Article

Performance Assessment of the Innovative Autonomous Tool CETOSCOPE© Used in the Detection and Localization of Moving Underwater Sound Sources

Yann Doh [1], Beverley Ecalle [1], Fabienne Delfour [1,2,3], Cyprien Pankowski [1], Gildas Cozanet [1], Guillaume Becouarn [1], Marion Ovize [1], Bertrand Denis [1] and Olivier Adam [1,4,5,*]

1   Association ABYSS, 1 Rue du Quai Berthier, 97420 Le Port, La Réunion, France
2   École Pratique des Hautes Études, Paris Science et Lettre, Centre de Recherches Insulaires et Observatoire de l'Environnement, CRIOBE EPHE-CNRS: UAR3278, Université de Perpignan, Bât. R, 58 Avenue Paul Alduy, 66860 Perpignan, France
3   Labex Corail, Centre de Recherches Insulaires et Observatoire de l'Environnement, CRIOBE, Papetoai, 98729 Moorea, French Polynesia
4   Institut Jean Le Rond d'Alembert, CNRS: UMR 7190, Sorbonne Université, 75005 Paris, France
5   Institut des Neurosciences Paris-Saclay, CNRS: UMR 9197, Université Paris-Saclay, 91400 Saclay, France
*   Correspondence: olivier.adam@sorbonne-universite.fr

Abstract: The detection and localization of acoustic sources remain technological challenges in bioacoustics, in particular, the tracking of moving underwater sound sources with a portable waterproof tool. For instance, this type of tool is important to describe the behavior of cetaceans within social groups. To contribute to this issue, an original innovative autonomous device, called a CETOSCOPE©, was designed by ABYSS NGO, including a 360° video camera and a passive acoustic array with 4 synchronized hydrophones. Firstly, different 3D structures were built and tested to select the best architecture to minimize the errors of the localizations. Secondly, a specific software was developed to analyze the recorded data and to link them to the acoustic underwater sources. The 3D localization of the sound sources is based on time difference of arrival processing. Following successful simulations on a computer, this device was tested in a pool to assess its efficiency. The final objective is to use this device routinely in underwater visual and acoustic observations of cetaceans.

Keywords: underwater passive acoustics; sound source detection; 3D localization; portable autonomous waterproof device

1. Introduction

The CETOSCOPE© is an original innovative autonomous waterproof device. It was designed and built to detect a sound event and to track an acoustic moving underwater source. The objective was to use this portable tool to visually and acoustically observe cetacean species in their marine environment and meet specific expectations such as identifying some individuals, analyzing their 3D movements or postures and describing their social interactions [1,2]. Focused on the passive acoustic part, this paper describes its architecture, the acoustic equipment and the different simulated and real tests conducted to measure its efficiency in water.

The goal of the CETOSCOPE© was to provide the visual and acoustic features in all directions and to build the 3D trajectories of the acoustic sources present around the device. To meet the expectations of users who are non-experts in electronics and/or teams with little financial funding, the chosen materials were easy to find and inexpensive.

Such a system relies on acoustic processing. Passive acoustics detection and localization of the acoustic sources are topics widely explored in many fields of science, such as robotics, computer vision, marine warfare and applied bioacoustics. Information on the

spatiotemporal locations of the acoustic sources can be derived from the acoustic propagation theory, combined with signal processing methods, and depends on the number of sensors and their relative positions from each other, and consideration of the acoustic characteristics of the marine environment and the sounds from these underwater sources.

Acoustic methods have already been developed for mounted and towed arrays of hydrophones. Using a pair of hydrophones can solve the bearing estimation. Inter-aural Level Difference (ILD) is inspired by human hearing by exploiting the variations in the acoustic paths between sensors [3], which cause variations in the magnitude, which can be linked to the angle of arrival. However, it provides only cues to the real positions, and the information must be combined with other parameters to reduce the uncertainty.

Based on the exploitation of the Time Difference Of Arrival (TDOA), static arrays are known to provide unambiguous, very precise locations of the acoustic sources [4–7]. The most-used TDOA resolution methods are based on finding geometric solutions for systems formed by hyperbolic intersections [8,9]. In one study [10], the authors reviewed four locators (in 2D space): Minimum Likelihood (ML), Weighted Sum (WS), Free Search (FS) and Hyperbolic Fixing (HF). In this review, ML returned the best results, and FS provided the lowest precision, for source locations.

In the development of the CETOSCOPE©, the approach was close to the HF method. Although this technique provided accurate results for remote sources, the detection of many nearby sources led to poor localization estimations. Furthermore, geometric resolution was not always possible, especially when intersections were located close to asymptotic areas or when the signal-to-noise ratio was degraded [11]. Complementary studies quickly concluded that HF was not efficient enough for short-distance localizations, and scientific research became oriented toward statistical approaches.

This paper firstly describes the architecture and the material of the CETOSCOPE©. Then, the method of sound source localization based on a 3D mesh is detailed. Thirdly, the article presents the results from numerical simulations in order to provide the 3D error map for source localization. Finally, the CETOSCOPE© was tested in a pool to assess its efficiency at localizing underwater sound sources whose exact positions were known. To conclude, the discussion deals with further perspectives and applications for dolphin behavior studies and the next technical steps to accomplish.

2. Materials and Methods

2.1. Material

Acoustic simulations were provided to select the best architectural design for the CETOSCOPE©, including the implementation of the 360° video camera, and to choose the optimal number of hydrophones and their relative positions. The final design was a tetrahedral shape with a total of 4 hydrophones, one at each corner of the tetrahedron, at a 3.32 m distance from each other. A frame of tubular stainless steel was chosen, to make the structure highly rigid and strong. Four 2 m long monoblock poles maximized the rigidity of the whole structure and ensured a fixed configuration when acoustic recordings, with no spatial variations and no vibrations of the hydrophone positions. Aquarian H2a hydrophones were fixed at the end of each pole, and fitted with anti-shock protection. According to the manufacturer, these are useful on a 10 Hz–100 kHz bandwidth, with a sensitivity of -180 dB re: 1 V/μPa $+/-$ 4 dB on 20 Hz–4 kHz.

To visually describe the 3D scene, a 360° video cam was required. Different tests were done with integrated video cameras in previous work [11], and finally, the 360° video system (rig) with 6 GoPro Hero 4s with 64 Go SDXC devices was selected because of the UHD resolution of each cam, and also because this system includes, in addition to horizontally arranged cameras, a top cam and a bottom cam that provide visual observations about the full 3D scene.

Fixed in the center of the device (Figure 1), a waterproof control box was also attached to the chassis, containing the ZOOM F8 digital recorder (192 kHz sampling frequency and 24-bit coding), a microcontroller to synchronize the audio recorders and the GoPro cameras,

and an auxiliary battery enabling them to be powered for about 1 h 30 min. The video and acoustic data were stored on a 1 Tb SD device.
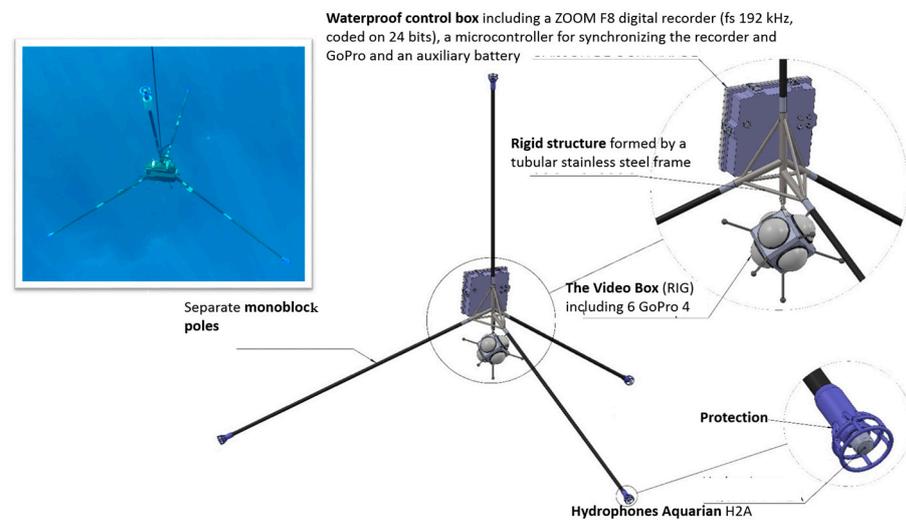

Figure 1. Description of the materials used for the CETOSCOPE©: a frame built with tubular stainless steel to make the overall structure rigid; 4 hydrophones, 1 placed at the extremity of each pole; the 360° video camera placed in the center; and the box including the digital recorder and the power battery.

The total weight in the air was 15 kg. The neutral buoyancy was set up at 5 m depth but can be manually adjusted depending on the use. The total hardware cost was less than USD 5000.

2.2. Detection

The detection algorithm was developed on various datasets, including simulated sounds and a real marine soundscape. Detections of the acoustic sources were verified by analyzing the videos when sound sources were close enough to the video cameras, and acoustic recordings provided the time-frequency features including the acoustic sources' intensities, types (transient or tonal), and bandwidths, and also metadata such as the starting time and the duration of their presence on the videos from each camera, named FC, BC, LC, RC, AC and BC, respectively, for "front cam", "back cam", "left cam", "right cam", "top cam" and "bottom cam". Because of the high resolution, it was possible to zoom in on the images to identify external features of the acoustic sources.

2.3. Localization

The 3D localizations of underwater acoustic sources were only possible on the videos because the distance was difficult to accurately estimate from a still image. Thus, the 3D localizations were only estimated from the analysis of the acoustic dataset.

For each detected acoustic event, the estimated time differences of arrival (TDOAs) were calculated from the cross-correlation applied on the 6 pairs of hydrophones. As presented in Figure 2, such a transformation highlighted the estimated TDOAs produced by all the sound sources. The time frame could be adjusted from 10 ms to 1 s. Short time sliding windows were adapted to impulsive/transient sounds; longer durations were adapted to continuous tonal sounds.
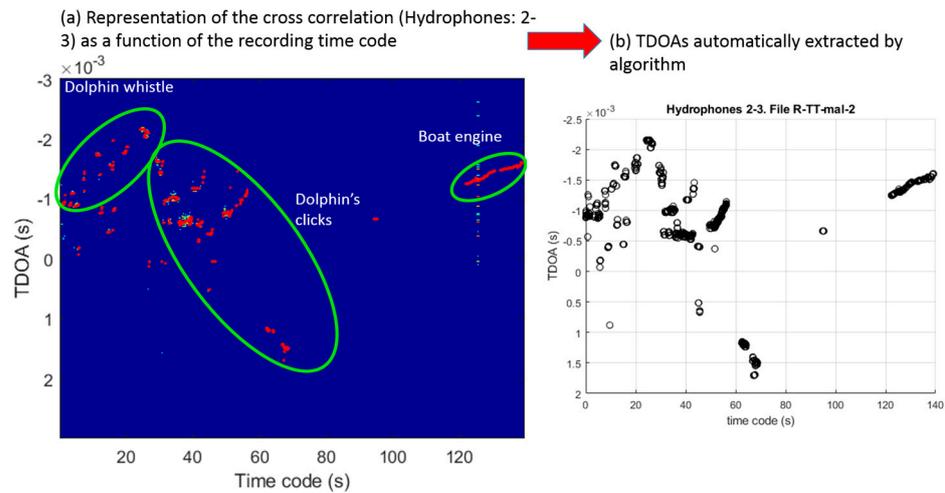
Figure 2. (**a**) Sliding cross-correlation of the signal over 130 s for a given pair of hydrophones. Time frame of 12.5 ms. Red dots represent the maximum of the cross-correlation by time frame and green ovals indicate the type of event confirmed by a human expert. (**b**) Automatically extracted TDOAs for the same sequence.

As an alternative to the theoretical method based on the intersection of 3 hyperbolas, the developed approach was inspired by mapping methods for Acoustic Source Location [12,13]. The proposed process compared the estimated Time Of Arrival (TOA) of a detected event with a grid of all the TOA provided by meshing the space around the CETOSCOPE©. It amounts to comparing the TDOAs provided by all these virtual positions from the 3D mesh to the estimated TDOAs (Figure 3). The dimensions of the mesh were chosen to optimize the space resolution for the estimated source positions and to take into account the calculation time. Obviously, the shorter the mesh length (distance between 2 nodes), the longer the computing time. This study was carried out taking into account that the calculations would be executed on a desktop computer or an embedded device.
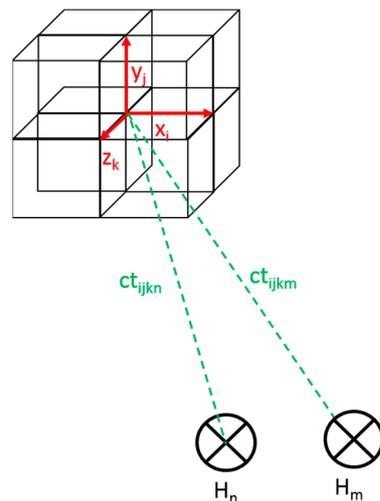

Figure 3. Geometric definition of the propagation distance between a given cubic mesh and a given pair of hydrophones n m.

In this study, for the acoustic simulation, the analyzed volume was chosen as an underwater cube of 40 m^3 with a 25 cm cubic mesh, and centered on the device. This volume can have a much larger size and a much smaller mesh length. The choices for fixing these 2 parameters (size of the volume and length of the mesh) depend on the objective of the study, based on the a priori information on the 3D positions of the sound sources and the required precision. Thus, in our study, the choices of the size of the prospecting volume

and the mesh length were based on visibility in clear sea water, estimated at around 20 m in normal conditions, and the size of the sound source was known to be greater than 0.5 m.

In addition, it is also necessary to take into account the computation time. The larger the volume and the smaller the cells, the longer the calculation time. In our study, the calculation related to more than 4 million possible positions for the sound source. Also, to save time, once the 2 parameters have been set, the TDOA matrix is calculated once and then stored in memory to be used in the following calculations.

The propagation time between a mesh and one hydrophone can be expressed as:

$$t_{ijkn} = \frac{\sqrt{\left(X_{H_n} - x_i\right)^2 + \left(Y_{H_n} - y_j\right)^2 + \left(Z_{H_n} - z_k\right)^2}}{c}, \tag{1}$$

where i, j and k are the indices of the mesh, and n is the index of the hydrophone and X_H, Y_H and Z_H its coordinates. c is the sound speed in the sea water.

Each point of the grid is associated with a combination of $\binom{N}{2}$ TDOAs, for an N-hydrophone array. The *TDOA* between 2 hydrophones n and m is:

$$TDOA_{ijknm} = t_{ijkm} - t_{ijkn}. \tag{2}$$

This results in T, a matrix $i \times j \times k \times \binom{N}{2}$.

The estimation of the position is done by comparing the estimated TDOAs $\hat{T}$ with TDOAs previously provided from the grid. The estimated position $(x, \hat{y}, z)$ is extracted by minimizing the quantity $E\{|T - \hat{T}|\}$, where E is the average [13]. A position *(x, y, z)* of the grid is therefore assigned to the acoustic event. The results can be represented in static form, displaying all the sources detected for a given sequence, or in a dynamic form (video), giving the successive estimated positions of the acoustic sources.

2.4. Precision and Accuracy of the 3D Localizations

To evaluate the performance of the CETOSCOPE©, its abilities to localize and precision were characterized. Given the small dimension of the array, and that the shape of the array was precisely known, measurable and undeformable, geometrical calibration was not required, as it is described in [14] for a wide array. The assessment was conducted following two experiments: the first one was numerical, based on a Monte Carlo approach. Random source positions were simulated and then estimated by the previously described method. This operation was reproduced on the whole mesh (# realizations = 10,000) and the error ε was calculated, defined as $\varepsilon = \|\hat{P} - P\|_2$, the Euclidian distance between the estimated vector $\hat{P}$ and the ground truth position P. For a scalar quantity, this expression becomes the absolute value of the difference. The CETOSCOPE© is virtually placed under the sea surface at $(x = 0, y = 0, z = -4.5m)$.

To compare the numerical simulation with real measurements, a second assessment was conducted in a closed controlled pool, where synthetic signals and the positions of the acoustic sources were exactly known. This test took place at the experimental pool of the French Institute for Ocean Science IFREMER (La Seyne-sur-Mer France, Figure 4a. Its length, width and depth were, respectively, $15 \times 10 \times 6$ m. Scientists and professional divers were in charge to conduct the protocol described in Figure 4b. The CETOSCOPE© was placed at 2.5 m depth, in the corner of the pool, suspended with a custom-made system, allowing a controlled rotation of the device. Speakers were placed in the diagonal of the pool at 5 m, 9 m and 13 m at 1.5 m and 3.5 m depth. Then, 45° successive rotations were carried out in order to rebuild the experiment detailed in Figure 4c. Six points at each angle were recorded for 8 angles (total number 48).
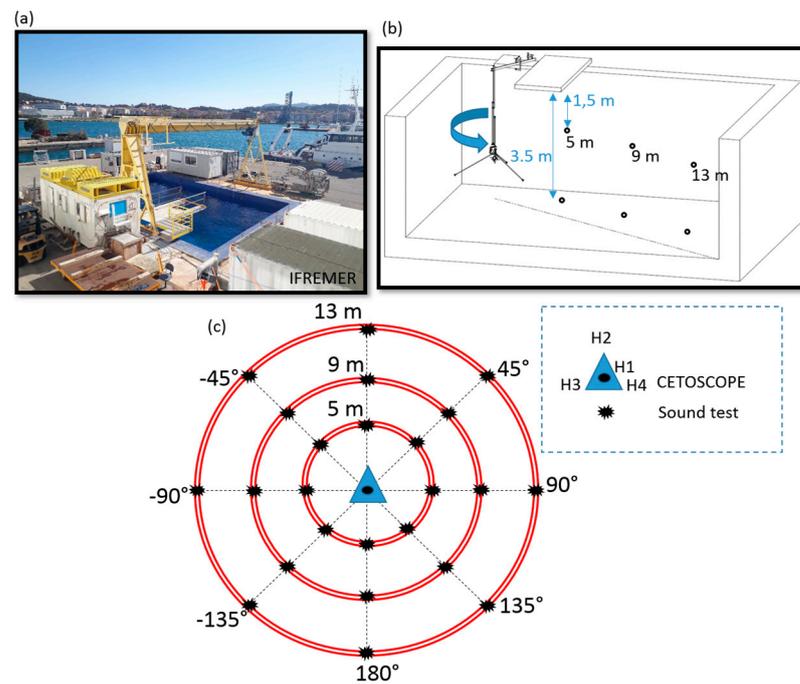
Figure 4. (**a**) view of the IFREMER (La Seyne/Mer) experimental pool, (**b**) side representation of the pool experiment protocol, (**c**) top view of the successive speaker positions.

Synthetic signals formed by 10 impulses at 1 s intervals were emitted at the precise locations previously mentioned. The durations and the magnitudes of the impulses were adjusted regarding the sound source characteristics: the underwater Lubell 916 speaker with a 200 Hz–23 kHz bandwidth (http://www.lubell.com/LL916.html, accessed on 6 April 2023) and the sensitivity of the hydrophones in the experimental conditions. TDOAs were extracted on received signals by human experimenters. To do that, they measured the number of signal samples on the 4 synchronized channels, and then compared their estimation with a cross-correlation. The extractions were done manually to ensure the best level of precision, and to avoid any confusion caused by the sound reflecting off the walls of the pool.

Then, the 3D positions were estimated following the algorithm described in this paper. Statistics were calculated on raw distributions and after filtering outliers when the modified Z-Score > 3.5 [15].

3. Results

3.1. Characterization of the Localization Error from Numerical Acoustic Simulation

Preliminary tests showed that the algorithm retrieved the true position with ($\varepsilon = 0$) in 100% of the cases, if the simulated position was right on the mesh nodes. Therefore, virtual true positions were all simulated out of the nodes. Figure 5 shows a decomposition of the error as a function of the Cartesian coordinates (x, y, z) and spherical projection azimuth, elevation and radius, respectively (a, e, r). Point clouds for Cartesian representation showed a relative isotropic behavior of the estimates. Because the acoustic source can only be located under the sea surface, the (z) plan is a semi-space compared to the (xy) plan. The mean error over (x, y, z) of the estimates varied between 0.8 m (std = 1.2 m) and 0.9 m (std = 1.3 m). It tended to increase with the distance to the center of the device. The distribution of the mean error for the whole coordinate system decreased regularly, with 50% of the estimates generating an error under 0.8 m. (std = 1.22 m).
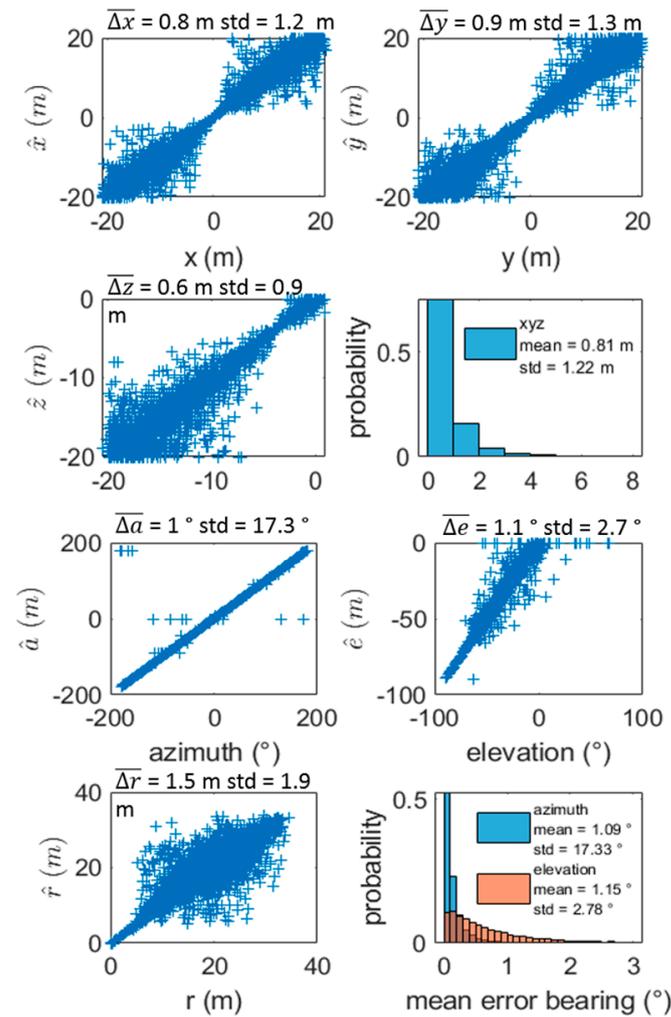
Figure 5. Point cloud estimates and true simulated positions. Up, Cartesian coordinates with the associated mean error distribution, down, spherical coordinates. Histograms are truncated for events as $p < 0.001$.

The azimuth point cloud was concentrated around the line $(x = y)$ with a mean error of $1°$ (with std $= 17.3°$) and did not show any dependency with the bearing. A few erratic values resulting from confusions in the sign highly impacted the std, representing less than 0.1% of the cases. These values were removed from the distribution representation for graphical representation reasons. Elevation showed a spreading profile of the error when the elevation was weak, i.e., when the source was near the surface. Mean error was evaluated at $1.1°$ (std $= 2.7°$). The distribution of the bearing error regularly decreased for the highest value and was very narrow for azimuthal estimates, as 97% generated an error under $1°$. Radius estimates showed a similar dependency on the distance, as revealed with Cartesian representations.

The next operation aimed to visualize a synthetic representation of the error in 3D. The field of the errors was computed using triangulation-based natural neighbor interpolation. Figure 6 provides constant (z) views for four different depths: 18 m, 10 m, 4.5 m (virtual depth of the device) and 1 m. The error varied within the 3D space, highlighting "shadow" areas (colored peaks on the graph). Source position estimates from those locations were degraded due to geometrical properties of the array. The highest values were scattered up to 8 m. For (z) views close to the device depth, peaks seemed to be aligned along 3 symmetrical plans defined by the base of the tetrahedral structure. A low-error area was

located at the center. This "crater" got larger as (z) views moved away from the depth of the device. Side effects were clearly visible for cuts next to the bottom or the surface.
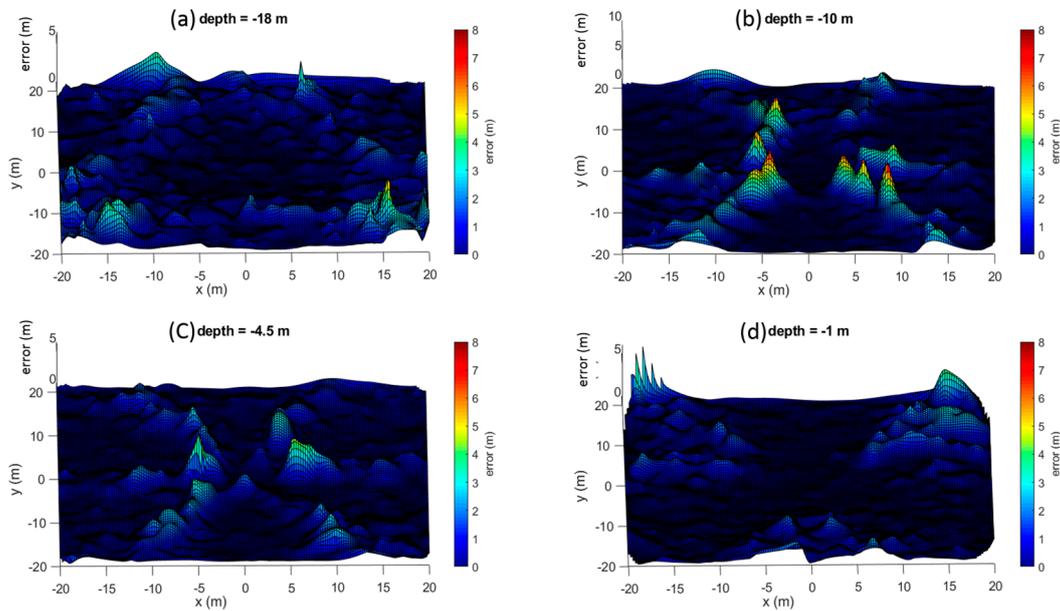

Figure 6. 3D representation of the estimated error for different (z) cuts as function of Cartesian coordinates (x,y): (**a**) z = −18 m, (**b**) z = −10 m, (**c**) z = −4.5 m, (**d**) z = −1 m. CETOSCOPE© virtually placed at 4.5 depth.

An ortho shape was studied with the same methodology. Although the tetrahedral shape provided a higher maximum error than the ortho shape, this profile had the advantage of being symmetrical, with the highest values focused close to the array (less than 5 m error over 5 m from the device).

3.2. Characterization of the Localization Errors from the Pool Experiment

The results and representation were rigorously prepared by human experts. However, higher fluctuations were expected for the pool experiment due to the difference in the number of realizations: 48 versus 10,000 for the simulation. At a global level, Figure 7 presents the histogram of the mean error between the retrieved positions and the true positions. For Cartesian coordinates, in Figure 7a, with a resolution of 0.83 m, the distribution spread out to an error of 21 m. Note that retrieved positions resulting in an error greater than 10 m represent 8/48 cases. Random events in the experimental setup or processing may have deteriorated these results. The mean error value was 2.7 m (std = 2.1 m) without considering the outliers (4.5 m, std = 4.8 m for the raw distribution). A total of 16/48 positions were retrieved with an error value equal to or smaller than 1 m. In Figure 7 b, the distribution of the error over the retrieved angles of arrival azimuth and elevation is given with a resolution of 1°. The mean value is 2.3° (std = 1.8°) for the azimuthal error and 2.5° (std = 1.5°) for the elevation (respectively, 2.3° (std = 1.9°) and 2.8° (std = 1.5°), without outliers' correction). A total of 38/48 retrieved positions presented an azimuthal error smaller than 3°, and 29/48 retrieved positions presented an elevation error smaller than 3°.
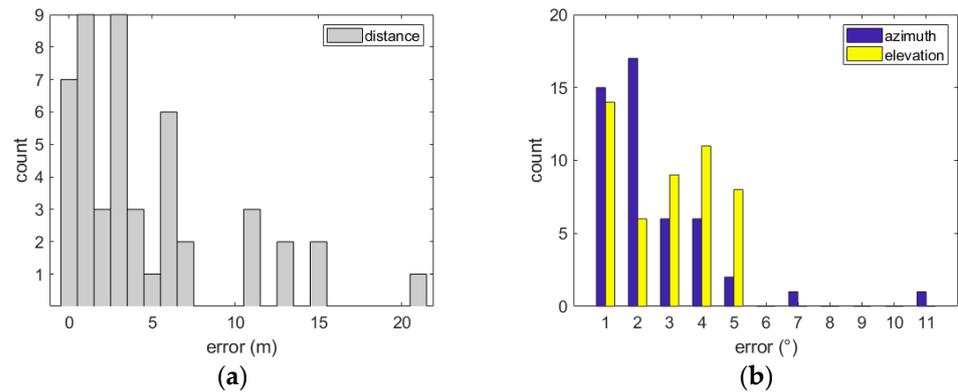
Figure 7. Distribution of the mean error: (**a**) on the Cartesian coordinates of the retrieved positions, (**b**) on the retrieved angles (azimuth and elevation).

Figure 8 shows a polar representation of the error in the different azimuths for all the measured Z plans, for the pool experiment and for the numerical simulation. These polar plots highlight the directionality of the error. Except for the z = 1.50 m (Figure 8e), results show a relative accordance between the experiment and the simulation by pointing out the bearing of the highest error. Given the considered plan, its direction changes between 225° and 270°.
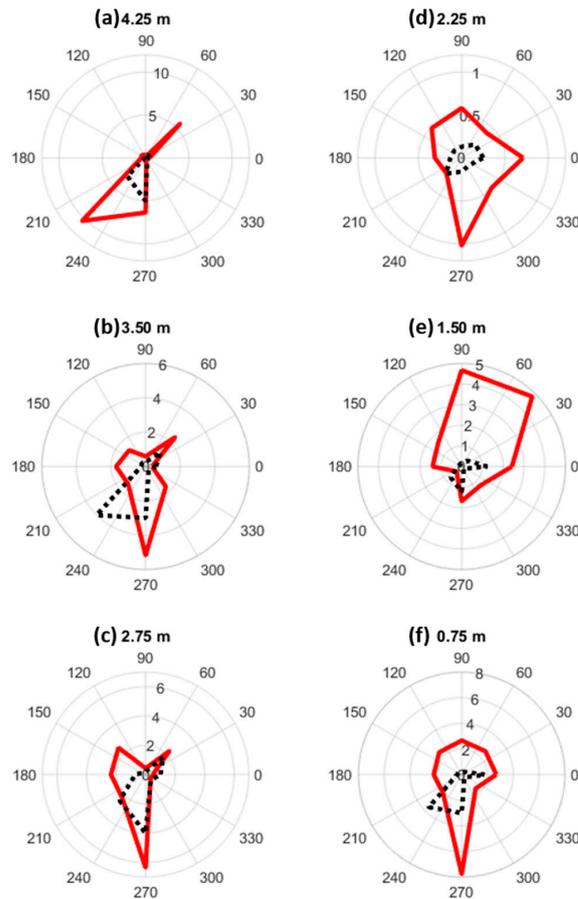
Figure 8. Polar representation of the error for different depths: (**a**) z = −4.25 m, (**b**) z = −3.50 m, (**c**) z = −2.75 m, (**d**) z = −2.25 m, (**e**) z = −1.50 m, (**f**) z = −0.75 m. Red line for pool experiment, black dots for the simulation.

4. Discussion

Error studies based on random position simulation and the pool experiment has revealed the deterioration of the estimate accuracy due to intrinsic factors. Numerical issues appeared due to the used algorithm, and geometrical constraints were defined by the shape of the array. This study did not quantify the impacting factors, neither the impact of the detection nor the signal-to-noise ratio (SNR); it aimed to provide the limits of this array configuration. The simulation and experiment were in accordance and validated that the error introduced by the array's geometry was highly spatially dependent.

For the numerical simulation, the magnitude of the error inn the retrieved position in Cartesian coordinates showed some fluctuations around 1 m with some local miscellaneous error peaks (until 8 m). Those values were similar to those found with small arrays. For three clusters of four hydrophones with 15 cm minimum spacing, the error was between 1 and 2 m in [16]. Using an array of 6 hydrophones with 1 m spacing, another study reported a range error between 0.1 m and 2 m [17]. The error tended to increase with the distance to the array, as reported by other papers [10,18] as well as for the terrestrial environment using [1].

3D cartography of the resulting error showed that the ability to estimate the position was closely linked to the relative position of the source regarding the depth of the device. Some of the high-error areas were located along axes containing a pair of hydrophones. The ambiguity appeared as the algorithm could converge for different possible solutions. A projection of the tetrahedron on the plan (xy) showed that error accumulated close to the projection, see Figure 9. Shadow areas were numerous and close to the device for plan located at the depth of device. However, some high error areas remained unexplained by this principle.
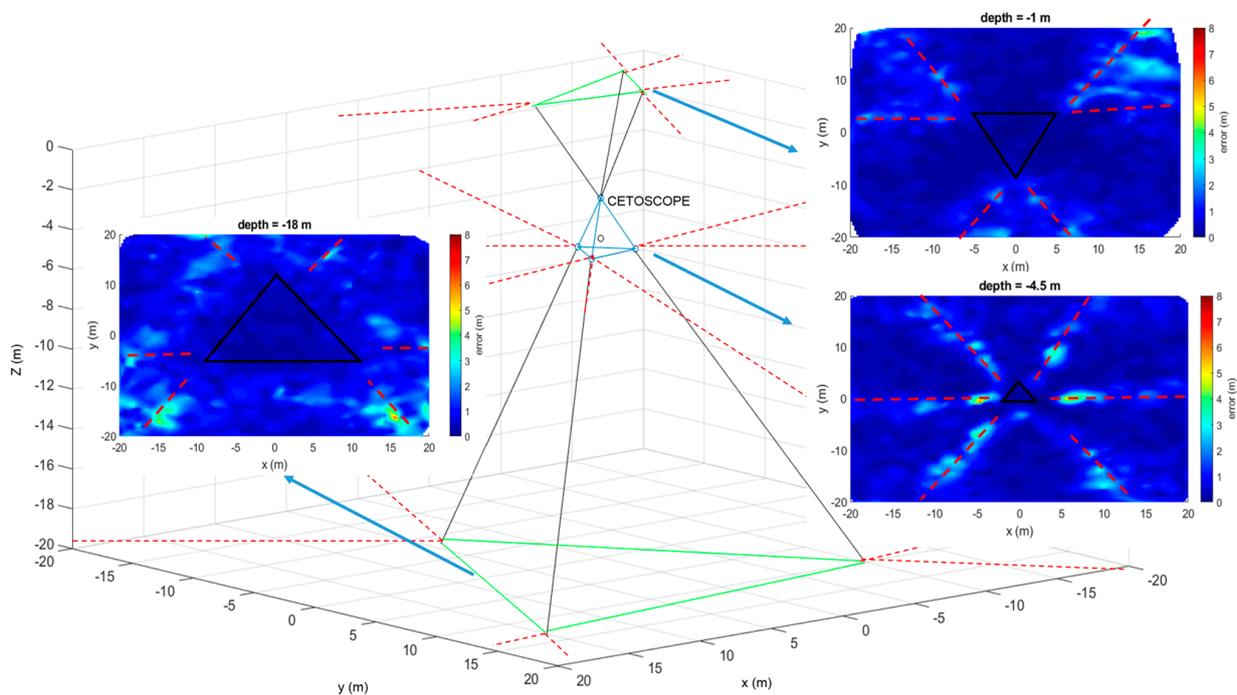

Figure 9. Relationship between geometrical projections of the array and changes in the shape of the shadow areas.

The results of both experiments are summarized in Table 1. The pool experiment showed an important deterioration of the error. This was expected due to statistical issues and also due to random and unpredictable events happening during the experiment. Despite this, both studies revealed the robustness of the bearing estimates. Though the distance between the device and the sound source was sometimes misestimated, the angle

of arrival was strongly reliable. Of all the positions, 33% to 75% were retrieved with an error of less than 1 m, and, in 40% to 99% of the cases, the CETOSCOPE© pointed in the true direction with an error of less than 3°.

Table 1. Summary of the results for the numerical simulation and for the pool experiment.

Simulation	Mean Error/Std	Error <1 m	Error <3°
Numerical simulation	Distance 0.8 m/1.2 m Azimuth 1.0°/17.3° Elevation 1.1°/2.7°	75%	Azimuth 99% Elevation 99%
Pool experiment	Distance * 2.7 m/2.1 m Azimuth * 2.3°/1.8° Elevation * 2.5°/1.5°	33%	Azimuth 79% Elevation 40%

* with outliers' correction.

The accuracy for the estimated 3D positions depends on the distance between the underwater acoustic sources and the CETOSCOPE© and also on the geometry of the CETOSCOPE© structure. The error map will be used to define a confidence criterion. For moving sources, these estimations could also be improved because they will cross areas with low accuracy. Also, post-processing could be applied to extract the trajectory from the successive estimated positions, and correct the error in the estimated positions by interpolation.

At the same time, the fusion of acoustic locations with the 360° video scene began. This step aimed to bring the geometry of the acoustic model in line with the field of cameras. It had to consider the optical deformations of each camera. A repositioning calculation was then necessary on the edges of the images. A calibration phase was necessary to characterize the conversion between both spaces. Although the accuracy of the distance can be uncertain, the bearing estimates are reliable on average, with an accuracy within 1–2.3°. Thus, the depth of field is impacted but the bearing separation is very accurate.

5. Conclusions

This work is the first characterization of the passive acoustic array component of the CETOSCOPE© audio video device used to detect moving acoustic sources, and its 3D position localization algorithm. This objective was reached by conducting sound source simulation experiments in a controlled pool and a Monte Carlo numerical simulation. At a first global level of analysis, in a volume of 40 m × 40 m × 40 m around the device, the mean localizing error was estimated for the simulation at 0.8 m (std = 1.2 m), and for the pool experiment at 2.7 m (std = 2.1 m). Analysis of the precision as a function of spherical coordinates showed that the mean error fell at about 1° for azimuth and elevation for the numerical simulation, and within 2.3–2.5° for the pool experiment. Both results were in relative accordance and highlighted that the localizing error was highly space-dependent. Numerical simulation allowed the computation of 3D error maps and defined with precision acoustical shadow areas that were confirmed by the pool experiment.

In conclusion, the rigid structure of the CETOSCOPE©, on which the 360° camera and the 4 microphones were fixed, was designed to optimize the estimations of the positions of sound sources. Dedicated software was created to provide the location and to build the 3D trajectories in THE case of moving emitters. The final goal will be to use the CETOSCOPE© to track cetaceans in their marine environment, for ethological purposes. Thus, the CETOSCOPE© will be an efficient tool to describe marine ecosystems for projects contributing to ocean conservation.

References

1. Mcgregor, P.; Dabelsteen, T.; Clark, C.; Bower, J.L.; Holland, J. Accuracy of a passive acoustic location system: Empirical studies in terrestrial habitats. *Ethol. Ecol. Evol.* **2010**, *1997*, 269–286. [CrossRef]
2. Fernandez, M.; Vignal, C.; Soula, H. Impact of group size and social composition on group vocal activity and acoustic network in a social songbird. *Anim. Behav.* **2017**, *127*, 163–178. [CrossRef]
3. Middlebrooks, J.C.; Green, D.M. Sound localization by human listeners. *Annu. Rev. Psychol.* **1991**, *42*, 135–159. [CrossRef] [PubMed]
4. Clark, C.W.; Ellison, W.T. Numbers and distributions of bowhead whales, Balaena mysticetus, based on the 1986 acoustic study off Pt. Barrow, Alaska. *Rep. Int. Whal. Comm.* **1989**, *39*, 297–303.
5. Spiesberger, J.L.; Wahlberg, M. Probability density functions for hyperbolic and isodiachronic locations. *J. Acoust. Soc. Am.* **2002**, *112*, 3046–3052. [CrossRef] [PubMed]
6. Alameda-Pineda, X.; Horaud, R. A Geometric Approach to Sound Source Localization from Time-Delay Estimates. *IEEE Trans. Audio Speech Lang. Process.* **2014**, *22*, 1082–1095. [CrossRef]
7. Baggenstoss, P.M. An algorithm for the localization of multiple interfering sperm whales using multi-sensor time difference of arrival. *J. Acoust. Soc. Am.* **2011**, *130*, 102–112. [CrossRef] [PubMed]
8. Spiesberger, J.L. Hyperbolic location errors due to insufficient numbers of receivers. *J. Acoust. Soc. Am.* **2001**, *109*, 3076–3079. [CrossRef] [PubMed]
9. Morrissey, R.; Ward, J.; DiMarzio, N.; Jarvis, S.; Moretti, D. Passive acoustic detection and localization of sperm whales (*Physeter macrocephalus*) in the tongue of the ocean. *Appl. Acoust.* **2006**, *67*, 1091–1105. [CrossRef]
10. Urazghildiiev, I.; Clark, C.W. Comparative analysis of localization algorithms with application to passive acoustic monitoring. *J. Acoust. Soc. Am.* **2013**, *134*, 4418–4426. [CrossRef] [PubMed]
11. Lopez-Marulanda, J.; Adam, O.; Blanchard, T.; Vallée, M.; Cazau, D.; Delfour, F. First results of an underwater 360° HD audio-video device for etho-acoustical studies on bottlenose dolphins (*Tursiops truncatus*). *Aquat. Mamm.* **2017**, *43*, 162–176. [CrossRef]
12. Baxter, M.G.; Pullin, R.; Holford, K.M.; Evans, S.L.; Delta, T. Source location for acoustic emission. *Mech. Syst. Signal Process.* **2007**, *21*, 1512–1520. [CrossRef]
13. Al-Jumaili, S.K.; Pearson, M.R.; Holford, K.M.; Eaton, M.J.; Pullin, R. Acoustic emission source location in complex structures using full automatic delta T mapping technique. *Mech. Syst. Signal Process* **2016**, *72–73*, 513–524. [CrossRef]
14. Vanwynsberghe, C.; Challande, P.; Marchal, J.; Marchiano, R.; Ollivier, F. A robust and passive method for geometric calibration of large arrays. *J. Acoust. Soc. Am.* **2016**, *139*, 1252–1263. [CrossRef] [PubMed]
15. Iglewicz, B.; Hoaglin, D. *How to Detect and Handle Outliers*; ASQC Quality Press: Milwaukee, WI, USA, 1993; pp. 11–13.
16. Gillespie, D.; Palmer, L.; Macaulay, J.; Sparling, C.; Hastie, G. Passive acoustic methods for tracking the 3D movements of small cetaceans around marine structures. *PLoS ONE* **2020**, *15*, e0229058. [CrossRef] [PubMed]

17.   Zhao, D.; Huang, Z.; Su, S.; Li, T. Matched-field Source Localization with a Mobile Short Horizonta Linear Array in Offshore Shallow Water. *Arch. Acoust.* **2013**, *38*, 105–113. [CrossRef]
18.   Macaulay, J.; Gordon, J.; Gillespie, D.; Malinka, C.; Northridge, S. Passive acoustic methods for fine-scale tracking of harbour porpoises in tidal rapids. *J. Acoust. Soc. Am.* **2017**, *141*, 1120. [CrossRef] [PubMed]