5.1 Mono-Channel Spectral Attenuation modeled by Hierarchical Neural Net Estimates Hydrophone-Whale Distance

Doh Y.
DYNI team
Aix-Marseille Université, CNRS, ENSAM, LSIS, UMR 7296, 13284 Marseille, France
Université de Toulon-CNRS, LSIS, UMR 7296, 83957 La Garde, France
yanndoh.m2@gmail.com

Glotin H.
IUF, Institut Universitaire de France
103 Bd. Saint-Michel, 75005 PARIS - France
DYNI team
Aix-Marseille Université, CNRS, ENSAM, LSIS, UMR 7296, 13284 Marseille, France
Université de Toulon-CNRS, LSIS, UMR 7296, 83957 La Garde, France
glotin@univ-tln.fr

Razik J.
DYNI team
Aix-Marseille Université, CNRS, ENSAM, LSIS, UMR 7296, 13284 Marseille, France
Université de Toulon-CNRS, LSIS, UMR 7296, 83957 La Garde, France
Joseph.Razik@univ-tln.fr

Paris S.
DYNI team
Aix-Marseille Université, CNRS, ENSAM, LSIS, UMR 7296, 13284 Marseille, France
Université de Toulon-CNRS, LSIS, UMR 7296, 83957 La Garde, France
sebastien.paris@lsis.org

Abstract

We aim to allow whale monitoring and anti-collision system using single hydrophone. We then propose a new model to estimate the range from wideband signals such as clicks emitted by odontocetes. We demonstrate that it is possible to link the intrinsic distortion of the signal with the distance of the acoustic path. We provide different models to establish the relationships between the signal energy and propagation distance. We deal with different energy scales: the global received energy of the signal $E_0$, the frequency bands energy and frequency bin energy. We then demonstrate that intermediate prediction of the whale orientation enhances the distance estimation, yielding to only 6% of relative error rate.

1 Introduction

Passive acoustics is one of the best ways to enhance the knowledge of marine mammals emitting sounds through various tasks: detection, classification, localization and density estimation. 3D whales localisation is mostly achieved using hydrophone arrays. Although these methods have been em-
ployed successfully for this task on mysticetes [6] and odontocetes [1, 3] with a high level of accuracy, they require the use of heavy and expensive hardware. Despite of the information loss, using a single, light and cheap hydrophone device, quick to deploy, could provide the necessary data to satisfy certain applications such as mobile listenning point and anti-collision system where a simple range estimation is sufficient. Therefore, in this paper, we decided to focus on single-hydrophone methods. Theoretically, it is possible to use the virtual hydrophone framework and the acoustic property of the water column in order to estimate the position of the whale [4, 7, 5]. This technique involves the acquisition of the direct acoustic path, the bottom reflection and the surface reflection. However, in practice, we often observe only a subset of the needed information. The above mentioned constraints led us to find a new model of range estimation applied to wideband signals such as clicks emitted by odontocetes. Specifically, this paper applies the proposed model on sperm whale recordings. In order to focus on range estimating in this first approach, and to avoid source separation problems, we choose to deal with single animal recordings.

It is well known that sound attenuation depends on frequency and propagation distance [8, 9]. The attenuation impacts the total energy and generates a distortion of frequency representation in the emitted signal that we can link to the distance between source and receiver. Madsen et al. in 2002, put forward the variation of the centroid with the distance et suggest a low pass effect provided by the acoustic propagation [18].

In this paper, we try to establish the expression of the relationship between the signal energy and propagation distance by an empirical model based on a neural network and by the theoretical model Inter-Frequency Attenuation (IFA) [10, 11].

Real data and their associated ground truth [3, 12, 13, 14], will allow us to develop this model by optimizing a few but important parameters. Also, an independent partition of the data will be useful to test the ability and limits of the proposed estimator.

2 Motivation

2.1 The relationship between the received signal and loss by transmission

Our goal is to extract information regarding the propagation distance, from the observed signal from a unique hydrophone. The relationship between Transmission Loss (TL) and distance, as provided by the passive sonar equation, is not well adapted to bioacoustics signal applications. Solving this equation depends mainly on the signal’s power at the origin Source Level (SL). Emitting with a variable sound level, a sperm whale is not a constant acoustic energy source. Moreover, we must also take into account other variability factors like the animal’s size, the Inter-Click Interval (ICI) and Inter-Pulse Interval (IPI), diving depth [18] or the directionality of the animal relative to the hydrophone position [2, 17, 20, 21].

First we introduce the expression of the received energy \( E \) at the distance \( r \) from the source, as a function of the energy source level \( E_{SL} \) and TL [15, 7] for a given frequency in dB. We consider the simple framework of omnidirectional spherical source:

\[
E(r, f) = E_{SL}(f) - TL(r, f),
\]

where the transmission loss TL can be decomposed by

\[
TL(r, f) = 20 \log_{10}(r) + \alpha(f)r,
\]

where \( r \) is the propagation distance (in m), \( f \) is the frequency (in Hz) and \( \alpha \) is the frequency attenuation coefficient (\( dB.m^{-1} \)). The first term of Eq. (2) is due to loss by geometric divergence of a spherical wave while the second term represents the frequency attenuation because of interactions between the wave and the medium.

In a first approximation, we can assume that the loss by divergence is predominant on frequency attenuation and TL does not depend on frequency which allows to consider the total energy \( E_0 \) of the signal.
The problem \( r = F(E_0) \) remains unsolved without a theoretical or statistical model of the energy source level. This function can be approximated and learned by a neural [24] network in particular a MLP, since MLPs are universal function approximators and will be described in a further section. It will be the basic model LER (for Loss Estimation Regression). But this function is empirical and very dependant on the data used to learn the model. Thus, it has motivated us to find a theoretical relationship only based on the frequency attenuation which could work without any knowledge of the total SL. This model imposes to consider not only the total energy but also the detailed frequency composition.

### 2.2 Data driven IFA regressions

![Diagram of different neural network models](image)

Based on the simplification done, we expect to have limited results using the theoretical model of the IFA. We also used a neural network algorithm [24] to learn a regression model to fit the relation between spectral informations and distance estimations such as:

\[
\hat{r} = h(|X(f; r); W|),
\]

where \( W \) represents the matrix of weights provided by the training step.

We trained a MLP to learn an empirical relationship between spectrum and radial distance. Spectrum bins \( |X_j(k\Delta F)| \) will be used as input to the network and ground truth radial distance \( d_j \) as the output. As presented in previous sections regarding the IFAT estimators, we will work on \( N = 256 \) samples. It will be the Model \( IFAR \). The used MLP is comprised of two hidden layers which are composed by \( 2N + 1 \) units. The MLP attributes the optimal weights describing the regression between spectrums bins and propagation distance.

We know that information on the animal directionality can be extracted from the click spectrum and its energy. Therefore we propose the model \( IFARH \) in order to enhance the model by associating in “waterfall” 2 models. A first one learns a regression estimating the 3 angles describing the position of the animal: off axis \( o_\circ \), azimuth \( a \) and elevation \( el \). A second one learn a regression function given the estimated angles \( \hat{o}, \hat{a} \) and \( \hat{el} \) such as:

\[
\]
\[ \hat{r} = h(\{|f; r|\}; \hat{\theta}; \hat{\alpha}; \hat{\ell}; \mathbf{W}). \]  

(4)

2.3 Proposition of a theoretical Model : Inter Frequency Attenuation (IFAT)

The proposed Theoretical Inter Frequency Attenuation (IFAT) [10] model aims to extract information from the source distance by taking advantage of the energy ratio between two frequency bands of the emitted signal. The derivation of the attenuation laws allowed us to establish the following relationship:

\[ r(B_1, B_2) = \frac{10 \log_{10} \left( \frac{E_1}{E_2} \right)}{\int_{F_1}^{F_2} \alpha(f) df - \int_{F_1}^{F_2} \alpha(f) df}. \]  

(5)

In this expression, \( r \) is the acoustic propagation distance, \( B_1 = [F_1, F_2] \) and \( B_2 = [F'_1, F'_2] \) are the frequency band involved, \( F'_1, E_1 \) and \( E_2 \) the energy of band 1 and 2. In this expression \( r \) does not depend on loss by divergence and energy at the origin, but it only depends on frequency attenuation.

3 Material

In this section, we present our dataset which is extracted from the Bahamas dataset distributed by AUTEC at the second DCL workshop in Monaco 2005. It consists of five hydrophones deployed off the Bahamas Island, and a total of 25 minutes of recording of one sperm whale with a sample rate of 48 KHz. The trajectory computed by LSIS/DYNI (Fig. 2) [3, 14] is similar to the one resulting by different methods by the scientific community [22, 23], and it will be considered as the ground truth.

Figure 2: The 2D trajectory (in \( x - y \) plane) of the single sperm whale observed during 25 min (LSIS/DYNI [3]) and corresponding hydrophone’s positions. The whale goes to south east. Supplemental material with the animated 3D tracking of this whale is available at http://glotin.univ-tln.fr/oncet and http://www.youtube.com/watch?v=0Szo3gdiTRk. We also give there in supplemental material with the file containing the \([x, y, z, t]\) whale positions, and the \([az, el, offaxis, t]\) files for each hydrophone H8.. H11.
Table 1: Mean ground truth distance, azimuth and hydrophone depth

<table>
<thead>
<tr>
<th></th>
<th>H11</th>
<th>H10</th>
<th>H9</th>
<th>H8</th>
</tr>
</thead>
<tbody>
<tr>
<td>depth</td>
<td>-1522 m</td>
<td>-1361 m</td>
<td>-1553 m</td>
<td>-1556 m</td>
</tr>
<tr>
<td>off axis</td>
<td>38 degrees</td>
<td>60 degrees</td>
<td>74 degrees</td>
<td>129 degrees</td>
</tr>
<tr>
<td>elevation</td>
<td>-14 degrees</td>
<td>-15 degrees</td>
<td>-14 degrees</td>
<td>-13 degrees</td>
</tr>
<tr>
<td>distance</td>
<td>3937 m</td>
<td>2900 m</td>
<td>4150 m</td>
<td>4716 m</td>
</tr>
<tr>
<td>std deviation</td>
<td>452 m</td>
<td>242 m</td>
<td>234 m</td>
<td>283 m</td>
</tr>
</tbody>
</table>

Considering the four hydrophones H8, H9, H10, H11, the range of distance between source and receiver is from 2500 m to 5500 m. The precise angle associated to the ground truth trajectory, has been calculated. The mean values are mentioned in Tab. 1.

We then divide the data in 2 partitions. Partition 1 will be used for the development and parameter optimization step. Partition 2 will be dedicated to estimation and predictions.

4 Results

4.1 IFAR and IFARH IFA estimation using MLP model

4.1.1 Training and development

The training step is employed on partition 1 of the data. The temporal order of the data is not taken into account when applied as the input of the MLP. During the development stage, we optimized only one parameter, which is the number of system iterations for the training session (early stopping) related to the quality of prediction.

Predictions are generated from partition 2 of the data and we assume independence with the training data. We were surprised to observe that the MLP learned the relationship between spectrum and propagation distance quite quickly. 300 iterations are sufficient to obtain satisfactory results for our prediction. In order to avoid over-fitting and to extract the most general predictor we keep the number of iterations low.

4.1.2 Distance prediction

In this section, we propose a temporal MLP prediction following $L E R$, $I F A R$ and $I F A R H$ on all hydrophone in the same data subset as section IV.A.2 for the IFAT estimator. It has been computed using the MLP for regression on the spectrum.

In Fig. 3 we can compare the fidelity of prediction and ground truth (results summarized in Tab. 3). Firstly, the prediction given by $I F A R$ and $I F A R H$ is better than $L E R$ one. It demonstrates the usefulness of considering spectrum inter bins and not only the global energy of the signal. $L E R$ shows some transitions which lead to a recess that others models seem to control. However, the bias seems to increase in the sections where the azimuth is the lowest (20 degrees). This behavior can be explained by a lack of an on-axis configuration during the training session. This assumption means that the regression law is different between an on-axis and off-axis configuration. It may also be caused by the different frequency structure of pulses (P or PJ) in a click following the receiver position [16].

Then, $I F A R H$ mean error is similar to $I F A R$. However, the dispersion of the predicted distances seems enhanced by the use of an intermediate MLP predicting the position angles. The Results of angle estimation are not presented directly in this paper. We noted that azimuth prediction was more precise than elevation prediction which could suggest that the spectrum shape is more dependent on azimuth.

4.2 Estimation of the radial distance from the theoretical model IFAT

We represent the final temporal estimation $r^*$ of radial distance between hydrophone 11 and the sperm whale. The computation of the estimators has been implemented on Partition 2 of the recordings (test set) according to $F_P^*$ and $N_{best}^*$ learned on the train set data (see previous section).

Figure 3: Temporal prediction of radial distance compared to ground truth using model LER, IFAR, IFARH.
Figure 4: An example of final IFAT temporal estimation of radial distance compared to ground truth radial distance and azimuth evolution on H11. fixed parameters: \( N_{\text{best}} = 4 \), \( F_{P125}^* = 8.5\,\text{kHz} \) and \( F_{P256}^* = 9\,\text{kHz} \).

In Fig. 4 we see that both curves have almost the same dispersion. Using \( N = 128 \) samples, the distances are overestimated, while underestimated using \( N = 256 \) samples.

The significant dispersion of our estimation cannot be due to the range variation. It can be related to the animal’s off-axis variation, meaning that IFAT does not cancel the animal’s directionality effects. Most of the errors of IFAT using \( N = 128 \) samples seem to accumulate on sections where the azimuth is high (> 20\,\text{deg}). When the animal is supposed to be on-axis, we observe that the estimation is converging to the ground truth.

On the other hand, for the \( N = 256 \) samples, the lowest error seems to match with high azimuth, and it results in a better estimation on high off-axis configurations. Thus the average between \( N = 128 \) and \( N = 256 \) samples is computed to provide a uniform behavior relative to the azimuth possibilities.

Finally, we also observe that the radial distance is decreasing which is in agreement with the directionality of the whale on the ground truth. IFAT correctly affirms that the whale is traveling towards the hydrophone.

5 discussion

According to the results (tables 2, 3) IFAT errors tend to vary between the different hydrophones, but the IFAT model and the neural network led to common conclusions. The performance of IFAR and IFARH confirms the usefulness to consider spectrum inter-bin property and not only the total energy which is more sensitive to natural and voluntary variations of the Source Level.

The IFAT, IFAR, IFARH and LER models provide us with differences regarding the azimuth. With the IFAT model, the performance of the estimator seems to be more impacted by the azimuth configuration. Since we developed the model without the consideration of the animal’s directionality, a future step may be the inclusion of this variable in the estimator expression.

The MLP demonstrates the relationship between spectrum and propagation distance in this data set on only 300 iterations. We demonstrated that the IFAT model could estimate distance with about 15% of mean relative error, while the MLP IFAR or IFARH reduces it to 6%.

As IFAT produces locale range estimates, a particle filtering process [19] could be efficiently added after IFAT in order to produce more reliable and complex estimates.

In the case of multiple emitting whale and a monohydrophone recording, IFA would play an important role in order to cluster the clicks and thus estimate the number of emitting whales. Moreover, in the case of multiple hydrophones, IFA can help in the localisation of each whale. The IFAR and
<table>
<thead>
<tr>
<th></th>
<th>H11</th>
<th>H10</th>
<th>H9</th>
<th>H8</th>
<th>MEAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFA T model</td>
<td>21</td>
<td>16</td>
<td>14</td>
<td>41</td>
<td>23</td>
</tr>
<tr>
<td>IFA model</td>
<td>20</td>
<td>21</td>
<td>9</td>
<td>46</td>
<td>16.5</td>
</tr>
<tr>
<td>average on</td>
<td>17</td>
<td>16</td>
<td>11</td>
<td>28</td>
<td>18</td>
</tr>
<tr>
<td>N = 128 and</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 256</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean total</td>
<td>19</td>
<td>17</td>
<td>11</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>relative error</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Absolute mean relative error of estimated distance in % for all hydrophones with IFAT theoretical model estimators

<table>
<thead>
<tr>
<th></th>
<th>H11</th>
<th>H10</th>
<th>H9</th>
<th>H8</th>
<th>MEAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFA T mean</td>
<td>14±6</td>
<td>20±10</td>
<td>8±9</td>
<td>15±5</td>
<td>14.2±7.5</td>
</tr>
<tr>
<td>IFA H mean</td>
<td>20±10</td>
<td>21±13</td>
<td>9±7</td>
<td>16±14.0</td>
<td>16.5±11</td>
</tr>
<tr>
<td>IFARH mean</td>
<td>4±2.5</td>
<td>11±3</td>
<td>4±3.5</td>
<td>4±2</td>
<td>5.75±2.75</td>
</tr>
</tbody>
</table>

Table 3: Absolute mean relative error of estimated distance by with MLP: IFAR/IFARH and standard deviation (in %)

IFARH models trained on a data set, can be applied on another recording set for similar species and hydrophones. Also one can model and run IFA for other species using biosonar, like bats.

![Figure 5: mean relative error between predictions and groundtruth. Different IFARH model versions have been tested: IFARH(o,a,el), IFARH(a,el), IFARH(o), IFARH(a), IFARH(el).](image)

Acknowledgments

The authors gratefully acknowledge the contribution of the Provence Alpes Cote d Azur region, Université Toulon Var, CESIGMA company and the anonymous referees.

References


