

Research Article

Automatic Indexing for Content Analysis of Whale Recordings and XML Representation

Frédéric Bénard and Hervé Glotin

Laboratoire des Sciences de l'information et des Systèmes (LSIS-UMR CNRS 6168), Université du Sud Toulon Var, BP 20132, 83957 La Garde Cedex, France

Correspondence should be addressed to Frédéric Bénard, frederic.benard83@gmail.com

Received 1 July 2009; Revised 24 November 2009; Accepted 26 January 2010

Academic Editor: Dirk Maiwald

Copyright © 2010 F. Bénard and H. Glotin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper focuses on the robust indexing of sperm whale hydrophone recordings based on a set of features extracted from a real-time passive underwater acoustic tracking algorithm for multiple whales using four hydrophones. Acoustic localization permits the study of whale behavior in deep water without interfering with the environment. Given the position coordinates, we are able to generate different features such as the speed, energy of the clicks, Inter-Click-Interval (ICI), and so on. These features allow to construct different markers which allow us to index and structure the audio files. Thus, the behavior study is facilitated by choosing and accessing the corresponding index in the audio file. The complete indexing algorithm is processed on real data from the NUWC (Naval Undersea Warfare Center of the US Navy) and the AUTECH (Atlantic Undersea Test & Evaluation Center-Bahamas). Our model is validated by similar results from the US Navy (NUWC) and SOEST (School of Ocean and Earth Science and Technology) Hawaii university labs in a single whale case. Finally, as an illustration, we index a single whale sound file using the extracted whale's features provided by the tracking, and we present an example of an XML script structuring it.

1. Introduction

Processing of marine mammal signals for passive acoustic localization is a problem that has recently attracted attention in scientific literature. Motivation for processing marine mammals signals stems from increasing interest in the behavior of endangered marine mammals. In [1–3], authors recently developed a robust algorithm for tracking multiple whales in real time. Based on this expertise, we intend here to index the whale recordings for a later detailed dynamic analysis of the whale behavior, like its hunting/foraging strategies. Our experiment is conducted on a real deep ocean recording of 25 minutes from the AUTECH Bahamas center using their open-ocean widely spaced bottom-mounted hydrophone array. This recording contains series of click sequences of sperm whales (*Physeter catodon*). The experiments in this paper consist of tracking the whale and extracting features from the audio files. Here, we prefer to use the SMF (Stochastic Matched Filter) detector [1] since Teager-Kaiser filtering is not efficient enough for accurate ICI (Inter-Click-Interval)

computation [2, 4–6]. After the positions estimation, ICIs (Figure 2) are automatically processed, as well as speed, distance to a selected hydrophone, trajectory angles, and click energy. These features constitute different values classified in blocks. The output of the method is an indexing of the hydrophone recording with the corresponding blocks.

2. Material and Method

Data are recordings from the ocean floor near Andros Island, Bahamas, provided with sound speed profiles and recorded in March 2002. Dataset 2 is sampled at 48 kHz and contains marine mammal clicks and whistles and background noise (e.g., distant engine boat noises). Dataset 2 is recorded on hydrophones 7 to 11 with 25-minute length (Table 1).

Source tracking is performed with continuous localization and a constant sound speed profile in 3D space using Time Delays Of Arrival (TDOA) estimates from four hydrophones. Similar results are described in [1]. In dataset 2, one sperm whale is localized.

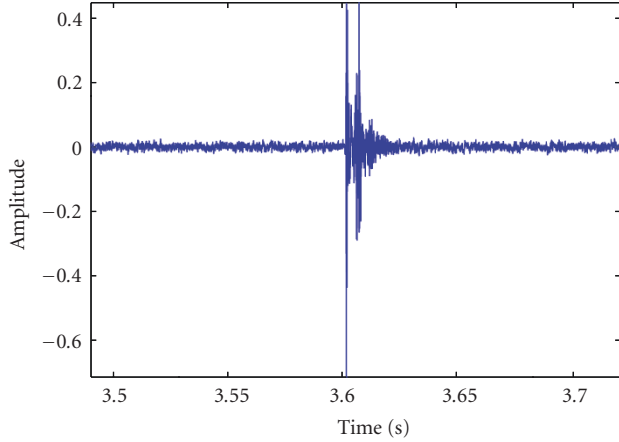


FIGURE 1: Example for a sperm whale click.

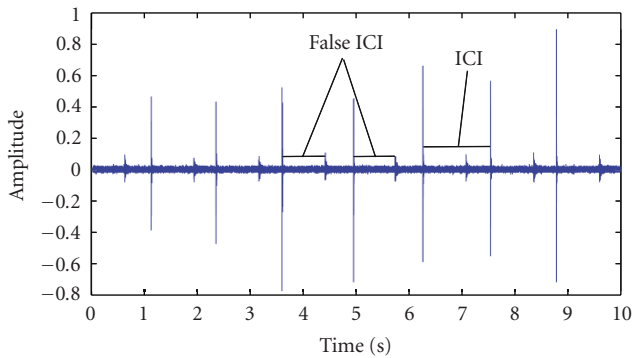


FIGURE 2: A click train with echoes from a single sperm whale. An ICI and a false ICI are shown.

TABLE 1: Hydrophones positions in dataset 2: Dist = Distance to barycenter (m).

Hydros	Dist	X (m)	Y (m)	Z (m)
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

2.1. Signal Filtering and Source Localization. A sperm whale click is a transient increase of signal energy lasting about 20 ms (Figure 1 shows the relative amplitude of a click). The clicks are generally emitted every 1 s forming groups called trains (Figure 2). Each click is usually followed by an echo, maybe due to the reflection of the click train at the ocean surface or bottom or different water layers. The methods used for filtering and tracking are described in [1]. In order to generate robust estimates of the TDOA, correlation-based techniques are generally avoided. In reality, the detector often detects the corresponding echo after each click (Figure 2). Therefore we use the SMF detector [1] with

a fast and efficient echo removal method. The output of the detector is the time of each click and some echoes. The echoes have a relative amplitude and frequency likeness with a click (especially at low SNR). The echo removal method is based on the temporal elongation of the echo, which is much longer than the click length because of the frequency energy dispersion in the different layers of the sea and smearing of the signal due to reflection. The echo removal is processed for each echo detected as a click by the SMF. The SMF is a filtering method employed here for detection. We chose the SMF because simple matched filter only takes into account the stochastic property of the noise, while in the SMF theory the signal of interest is also stochastic. The motivation of this hypothesis is that the sperm whale click is not a deterministic signal since the frequency and amplitude properties change according to the whale position relative to the hydrophone. Moreover, the whale modulates the signal frequency and amplitude for echolocation or communication. The experiment on a thousand clicks shows that a click received on a hydrophone can be considered gaussian. Therefore, we consider the click received on the hydrophone as a stochastic process. The clicks and sea noise are considered as zero-mean gaussian stochastic processes. For the SMF, we consider a stochastic process of interest, s , of length N , with a covariance matrix noted as $E(ss^T) = A$. The noise b is additive, centered, and independent. The covariance matrix $E(bb^T)$ is noted as B . The SMF theory says that the filter impulse response h for an FIR filter (length N) that maximizes the SNR (Signal-to-Noise Ratio) is the eigenvector solution of

$$Ah = \lambda Bh, \quad (1)$$

associated to the greatest eigen value λ_0 . The function used for the detection is

$$\Lambda(x) = [h^T x]^2, \quad (2)$$

with x as an N -length window. We utilize 20 ms windows which correspond to the mean click length. As we can see in (1), h is function of the sea noise process, so there are different filtering vectors for each channel (the hydrophones are separated by several kilometers and the level and properties of the noise are different for each channel). After h calculations for each channel, we filter the signal with one bin of shifting between each window. Finally, thanks to the measured TDOA and an acoustic model based on a constant sound speed profile, the least squares cost function determines the marine mammal positions using a multiple nonlinear regression with Gauss-Newton method (Levenberg-Marquardt) [4, 8, 9]. The residuals are approximated using a Chi-square distribution with $Nc - d$ degrees of freedom, noted as X_{Nc-d}^2 , where Nc is the number of hydrophone couples considered and $d = 3$ the number of unknowns (x, y, z) coordinates. The position is accepted if the residual is less than a threshold α , calculated by solving $P = \text{prob}(X_{Nc-d}^2 > \alpha)$ with $P = .01$ (we keep 99% of the estimates).

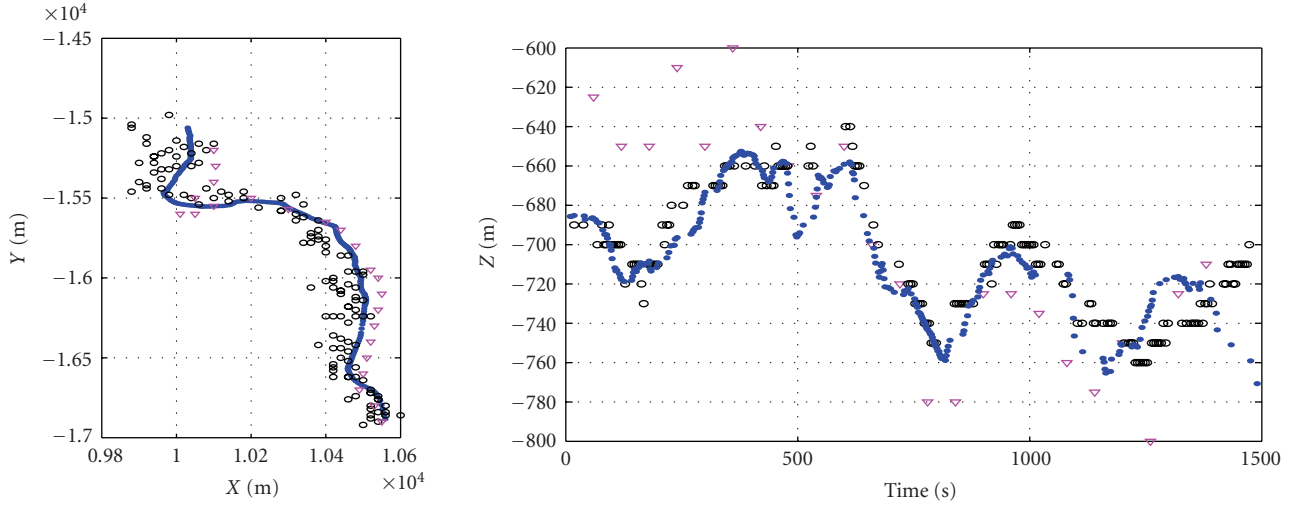


FIGURE 3: Plan view and diving profile of the marine mammal in dataset 2, estimates with a constant profile, and the SMF method (\cdot); and estimates from Morrissey's (∇) and from Nosal's methods (\circ).

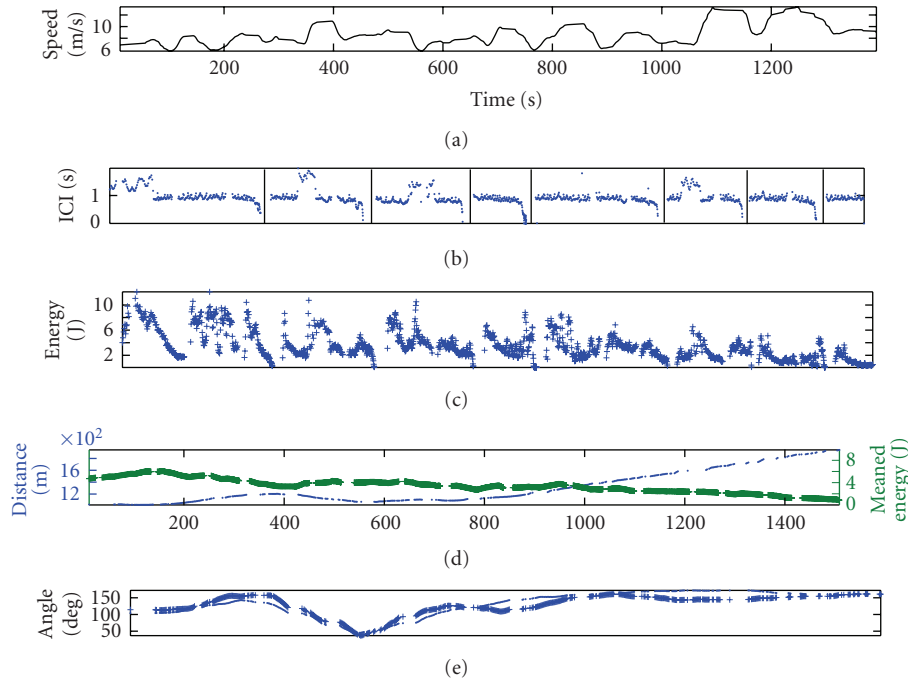


FIGURE 4: All features versus time of recording in dataset 2. (a) Speed of the whale. (b) ICI. Lines in (b) underline ingestions of the preys. (c) Energy of the clicks. In (d), distance (\cdot) and meaned energy ($+$) are plotted, so we can see the influence of the distance to the energy bias, and the energy decreases while the distance increases. (e) Theta1 ($+$), Theta2 (\cdot). Theta angles are smoothed on 10 s windows, so we cannot notice yaw and pitch variation, visible at 1 s scale.

2.2. Review of the Results and Comparisons of the Tracking. For dataset 2 (Figure 3), the results are comparable to Morrissey's [10] and Nosal's [11] methods. Moreover, with our constant sound speed model, the results are about the same as Morrissey's and Nosal's, who used profiles corresponding to the period and place of the recordings. The confidence regions are computed for the dataset with a Monte Carlo method. The ellipse maxima (30 m) are similar to the marine mammal length ($\hat{2}$ 0 m).

3. Feature Extraction

3.1. Labeling. The SMF method is not accurate enough for a direct ICI indexing (there could be spurious echoes after the echo removal process) [12]. Therefore, we use the labeling algorithm described below. The ICIs are detected on a selected hydrophone. Here, we chose H7, the one with the best SNR. Knowing the TDOA of a corresponding whale position at a given time, signals on H7 and H8 are shifted

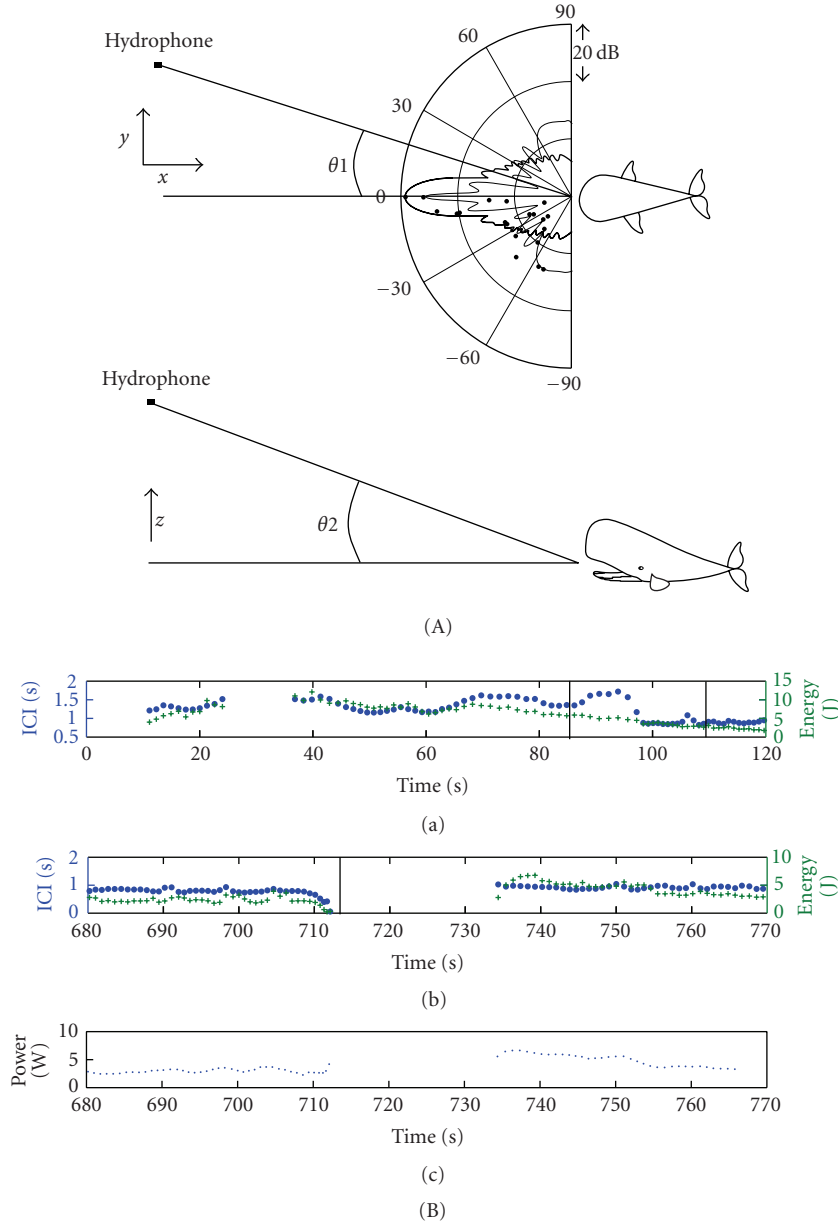


FIGURE 5: (A) Theta angles of a sperm whale. (B) Time zooms on ICI Figure 4(b) (\cdot) and energy (+) on a foraging (a) and a prey attack (b) behavior (silence after the ICI inferior or equal to 0.45 s is due to ingestion). The variation between ICI_{\min} , ICI_{\max} emphasizes a foraging behavior. In (a), during the first and the third phases between vertical lines, we can see a high correlation between energy (E) and ICI, respectively, 0.82 and 0.75, which means that pitch and yaw do not affect the ICI and the energy with a possible power estimation, whereas in the second phase, correlation is 0.47, and the whale is pitching and/or yawing. In (b) the first phase also denotes a good correlation (0.77), and in the second the correlation decreases and also emphasizes a behavior change. (c) is the train click power versus time averaged on 3 s.

from the TDOA values. We then process the SMF detector and the echo removal algorithm on H7 and H8. Both signals are summed in order to display the high-energy clicks and to increase the SNR. The summed signal is thresholded, to keep the remaining true time of the clicks. ICI computation is automatic from the click times.

3.2. Indexing. The features extracted through the localization are the speed, ICI, energy of each click, distance, and

head's angles to a given hydrophone. We process the features on dataset 2, which corresponds to the whale at depth. Some of these features are directly computed using the positions of the whale (speed, angles, and distance) and others with the labeling of the records (ICI and energy). The quality of the feature estimation is directly dependent on the labeling. Compared to a manual labeling, the match is perfect for the usual clicks, except for creaks which are very low-energy clicks, and are difficult to label. But

the influence of the creaks is weak on the ICI and the energy results. The feature estimation is also dependent on the position estimation (the maxima of the confidence regions are 30 m, but we do not need higher accuracy for a correct features extraction). Each feature can be computed independently.

- (i) Speed is the speed norm in the space and is computed similarly to the distance, during the time of recording directly from and to the positions estimation (Figures 4(a) and 4(d)).
- (ii) ICI (Figure 4(b)) corresponds to the interval between two clicks and is typically about 1 s for foraging behavior, and decreases ($ICI \leq 0.45$ s; clicks are called creaks) during a prey (generally squid) detection [13]. A short silent concludes the sequence (the whale does possibly not emit click during prey ingestion). The ICI and the click energy (Figures 4(b) and 4(c)) are directly computed from the labelling described in Section 3.1.
- (iii) Power estimation is computed with the following formula: $energy \times 1/ICI$. Power is averaged on 3 s windows.
- (iv) The direction of the whale is defined by two angles, Theta1 and Theta2 (Figure 4(e)). Theta1 is the off-axis angle of the whale and is defined as the angle separating the whale dorsorostral axis to the line joining the whale to the hydrophone on the (x, y) plane. Theta2 is calculated similarly after a projection onto the z -axis. Both angles underline the features of the whale's emission cones. For Theta1, beam patterns [14] are plotted. Indeed, when angles increase toward 180 degree, the whale faces away from the hydrophones, and the energy decreases.

Figure 6 depicts the diagram to obtain the different computed features and the connections with the tracking algorithm. Measurements are correlated with each other. The energy is mainly function of the ICI, with the whale emitting high-energy clicks during a foraging period, whereas during a prey track energy is lower (Figure 5(B)). Moreover, the energy is also a function of the distance (due to transmission loss), over the whole energy plot. Nevertheless, energy and ICI variations do not correlate well and depend on the hydrophone position. Several models exist to describe the correlation, but in [14] the preferred model stipulates that the whale echolocates at slow-rate ICI_{max} on one side and faster ICI_{min} on the other.

4. Content Description

4.1. Behavior Analysis. When searching for prey at depth, the sperm whale made a slow pitch movement and created

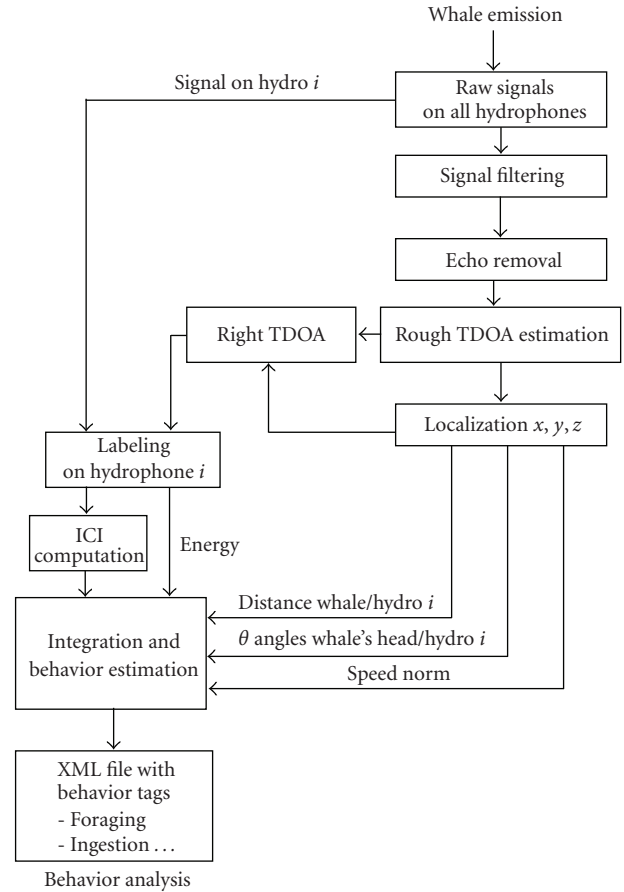


FIGURE 6: Block scheme of the features extraction. The forward arrow of the localization box points the robust TDOA filtering.

a faster pitch or yaw movement in synchronization with the clicking activity. [7] has suggested that sperm whales, at depth, make an asymmetric scan of the surrounding water and an ICI modulation [15], and that during the search phase sperm whales methodically scan a cone-shaped mass of water searching for prey. This scanning behavior suggests that each sperm whale click is generated to aim in a specific direction at a specific range. Sperm whales could move physically to change the click beam direction, and control level and ICI to change the click target range. [7] then points out a correlation between click level variations and ICI. A hypothesis explaining such a correlation relies on click level control: sperm whales click slowly at a high source level and faster at a lower source level. In [7] the proposed echolocation model specifies that sperm whales aim at closer targets while emitting clicks with lower source level at higher rhythmic rate, and aim at farther targets emitting clicks with higher source level and slower rhythmic rate. Finally, [15, 16] suggest that sperm whales move at constant heading and speed while hunting. The synthesis of all of the results leads the authors to suggest that sperm whales follow a dive-scale strategy when foraging. Sperm whales preanalyze their prey layer while descending, and cross this layer by repeating a methodic scanning/catching

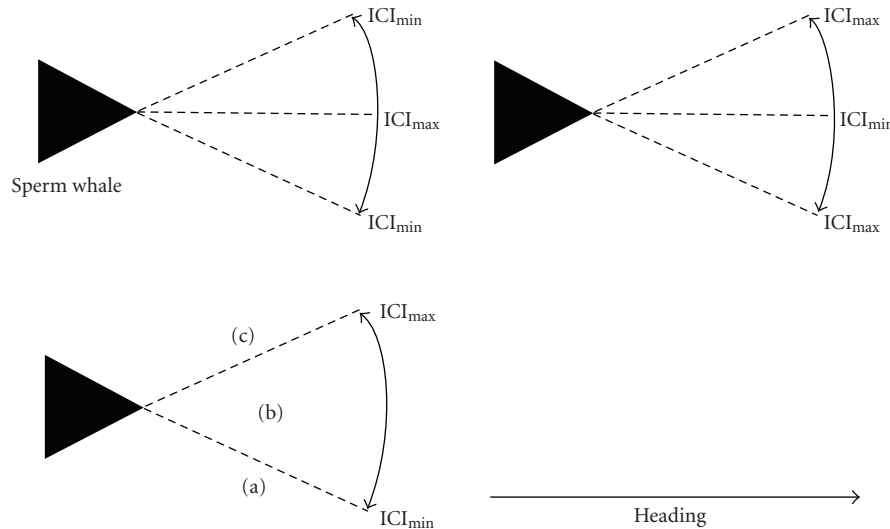


FIGURE 7: Three models describing the fast, periodic movement of the whale in sync with its ICI. Only the third model could lead to the synchronization $ICI_{\max} \Leftrightarrow \text{amplitude maximum}$ using a hydrophone located either in front of or behind the whale (it is located in (a), on the right of the whale). The click apparent level and the ICI are in sync ($ICI_{\max} \Leftrightarrow \text{Energy}_{\max}$) if the hydrophone is located in (c) (on the left of the whale). There is no apparent synchronization between the ICI, the off-axis angle, and the click energy if the hydrophone is in sector (b) (straight). The figure is from [7].

technique, before resurfacing. In Figure 5, we can see that pitch and yaw are present during an active search, and also a high correlation between ICI and energy. The gaps (silences) denote a simple direct catch of the prey. There is always a decrease of energy and ICI during the presumed prey attack. Pitch and yaw cannot be estimated because of being very weak.

Each scan would last 5 to 15 s and analyzes the water twice between the upper ranges r_{\max} , the lower range r_{\min} , and the upper range r'_{\max} . This scan suggests that each sperm whale click is generated to aim in a specific direction at a specific range. Sperm whales move physically to change the click beam direction, and control level and ICI to change the click target range.

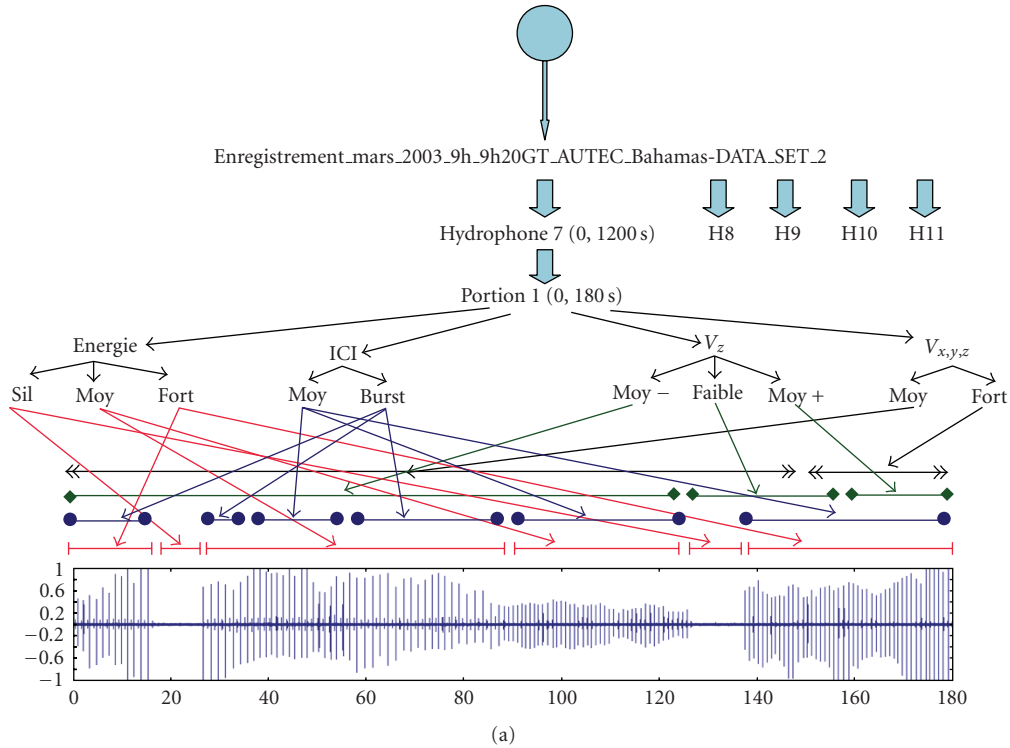
Three models were proposed in [7], describing the fast, periodic, rotational movement of the whale in sync with its ICI (Figure 7). Such a rotational movement is assumed to be planar in the following. The first model stipulates that the whale echolocates at slow rate ICI_{\max} in front of it and faster ICI_{\min} on its sides. The second model stipulates that the whale echolocates at high-rate ICI_{\min} in front of it and slower ICI_{\max} on its sides. The third model stipulates that the whale echolocates at slow-rate ICI_{\max} on one side and faster ICI_{\min} on the other. The models will have to be verified on more data.

4.2. XML Representation. All of the features are computed versus time, so we can index the sound file. Figure 8(a) is a graphical XML (Extensible Markup Language) representation of a system which aims to seek in a recording signal blocks values or descriptions for the behavior analysis of

the whale in a sound file, while observing the temporal signal. XML structure is an easy way to rapidly search for information in large whale recording databases. Figure 8(b) is an example of an XML structure script for the whole system. Several tags describe the time and duration of the events as well as values on behavior such as speed, ICI values, energy, distance, and angles. For a correct analysis, behavior interpretation with the tags must be done by the operator (cetologist).

5. Conclusion

A real-time tracking algorithm was used to extract parameters in order to index hydrophone audio files for one or more vocalizing sperm whales. Results indicate that only one sperm whale was present in the area, unless other whales in the area were quiet during the selected 25-minute period. Thanks to the localization, the signals could be labeled and features extracted. The features are used to index the raw audio files. The indexing leads to an XML structure (Figure 8) with behavior tags of the file that allow rapid access of information such as prey ingestion, prey search phase. Our results for the localization are validated by similar results from the US Navy and the University of Hawaii labs in the case of one whale. The models described in [7] fit our observation of the ICI and the energy modulation. Future works will consist in expanding the current algorithm to multiple whales, but in the single whale case this structuring method is robust for analysis and study of sperm whales and more generally for clicking marine mammal.



```

<?xml version="1.0"?>
<DATABASE>
<DATABASE id="SET2" desc="Dataset containing clicks of one whale"
date="March 2003" start_time="9h" end_time="9h20" location="AUTEC
Bahamas"/>
</DATABASE>

<HYDROPHONE>
<HYDROPHONE id="HYDR_7" length="5min" sampling_frequency="48kHz"
cartesian_coordinates_x_y_z="10658 -14953 -1530" desc="Hydrophone on
dataset2"/>
<HYDROPHONE id="HYDR_8" length="5min" sampling_frequency="48kHz"
cartesian_coordinates_x_y_z="12788 -11897 -1556" desc="Hydrophone on
dataset2"/>
<HYDROPHONE id="HYDR_9" length="5min" sampling_frequency="48kHz"
cartesian_coordinates_x_y_z="14318 -16189 -1553" desc="Hydrophone on
dataset2"/>
<HYDROPHONE id="HYDR_10" length="5min" sampling_frequency="48kHz"
cartesian_coordinates_x_y_z="8672 -18064 -1361" desc="Hydrophone on
dataset2"/>
<HYDROPHONE id="HYDR_11" length="5min" sampling_frequency="48kHz"
cartesian_coordinates_x_y_z="12007 -19238 -1522" desc="Hydrophone on
dataset2"/>
</HYDROPHONE>

<Portion HYDROPHONE="HYDR_7" start_time="0s" end_time="1200s">
<Section start_time="0s" end_time="180s">
<Energy value="Silence" desc="Energy=0">
Start_time="15s" end_time="25s"
Start_time="130s" end_time="140s"
...
<Speed="Mean+" desc="Speed>12m/s">
NULL
<ICI="Burst" desc="ICI>1.5s">
Start_time="0s" end_time="19s"
Start_time="22s" end_time="30s"
...
<Distance="Low" desc="Distance<1000m">
Start_time="0s" end_time="180s"
...
<Angles="Low-Low" desc="Theta1<30, Theta2<30 degrees">
NULL
...
<Behavior="Foraging" desc="If ICI>0.8s and energy>5 ">
Start_time="0s" end_time="19s"
Start_time="22s" end_time="30s"
...
<Behavior="Ingestion" desc="If Silence and previous ICI<0.5s">
Start_time="130s" end_time="140s"
...
</Section>
</Portion>

<Portion HYDROPHONE="HYDR_8" start_time="0s" end_time="1200s">
...

```

FIGURE 8: At (a), XML structure of the hydrophones recordings. At (b), part of an XML whale record analysis including behavior tags. For each tag is attributed a quantitative or qualitative value and also the start and end times.

Acknowledgments

The authors thank AUTECH and NUWC for recording and distributing the dataset, Eva-Marie Nosal and Pascal Giraudet for their help, and Olivier Adam for the figure. 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)”. This work is funded by the “Conseil Régional Provence-Alpes-Côte d’Azur,” France, and partly by Chrisar Software Inc.

References

- [1] F. Caudal and H. Glotin, “Stochastic matched versus teager-kaiser-mallat filters for tracking simultaneous clicking whales,” in *Proceedings of the International Workshop on New Trends for Environmental Monitoring Using Passive Systems (PAS-SIVE ’08), IEEE Xplore*, IEEE, Hyeres, France, October 2008.
- [2] H. Glotin, F. Caudal, and P. Giraudet, “Whale cocktail party: real-time multiple tracking and signal analyses,” *Canadian Acoustics*, vol. 36, no. 1, pp. 141–147, 2008.
- [3] F. Caudal and H. Glotin, “Automatic inter-clickinterval (ICI) and behavior estimation for one emitting sperm whale,” in *Proceedings of the International Workshop on New Trends for Environmental Monitoring Using Passive Systems (PAS-SIVE ’08), IEEE Xplore*, IEEE, Hyeres, France, October 2008.
- [4] P. Giraudet and H. Glotin, “Real-time 3D tracking of whales by echo-robust precise TDOA estimates with a widely-spaced hydrophone array,” *Applied Acoustics*, vol. 67, no. 11-12, pp. 1106–1117, 2006.
- [5] V. Kandia and Y. Stylianou, “Detection of sperm whale clicks based on the Teager-Kaiser energy operator,” *Applied Acoustics*, vol. 67, no. 11-12, pp. 1144–1163, 2006.
- [6] J. F. Kaiser, “On a simple algorithm to calculate the ‘energy’ of a signal,” in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP ’90)*, vol. 1, pp. 381–384, Albuquerque, NM, USA, 1990.
- [7] C. Laplanche, O. Adam, M. Lopatka, and J.-F. Motsch, “Measuring the off-axis angle and the rotational movements of phonating sperm whales using a single hydrophone,” *The Journal of the Acoustical Society of America*, vol. 119, no. 6, pp. 4074–4082, 2006.
- [8] P. R. White, T. G. Leighton, D. C. Finfer, C. Powles, and O. N. Baumann, “Localisation of sperm whales using bottom-mounted sensors,” *Applied Acoustics*, vol. 67, no. 11-12, pp. 1074–1090, 2006.
- [9] D. W. Marquardt, “An algorithm for least-squares estimation of nonlinear parameters,” *SIAM Journal on Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.
- [10] R. P. Morrissey, J. Ward, N. DiMarzio, S. Jarvis, and D. J. Moretti, “Passive acoustic detection and localization of sperm whales (*Physeter macrocephalus*) in the tongue of the ocean,” *Applied Acoustics*, vol. 67, no. 11-12, pp. 1091–1105, 2006.
- [11] E.-M. Nosal and L. N. Frazer, “Track of a sperm whale from delays between direct and surface-reflected clicks,” *Applied Acoustics*, vol. 67, no. 11-12, pp. 1187–1201, 2006.
- [12] F. Bénard and H. Glotin, “Sperm whales records indexation using passive acoustics localization,” in *Proceedings of the 7th International Workshop on Content-Based Multimedia Indexing (CBMI ’09)*, pp. 207–212, Crete, Greece, June 2009.
- [13] P. T. Madsen, R. Payne, N. U. Kristiansen, M. Wahlberg, I. Kerr, and B. Møhl, “Sperm whale sound production studied with ultrasound time/depth-recording tags,” *Journal of Experimental Biology*, vol. 205, no. 13, pp. 1899–1906, 2002.
- [14] B. Mohl, M. Wahlberg, and P. T. Madsen, “Ideal spatial adaptation via wavelet shrinkage,” *The Journal of the Acoustical Society of America*, vol. 114, pp. 1143–1154, 2003.
- [15] C. Laplanche, O. Adam, M. Lopatka, and J.-F. Motsch, “Male sperm whale acoustic behavior observed from multipaths at a single hydrophone,” *The Journal of the Acoustical Society of America*, vol. 118, no. 4, pp. 2677–2687, 2005.
- [16] C. Laplanche, *Etude par acoustique passive du comportement de chasse du grand cachalot*, These de doctorat, Université Paris XII Val-de-Marne, Paris, France, 2005.