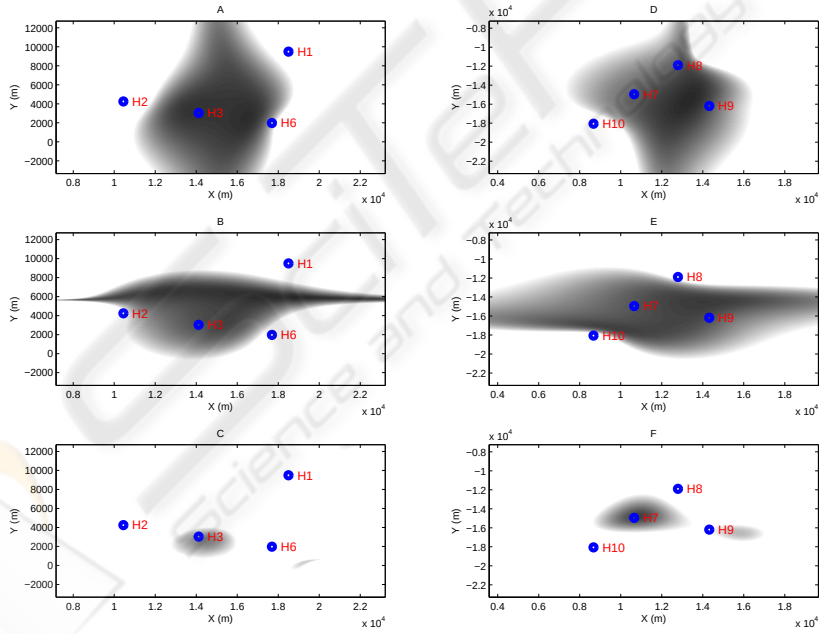

Figure 7: CRLB values scaled in gray colors with a plan view: black means a null CRLB and white a CRLB≥10 m. For all the figures, a depth of 500m was chosen. (A): CRLB on x values, dataset D1. (B): y, D1. (C): z, D1. (D): x, D2. (E): y, D2. (F): z, D2.

3.2 Confidence Regions

The confidence ellipse generalizes the notion of confidence interval in case of gaussian vector in two dimensions. Considering a stochastic gaussian vector $X \sim N(\mu, \Sigma)$ of two dimensions with $\Sigma > 0$. We have to determine the level curves of the gaussian distributionl, i.e. the points x of the ellipse

$$(x-\mu)^{*}\Sigma^{-1}(x-\mu) = r$$

Being a symmetrical positive matrix, Σ can be spectraly factorized as $\Sigma = PDP^*$ where P is the orthonormal matrix of the eigen vectors and D the diagonal matrix of the eigen values. For each x, the point $\tilde{x}$ is defined with $\tilde{x} = D^{-1/2}Px$, then $|\tilde{x}|^2 = r$, i.e. all the $\tilde{x}$ is the circle centered with radius $r : \tilde{x}(\rho) = r[\cos(\rho)\ \sin(\rho)]^*$ for $0 \leq \rho < 2\pi$. The parametric equation of the ellipse is:

$$x(\rho) = \mu + rP^*D^{1/2} \begin{bmatrix} \cos(\rho) \\ \sin(\rho) \end{bmatrix}, 0 \leq \rho < 2\pi$$

The quadratic norm of the stochastic vector $\tilde{X} = D^{-1/2}PX$ is distributed with the chi-square distribution with 2 degrees of freedom. Putting r = 5.99, then

$$\mathbb{P}(X \in x \in \mathbb{R}^2 : (x-\mu)^*\Sigma^{-1}(x-\mu) \leq r) = 0.95$$

This means the real position is 95% likely to be inside the ellipse.

4 RESULTS

4.1 Tracking Comparisons

For D2 (Fig.4), a constant and a linear sound speed profile were used (Caudal and Glotin, 2008) and the results are similar with the Morrissey's (Morrissey et al., 2006) and Nosal's (Nosal and Frazer, 2006) methods. The diving profile underlines a bias of about 50 to 100m between the linear and the constant profiles results, emphasizing 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 on D1 are shown in Fig.6-5 for a linear sound speed profile. We thus localize 5 MM. The confidence regions are computed for the two datasets with a Monte Carlo method. The ellipses maxima (30m) fit with MM length (20m).

4.2 CRLB Computation

The CRLB provide the maximum acurracy for the localization. Here, we computed the CRLB (in meter) in the space (x,y,z) and plot the values for both datasets (Fig.7). We consider that the standard deviation of the noise is equal to the quantification noise with a sampling frequency of 480Hz. The main dependencies of the bounds are the noise and the array configuration.

In figures 8 and 9, the CRLB on y and z is shown for a depth of 1000m, and is just about the same for a depth of 500m. Thus, as the mean depth of a sperm whale is in this range, the CRLB likewise.

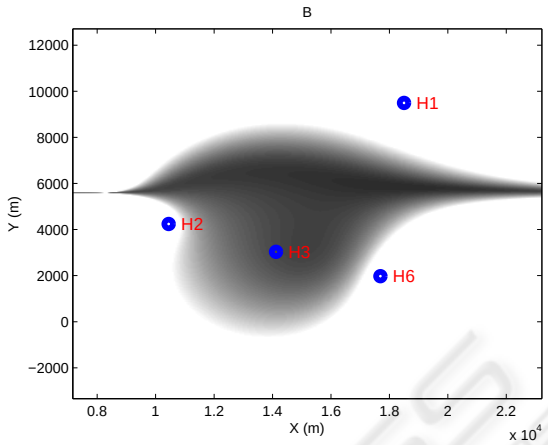

Figure 8: CRLB on y axis, plan view, dataset D1, depth=1000m.

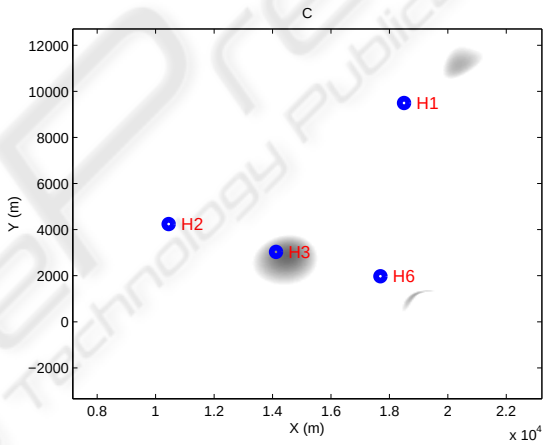

Figure 9: CRLB on z axis, plan view, dataset D1, depth=1000m.

4.3 The Confidence Regions

We apply a Monte Carlo method with a gaussian distribution noise with the standard deviation described above. For each $\widetilde{T}$ realization, the source 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 mean values of the confidence intervals on X, Y, Z axes are about 18, 16 and 30 m (Fig.10). 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

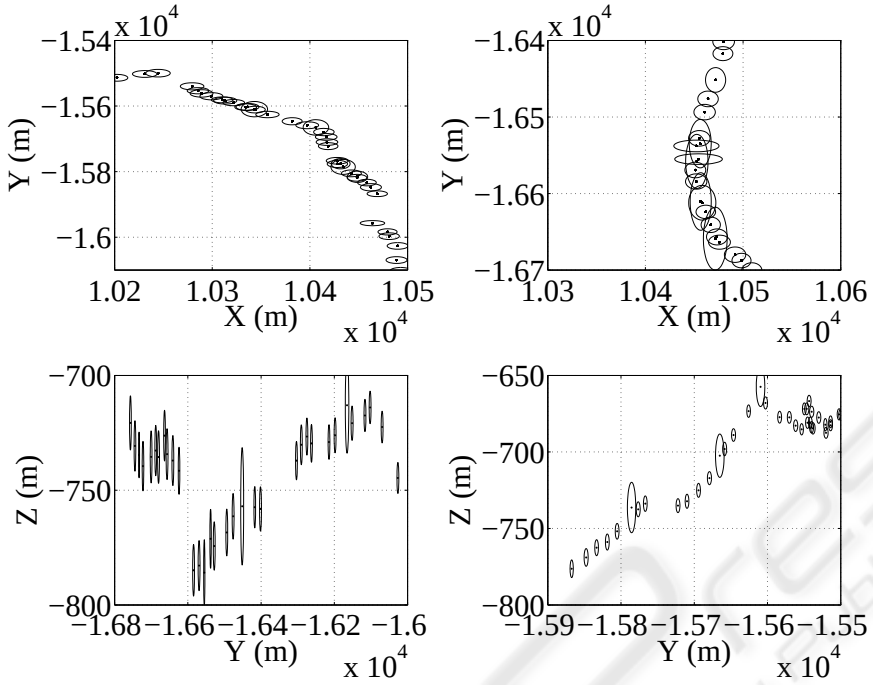

Figure 10: Confidence regions projection on X and Y and on Z and Y axes for D2 trajectory.

are meaningfully higher than the other axes because the distance between each hydrophones in this direction (maximum difference on the Z axis between hydrophones is 200m) is smaller. The D1 results obtained with a linear profile (Fig.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 (Fig.12), while the whales close to the center (Fig.11) exhibit an error of about 10 to 20m like for D2 (Fig.6). Those uncertainties are reasonable according to the sperm whale length.

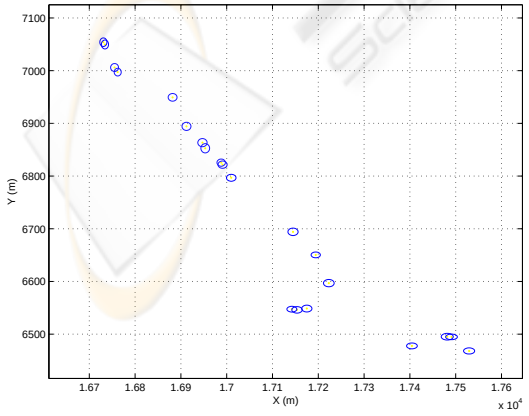

Figure 11: Confidence regions projection on X and Y for the whale in the middle (* symbol), D1 trajectory.

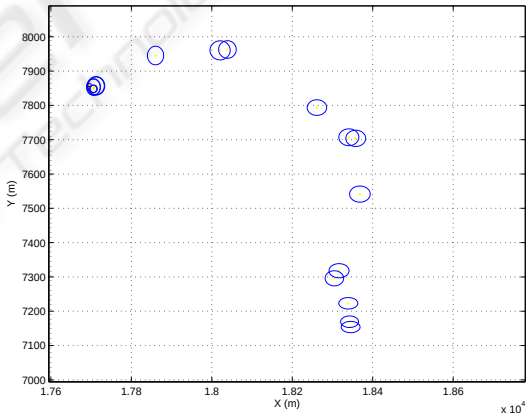

Figure 12: Confidence regions projection on X and Y for the whale in the right (x symbol), D1 trajectory.

4.4 Performance Comparisons

Comparing the CRLB and the ellipses, we denote the correlation between maximum accuracy and confidence regions. Figures 7-C and 7-F show the accuracy on z axis is $\geq$10 m for both datasets in the tracking regions which is consistent with the ellipses results. The CRLB (in D1, figures 6-D,E) also explains that the farthest whale has larger confidence regions. But for both datasets, the CRLB on x and y is far inferior to 10m inside the array whereas our ellipses are about 10 to 20m. Maybe other parameters like the approxi-

mated celerity profile are involved in.

5 CONCLUSIONS

The tracking algorithm presented in this paper is real-time on a standard laptop and works for one or multiple emitting sperm whales. Depth results with constant speed contains a bias errors due to the refraction of the sound paths from the MM to the receivers what a linear speed corrects. An other way to tackle the speed profile issue (Glotin et al., 2008) is to estimate it as a fourth unknown in the regression. Our algorithm has no species dependency as long as it processes all transients. At this time, only our algorithm gives results with typical speed 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 estimation based on click clustering, averaged number of MM for each 5min chunks on D1 is $[4.3; 5.3; 4; 3.6]$ (Halkias and Ellis, 2006) similar to ours ($[4; 4; 4; 3]$). The localization accuracy is computed via the CRLB and the confidence ellipses which are correct considering the MM length. The high depth error is mainly due to the low precision reachable with the CRLB considering the array configuration. Our method provides thus robust online detecting/counting system of clicking MM in open ocean.

REFERENCES

Caudal, F. and Glotin, H. (2008). Multiple real-time 3d tracking of simultaneous clicking whales using hydrophone array and linear sound speed profile. In *ICASSP IEEE*, volume 4p.

Donoho, D. L. (1995). De-noising by soft thresholding. *IEEE Trans. IT*, 41:613–627.

Giraudet, P. and Glotin, H. (2006a). Echo-robust and real-time 3d tracking of marine mammals using their transient calls recorded by hydrophones array. In *ICASSP IEEE*.

Giraudet, P. and Glotin, H. (2006b). Real-time 3d tracking of whales by echo-robust precise tdoa estimates with a widely-spaced hydrophone array. *Applied Acoustics*, 67:1106–1117.

Glotin, H., Caudal, F., and Giraudet, P. (2008). Whales cocktail party: a real-time tracking of multiple whales. *International Journal Canadian Acoustics*, page 7p.

Halkias, X. and Ellis, D. (2006). Estimating the number of marine mammals using recordings form one microphone. In *ICASSP IEEE*.

Kandia, V. and Stylianou, Y. (2006). Detection of sperm whale clicks based on the teager-kaiser energy operator. *Applied Acoustics*, 67:1144–1163.

Kay, S. (1993). *Fundamentals of statistical signal processing*. PTR.

Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. volume 11, pages 674–693. IEEE Transaction on Pattern Analysis and Machine Intelligense.

Morrissey, R., Ward, J., DiMarzio, N., Jarvisa, S., , and Moretti, D. (2006). Passive acoustics detection and localization of sperm whales in the tongue of the ocean. *Applied Acoustics*, 62:1091–1105.

Nosal, E. and Frazer, L. (2006). Delays between direct and reflected arrivals used to track a single sperm whale. *Applied Acoustics*, 62:1187–1201.

White, P., Leighton, T., Finfer, D., Powles, C., and Baumann, O. (2006). Localisation of sperm whales using bottom-mounted sensors. *Applied Acoustics*, 62:1074–1090.