WHALE COCKTAIL PARTY: REAL-TIME MULTIPLE TRACKING AND SIGNAL ANALYSES

Hervé Glotin[1]**, Frédéric Caudal**[1]**, Pascale Giraudet**[2]
1-System & Information Sciences Laboratory (LSIS - UMR CNRS 6168)
2-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 underwater acoustic method to track multiple emitting 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 multiple tracking algorithm is developed. 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, linear sound speed profile or an estimated. The complete algorithm is tested on real data from the NUWC[1] and the AUTEC[2]. Our model is validated by similar results from the US Navy[3] and SOEST[4] Hawaii univ labs in the case of one whale, and by similar whales counting from the Columbia univ. ROSA[5] 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 emitting whales.

RESUME

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. Etant 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, selectionnées, 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 NUWC[1] et de l'AUTEC[2]. Notre modèle est validé par des résultats similaires de l'US Navy[3] et du laboratoire SOEST[4] de l'université d'Hawaii dans le cas d'émissions simples, et par une estimation du nombre de baleines du laboratoire ROSA[5] 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 the AUTEC and the 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 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.

This papers 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 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 im-

[1] Naval Undersea Warfare Center of the US Navy
[2] Atlantic Undersea Test & Evaluation Center - Bahamas
[3] NUWC
[4] School of Ocean and Earth Science and Technology
[5] Recognition and Organization of Speech and Audio

prove 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 - Bahamas[3], 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 $c = 1500ms^{-1}$ and estimated celerity profile, 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 (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]:

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

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 which 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's 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 $D(u_k, \lambda) = sgn(u_k)max(0, |u_k| - \lambda)$, with u_k the wavelet coefficients, $\lambda = \sqrt{(2log_e(Q))}\sigma_N\sigma_{\tilde{N}}$ and Q is the length of the resolution level of the signal to denoise [1]. The noise variance σ_N is calculated on each 10s windows on the raw signal with a maximum likelihood criterion. $\sigma_{\tilde{N}}$ is the variance of the wavelet coefficients on a resolution level of a generated, reduced and centered 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 ($\widetilde{T}$) 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 $T(i, j) = t_j - t_i$. Its estimate $\widetilde{T}(i, j)$ 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 (Nb_T) highest peaks on each CC to determine the corresponding $\widetilde{T}(i, j)$ (see Figure 1 for detail) . The filtered signals give a very fast rough estimate of $\widetilde{T}$ (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 $\widetilde{T}$. Without filtering, CC generates spurious delay estimates and the tracks are not correct. The raw CC shows more $\widetilde{T}$ 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 for the detection of additional $\widetilde{T}$ 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 $\widetilde{T}$ correspond to a multiple of the echo.
For each pair of hydrophones (i, j), all $\widetilde{T}_a(i, j)$ satisfying one of the following equations are removed, $k \in \{1..4\}, a \in \{2..Nb_T\}$:

$$\widetilde{T}_a(i,j) - \widetilde{T}_1(i,j) = k * E(i) \pm \xi,$$
$$\widetilde{T}_a(i,j) - \widetilde{T}_1(i,j) = -k * E(j) \pm \xi.$$

where $\xi = 2ms$.

2.4 $\widetilde{T}$ transitivity and filtering

Once many $\widetilde{T}$ for each pair of hydrophones have been eliminated, the remaining $\widetilde{T}$ are combined every 10s to select all quadruplets of hydrophones whose $\widetilde{T}$ independent triplet correspond to the same source. Thus we consider that a quadruplet of hydrophones (i, j, k, h) localized the same source with

[3] 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);H1=(8672,-18064,-1361);H11=(12007,-19238,-1522)

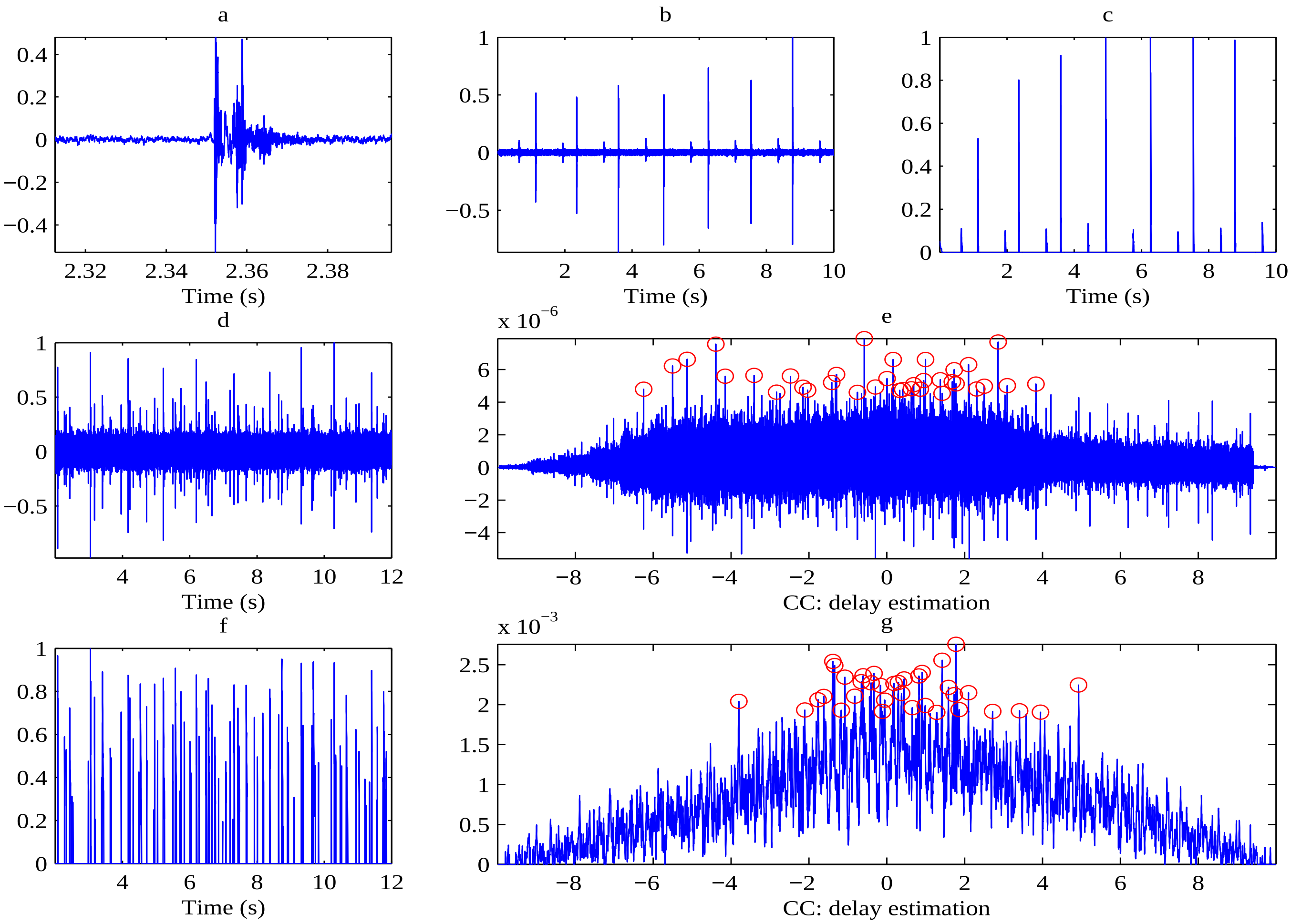

Figure 1: (a): detail of a click on the normalized .wav file format. (b): raw signal (D2) of hydrophone 7 (H7) during the first 10 seconds of recording, containing 7 clicks and their echoes. (c): (b) after filtering. (d): raw signal (D1) of H3 during the first 10 seconds of recording showing multiple 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.

the $\widetilde{T}_{a,b,c,d,e,f}$ if the 4 following equations are verified [3, 2] for each time t:

$$\begin{aligned}
\widetilde{T}_a(i,j)+\widetilde{T}_b(j,k) &= \widetilde{T}_d(i,k)\pm\delta,\\
\widetilde{T}_a(i,j)+\widetilde{T}_c(j,h) &= \widetilde{T}_f(i,h)\pm\delta,\\
\widetilde{T}_d(i,k)+\widetilde{T}_e(k,h) &= \widetilde{T}_f(i,h)\pm\delta,\\
\widetilde{T}_b(j,k)+\widetilde{T}_e(k,h) &= \widetilde{T}_c(j,h)\pm\delta.
\end{aligned}$$

$\widetilde{T}$ has been estimated with 2 ms precision, moreover $\widetilde{T}$ transitivity only works for an isospeed model which means sound rays move in a straight line. We consider the error $\delta = 6ms$. The distribution of the maximum $\widetilde{T}$ 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.

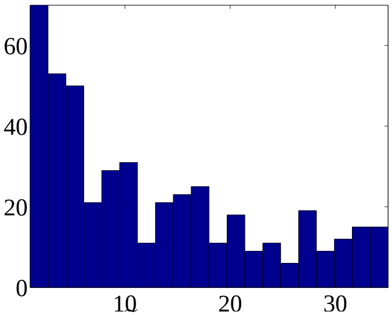

Figure 2: Maximum $\widetilde{T}$ CC rank histogram for each triplet

The three independent $\widetilde{T}$ of each hydrophones $(i,\ j,\ k,\ h)$ quadruplet measured on the windows t are noted:

$$\{\widetilde{T}_a(i,j,t), \widetilde{T}_d(i,k,t), \widetilde{T}_f(i,h,t)\}.$$

The modeled delays are:

$$T_a(i,j,t) = \frac{\|X_t,H_i\| - \|X_t,H_j\|}{c},$$
$$T_d(i,k,t) = \frac{\|X_t,H_i\| - \|X_t,H_k\|}{c}, \quad (2)$$
$$T_f(i,h,t) = \frac{\|X_t,H_i\| - \|X_t,H_h\|}{c},$$

where $\| \ \|$ denotes the Euclidian norm. We assume that the precision errors of the T due to the decimation are modeled with a Gaussian, centered, additive, and uncorrelated noise between sensors, noted εconsidered the same on each of the windows t and with a variance $\sigma^2 = (\frac{\xi}{3})^2$ (σ contains 68% of the Gaussian distribution).

$$\begin{aligned}\widetilde{T_a}(i,j,t) &= T_a(i,j,X_t) + \varepsilon_{i,j,t},\\ \widetilde{T_d}(i,k,t) &= T_d(i,k,X_t) + \varepsilon_{i,k,t},\\ \widetilde{T_f}(i,h,t) &= T_f(i,h,X_t) + \varepsilon_{i,h,t},\end{aligned} \quad (3)$$

X_t is estimated with a least square method. The least square criteria to minimize is given by:

$$\begin{aligned}Q(X_t) &= \frac{1}{2}\left[\frac{\widetilde{T_a}(i,j,t) - T_a(i,j,X_t)}{\sigma^2}\right]^2\\ &+ \frac{1}{2}\left[\frac{\widetilde{T_d}(i,k,t) - T_d(i,k,X_t)}{\sigma^2}\right]^2\\ &+ \frac{1}{2}\left[\frac{\widetilde{T_f}(i,h,t) - T_f(i,h,X_t)}{\sigma^2}\right]^2.\end{aligned}$$

This case is a non linear criteria minimization. Indeed, $Q(X_t)$ contains the non linear function $\| \ \|$ (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 $\Delta = 10^{-6}$.

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. Five hydrophones are necessary, which is the case in D2, to calculate four independent $\widetilde{T}$. 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.

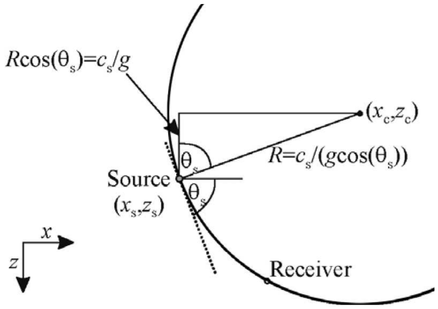

Figure 3: Geometry for a source and receiver in a linear sound speed profile [13]

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 (xr, zr). From the geometry shown in Figure 3, the center of the circle, (xc, zc), along which the ray path is an arc, can be shown to be:

$$\begin{aligned}x_c &= \frac{x_s + x_r}{2} + \frac{(z_s - z_r)}{2(x_s - x_r)}(z_r - z_s + \frac{2c_s}{g}),\\ z_c &= z_s - \frac{c_s}{g}.\end{aligned} \quad (4)$$

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.

3 Results

For D2, three sound speed profile were used:a constant; or an estimated; or a linear. The results are compared with the Morrissey's [11] and Nosal's [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).

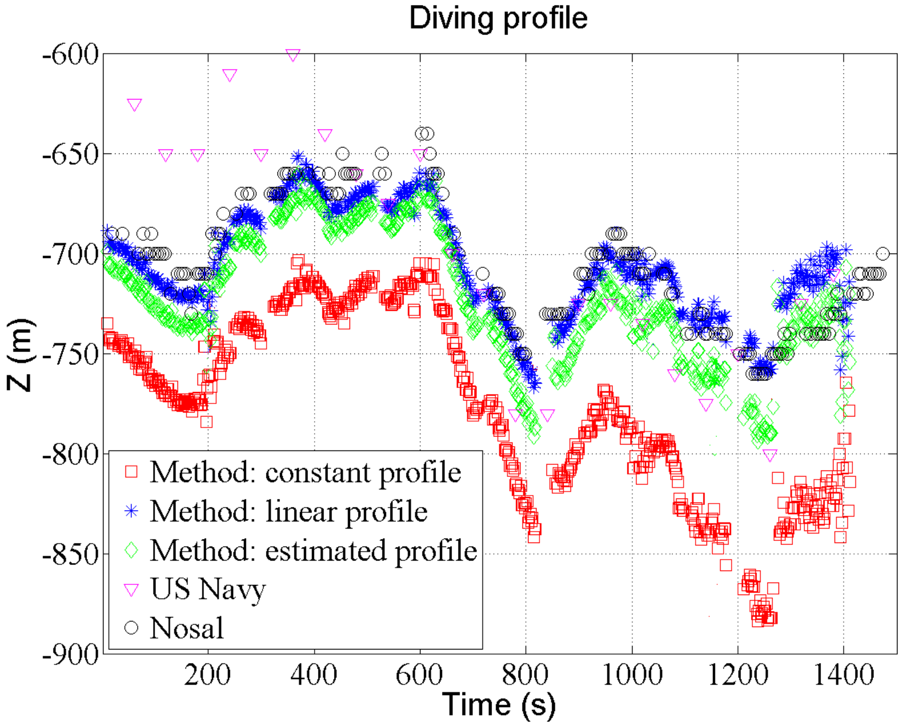

Figure 5: Diving profile of the MM in D2, our estimates with a linear (⚹), a constant profile (□) and an estimated profile (◊); and estimates from Morrissey's [11] (▽) and from Nosal's [12] methods (o).

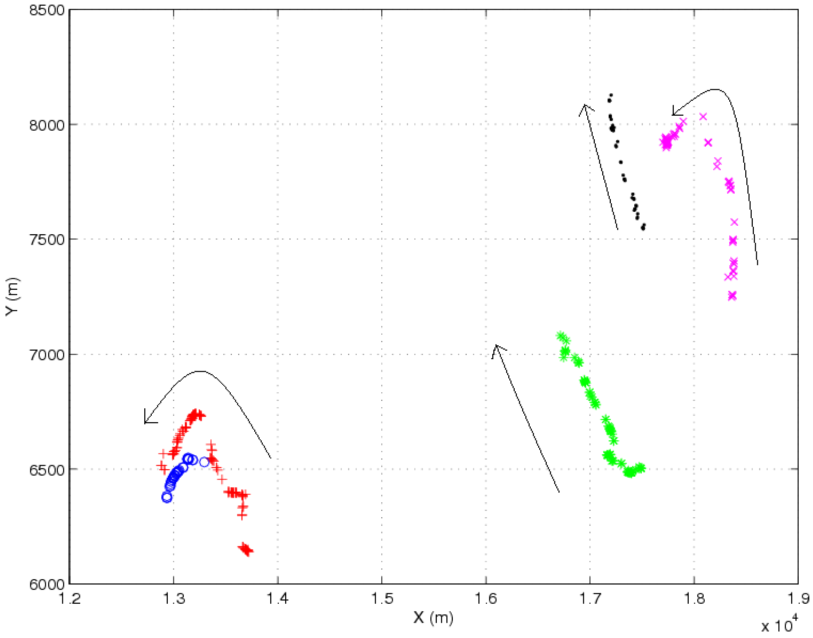

Figure 6: Plan view in D1. Each symbol correspond to one of the five whales. The arrows stress the directions of each whale. See Figure 7 for their diving profile. Whale 1:(o), 2:(+), 3:(⚹), 4:(·), 5:(x). Recording duration: 20min.

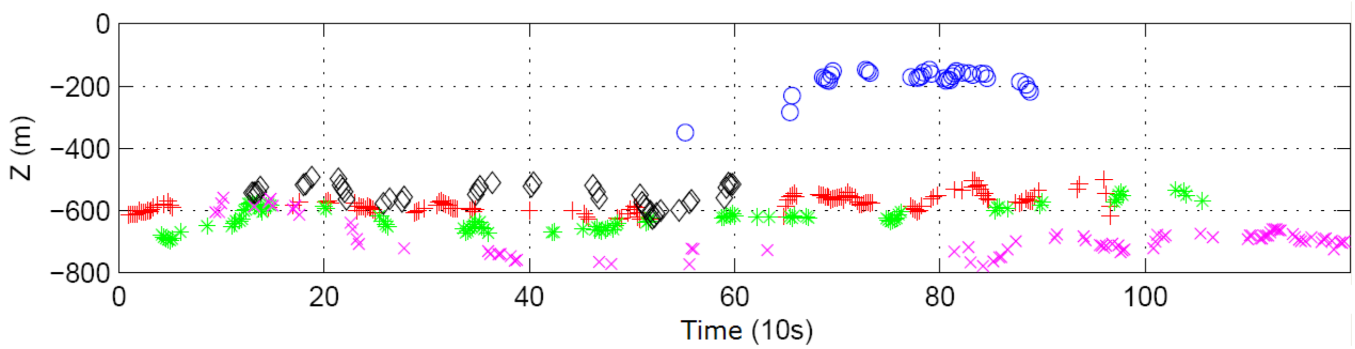

Figure 7: Averaged diving profile in D1. Each symbol correspond to one of the five whales. Whale 1:(o), 2:(+), 3:(∗), 4:(·), 5:(x).

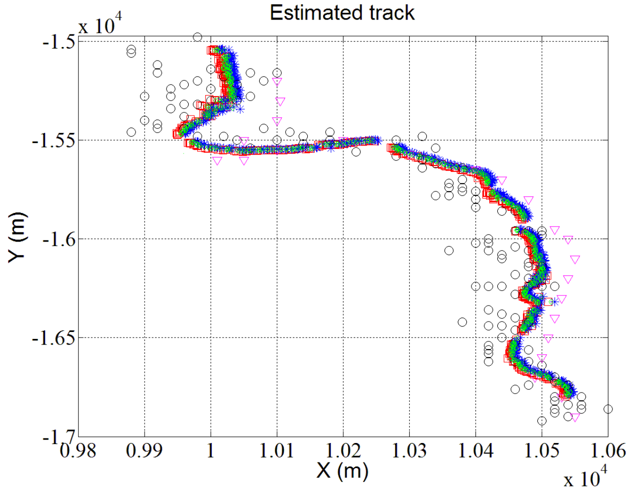

Figure 4: Plan view of the MM in D2, our estimates with a linear (∗), a constant profile (□) and an estimated profile (◊), threesome are almost merged; and estimates from Morrissey's [11] (▽) and from Nosal's [12] methods (o). Note the variance of the positions with Nosal's method. The whale direction 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 noise of engine boats.

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 raw 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 raw is the difference between PIMC and ROSA Lab estimations.

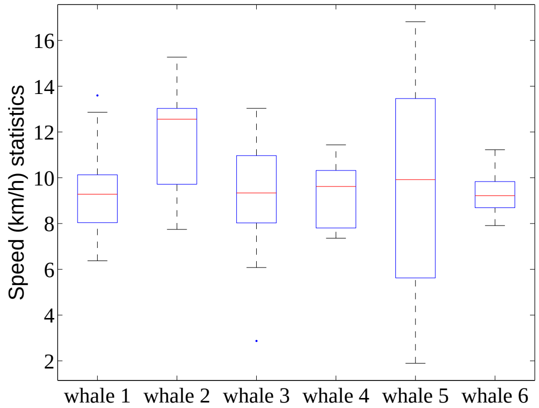

Figure 8: Speed (averaged on 30s windows) statistics on the whole set for each whale in D1 (whales 1 to 5) and D2 (whale 6). The central line of the box is the median of speed and the lower and higher lines are the quartiles. The whiskers show the extent of the speed. Whale 5 seems to stop a moment at the end of the track (See Figure 7).

3.1 The confidence regions

In section 2.5, because we consider a gaussian distribution, the standard deviation of the noise is $\frac{\xi}{3}$. Then, we apply a Monte Carlo method. 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 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 hydrophones 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 uncer-

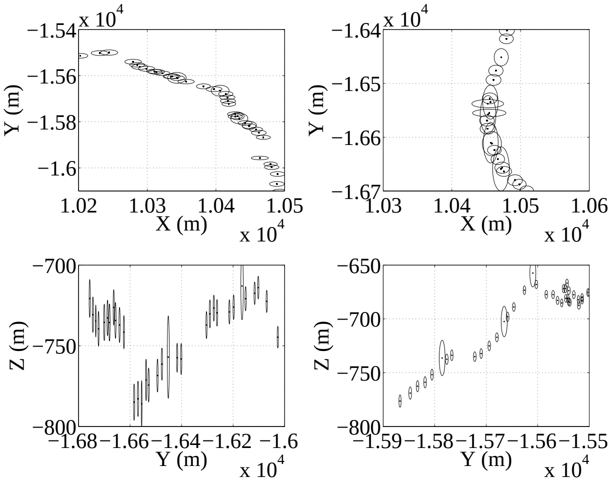

Figure 9: Confidence regions projection on X and Y and on Z and Y axes for D2 trajectory.

tainties 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 chunks in D1 is similar to ours. Our method provides thus robust online passive acoustics detecting/counting system of clicking MM groups in open ocean. Further studies will be conducted for click labeling and inter click analyses.

5 Acknowledgments

We thank AUTEC 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 Multimedia data for Cetology (PIMC)". We thank also H. Pettis for organizing the third international workshop on "Detection & Localization of MM Using Passive Acoustics" at Boston in 2007.

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 Adaptative methods in speech recognition, Nice*, Sept 2001.

[5] H. Glotin, D. Vergyri, C. Neti, G. Potamianos, and J. Luettin. 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 form 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] JF. 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 Intelligense, 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 bottom-mounted sensors. *Appl Acoustics*, 62:1074–1090, 2006.