

A NOVEL APPROACH TO REAL-TIME RANGE ESTIMATION OF UNDERWATER ACOUSTIC SOURCES USING SUPERVISED MACHINE LEARNING

Ludwig Houégnigan¹, Pooyan Safari², Climent Nadeu², Mike van der Schaar¹, Michel André¹

¹ Laboratory of Applied Bioacoustics (LAB), Polytechnic University of Catalonia (UPC), Spain

² TALP Research Center - Dept. TSC, Polytechnic University of Catalonia (UPC), Spain

ABSTRACT

The proposed paper introduces a novel method for range estimation of acoustic sources, both cetaceans and industrial sources, in deep sea environments using supervised learning with neural networks in the context of a single sensor, a compact array, or a small aperture towed array. The presented results have potential both for industrial impact and for the conservation and density estimation of cetaceans. With an average error of 4.3% for ranges up to 8 kilometers and typically below 300 meters, those results are challenging and to our knowledge they are unprecedented for an automated real-time solution.

Index Terms— range estimation, neural networks, source localization, array processing, acoustics, density estimation

1. INTRODUCTION

In underwater acoustics and in particular in passive acoustic monitoring, range estimation capability is a much appreciated asset. Range estimation is indeed a key information to mitigate, study, analyze and model the propagation and effects of anthropogenic sound. In geophysical surveys, legal regulations were enforced requiring the estimation of an exclusion zone, a perimeter of a certain diameter free of cetaceans, so as to mitigate the potential impact of exploration.

Range hence becomes a key parameter with the capability to interrupt costly activities at sea. In other circumstances, such as at windmill farms, while noise impact on the marine fauna is being monitored, detection and range estimation of distant sources such as marine mammals provides important information on the interaction between marine mammals and anthropogenic noise. Range is also central in the estimation of the abundance of marine mammals based on their vocalizations: in ecological studies, the range of vocalizing animals allows to derive a detection function that is central in density estimation.

Localization experiments and animal telemetry studies have allowed to derive models describing the species-specific probability of detecting animal vocalizations with regard to range, yet the scarcity and cost of this data is such that Monte Carlo models of detection probability have been constructed. The proposed method could participate in the design of density estimation models as it updates the expected efficiency of range estimators and their statistical properties. Furthermore this supervised estimation method is empirically related to the specific

propagation characteristics of the environment as opposed to general and simulated.

The proposed approach also tackles, all together, the limitations due to large deployment or maneuvering costs, and the challenges of automation, real-time and generalization.

2. APPROACHES TO RANGE ESTIMATION

2.1 Range estimation from multiple widely-spaced hydrophones

Range estimation is done most easily with multiple widely-spaced hydrophones offering an adequate aperture. This is the less adverse yet not a frequent solution due to deployment and data management costs. In such conditions some difficulties may still arise in particular if (1) the signal of interest cannot be captured on all of the sensors if (2) the acquired signals have synchrony problems or when (3) associating multiple signals on multiple hydrophones that may or may not have been produced by several animals. Furthermore a signal impinging on widely-spaced sensors may have undergone different propagation paths, with reflection, diffraction and reverberation phenomena yielding a complex pattern recognition problem of noisy, incoherent and distorted pairs of signals producing noisy time-delay estimates for which robust methods are fundamental [12].

A set up of widely spaced sensors is therefore not so frequent, and for maneuverability and deployment costs, single sensors, towed arrays or compact volumetric arrays are more often preferred. Those apertures do not typically allow to sense range in a straightforward manner, hence the development of several methods to extract from recordings that essential variable.

2.2 Using surface and/or bottom reflections

In [10] Thode et al. proposed a creative solution using surface reflections to estimate the range and depth of acoustic sources. This was followed by further practical work applied to marine mammals by Zimmer[1], Nosal et al[13], Thiemann et al[9].

Zimmer proposed 2 original methods using compact arrays. First by combining the estimated angle of incidence of the direct path and the time-delay to a surface reflection, using constant sound speed and geometrical derivation to obtain depth and range and secondly using the same derivation with a ray tracing model and a realistic sound speed profile. Zimmer forecasted an average error

on range below 200m but that error may greatly vary based on the quality of direct and reflected time-delay estimates.

Even though the developed methods can perform in real-time, the presence of reflections and their automatic detection cannot be guaranteed and have to be opportunistic while the estimation of the time of arrival/time-delay of surface reflections was found to be a major potential contributor to errors on range estimates.

2.3. Modal decomposition

In the lower frequency range (below 500 Hz), the dispersive effects of propagation- for example on the vocalization of baleen whales- can be described through normal mode theory. The different group speed propagation of each mode renders the time-frequency dispersion of the modes characteristic of the call's range and permit its reconstruction as proposed by Bonnel in [1]. Among multiple sensors of a widely spaced vector sensor array, a single hydrophone was used for range estimation based on modal propagation and the obtained results were contrasted with the ground truth obtained with the other remaining sensors. A key aspect of the method is the use of a time-domain warping algorithm which allows a separation of the propagation modes and thus an accurate description of the waveguide at stake. However the warping method is for now restricted to monotonic frequency-modulated calls-“upsweeps” and “downsweeps”- and, most importantly, automation is rendered difficult firstly by the identification of instantaneous frequencies and secondly by the difficultly predictable number of iterations necessary to converge towards a correct modal separation.

3. SUPERVISED APPROACHES TO ACOUSTIC SOURCE LOCALIZATION AND RANGING

Advances in machine learning and most notably in Deep Learning [19] invite us to revisit some signal processing problems and applications in order to challenge well-established techniques and set new performance results. In that frame, over the past years, the authors of this paper, through different publications [11,12] have proposed new approaches using machine learning and data-specific modelling in order to solve problems related to both air and underwater acoustic source localization (biological sources such as cetaceans, or artificial ones such as pingers, ships, navy sonar, etc).

3.1 Supervised time-delay estimation

In [12] time-delay estimation was reformulated as a pattern recognition problem to which supervised learning using multilayer perceptrons provided a particular solution, a reformulation and extension of earlier work by Shaltaf [14]. A comparison with standard methods demonstrated that a well-trained neural time-delay estimator could account for significant non-random data structure and reduce bias and variance error by up to 90% in extreme cases where traditional methods such as the generalized cross-correlation and adaptive eigenvalue decomposition typically provided random and inconsistent results and were thus unsuitable for automated localization and tracking.

3.2 Supervised multilateration

It was also found [11] that adequately trained neural networks could provide a robust solution to multilateration. The supervised

approach artificially generated noisy time-delays as high level input features that were associated to pre-determined spatial locations and sound speed profiles. This provided consistent closed-form solutions, an enjoyable feature for real-time localization, and is a preliminary step providing robust position estimates for spatial tracking. This approach was demonstrated on data from the ANTARES deep sea neutrino telescope which was also used to detect the presence of cetaceans and in particular helped in the long-term monitoring of sperm whales.

3.3 Supervised range estimation using neural networks

Range estimation using neural networks comes as a logical follow up to previous approaches to learn meaningful structure from time and spatial data and to re-express localization as a pattern recognition problem that is by no means different from detection and classification. This allows the use of similar machine learning tools for all passive monitoring tasks. The fundamental intuition behind the proposed range estimation scheme relates to the physical effects of propagation through absorption and transmission loss, through which signal frequency distribution is altered in a very specific way which is related to the travelled range (eq.1). This supervised approach proposes to use propagation effects (transmission loss and absorption) as a form of “audio watermark” from which range can be partly reconstructed.

The sonar equation (eq.1,[3]), expresses the relationship between received levels and, on the one hand, propagation and environmental effects such as background noise, transmission and directional loss at emission, as well as, on the other hand, acquisition effects :

$$RL(R, \theta, f) = ASL(\theta, f) - TL(R, f) - NL(f) + AG + PG \quad (eq.1)$$

Where:

- $RL(R, \theta, f)$ represents the received level for respective range R , direction of emission θ , and frequency f .
- $ASL(\theta, f)$ represents the apparent source level as a function of direction of emission θ , and frequency f . For multiple cetaceans species beampatterns have been empirically derived as in [15,16,17] and allow to express the variation of level with regard to angle.
- $NL(f)$ is the frequency-related background and/or ambient noise level.
- AG and PG represent respectively the hydrophone (or array) gain and processing gain.

The transmission loss TL has both a range and a frequency dependent term. For spherical spreading we have

$$TL(R, f) = 20 \log(r) + \alpha_f(f)R \quad (eq. 2)$$

The finite response of sea water to sound pressure produces the absorption loss $\alpha_f(f)$ in sound. It has been empirically related to multiple parameters among which depth, frequency, salinity and acidity (pH).

A formula derived by François and Garrison [6,7], (eq. 1), divides the absorption coefficient α_f expressed in dB/km into three main contributions from boric acid, magnesium sulphate and water viscosity, at a depth z and at a sampling frequency f (kHz), sound velocity c (m/s), and salinity S (parts per thousand):

$$\alpha_f = A_1 P_1 \frac{f_1 f^2}{f_1^2 + f^2} + A_2 P_2 \frac{f_2 f^2}{f_2^2 + f^2} + A_3 P_3 f^2 \quad (eq. 3.0)$$

, where the boric acid term contains

$$P_1 = 1 \text{ and } A_1 = \frac{8.86}{c} 10^{(0.78pH-5)} \text{ (eq. 3.1)}$$

, where the magnesium sulfate term contains

$$A_2 = 21.44 \frac{S}{c} (1 + 0.025T)$$

$$P_2 = 1 - 1.37 \times 10^{-4} z + 6.2 \times 10^{-9} z^2,$$

$$\text{and } f_2 = \frac{8.17 \times 10^{\left(\frac{8 - \frac{1990}{273+T}}{1+0.0018(S-35)}\right)}}{1+0.0018(S-35)}, \text{ (eq. 3.2)}$$

For details about pure water viscosity we refer to [3] which displays the terms A3 and P3 in function of temperature and depth.

Ainslie and Mccolms provided simplified equations based on the same data and made explicit the different contributions while closely matching empirical data and providing a clearer physical insight with (eq.2):

$$\alpha_f = 0.106 \frac{f_1 f^2}{f_1^2 + f^2} e^{\frac{(pH-8)}{0.56}} + 0.52 \left(1 + \frac{T}{43}\right) \left(\frac{S}{35}\right) \frac{f_2 f^2}{f_2^2 + f^2} e^{-\frac{D}{6}} + 0.00049 f^2 e^{-(T/27+D/17)} \text{ (eq. 4)}$$

Table 1 presents absorption coefficients numerical values in dB/km predicted by (eq. 3.0) and (eq. 4) for the following environmental conditions: T=15 degree Celsius, depth is 5 m. salinity 35 (ppt), acidity 8 (pH).

Frequency	100 Hz	200Hz	400Hz	800Hz	1kHz
François et al. (eq. 3.0)	0.001	0.003	0.012	0.041	0.056
Ainslie et al. (eq. 4)	0.001	0.003	0.012	0.041	0.057
Frequency	1.5 kHz	5 kHz	10 kHz	100KHz	
François et al. (eq. 3.0)	0.092	0.304	0.836	37.058	
Ainslie et al. (eq. 4)	0.095	0.316	0.856	37.994	

Table 1, numerical values of absorption coefficients derived from eq 3.0 and eq 4. for different frequency values

From (eq.1), (eq.3.0) and (eq.4), it can be understood that observed noisy received levels in multiple frequency bins are significantly dependent on combinations of absorption coefficients that have a typical frequency distribution. Sensing range therefore can be seen as sensing the contribution of the absorption pattern in received signals. Broadband signals obviously contain more information to average noise and error from and reconstruct range from, however closely spaced frequency bins of tonal or harmonic signals typically share similar directional loss and also contain significant information. In the case of a multi- or wideband signal or for an unknown source level, a single noisy frequency bin is not sufficient to reconstruct range, however a combination of multiple

noisy frequency bins could be used to estimate the most likely range.

Numerical optimization or a Hidden Markov Model, where the hidden state would be the actual range and the observations could be a non-linear combination of the energy received in each frequency bins, could be derived. Which combination is optimal to reconstruct range is not trivial and for the sake of concision we leave this derivation for a more extended publication for wideband and multi-band signal. Instead this paper focuses on empirical measurements which seem to validate this hypothesis. Given the high and variable noise levels and the possible variability in source levels, a supervised learning approach was thus preferred. This approach should allow mapping non-linear and non-trivial relationships between a non-pre-determined number of frequency bands and may allow to include non-previously considered features of interest for range estimation.

4. EXPERIMENTAL SET-UP AND RESULTS

An aluminum array with four mounted hydrophones and a system for data acquisition were designed to perform real-time acoustic source detection and localization. Sensed signals were pre-amplified using 4 dedicated phantom power supplies and a 4 channels A/D converter was used to acquire at a rate of 60 kHz. This allowed for an acceptable resolution for time-delay estimation and angle estimation in a range of 250 samples. The compact case and array aperture permitted to deploy the whole system with ease from a small research vessel as depicted in figure 1.



Fig.1.1 Acquisition case with preamplifiers, laptop, A/D converter



Fig.1.2 View of the mounted antenna and acquisition case.

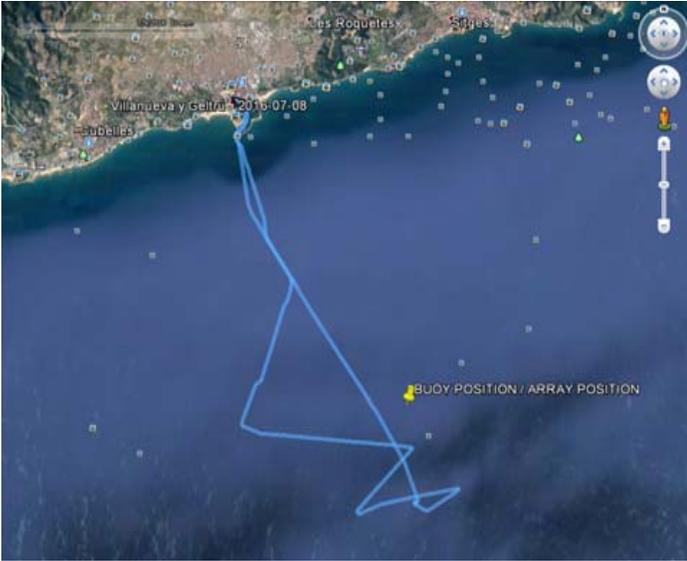


Fig.2: Track of catamaran-mounted loudspeaker along the coast of Garraf and fixed position of the array during acquisition (pin).

A central aspect of this work consisted in ensuring independence of the extracted spectral and cepstral features from sound levels so that range detection would hence only depend on the inter-frequencial and inter-cepstral relationships. Received levels are related to source level through transmission loss and absorption but the great variability of source levels for most marine mammals species [18] could strongly bias any range estimate. Detrending of the spectrum from received levels was achieved by applying intermediary normalizations to the frequency spectrum before reaching the final features and testing various feature normalizations (linear range scaling, linear scaling to unit range and linear scaling to unit variance) as depicted in figure 3, with the objective of separating the influence of source level on range estimation and make the estimation solely rely on the relationship between multiple frequency bands.

The style of the extracted features depicted in fig. 3.1 to 3.3 is reminiscent of cepstral features where the logarithmic function has been replaced by a normalization function. An earlier attempt at using the true cepstral features to feed the neural architecture did not however provide satisfying range prediction capabilities. Linear range scaling and linear scaling to unit variance produce first features of significantly higher amplitude while linear ranging to unit variance seems to produce a more random feature distribution.

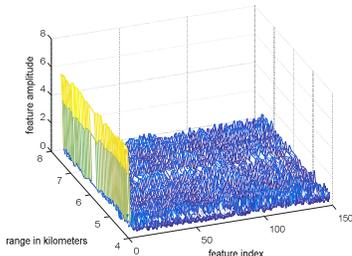


Fig. 3.1: Feature space (Linear range scaling) with respect to range (in km)

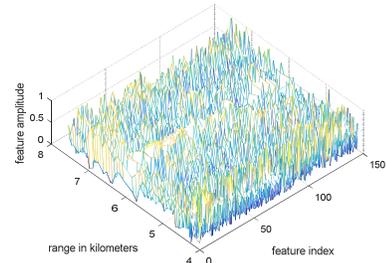


Fig. 3.2: Feature space (normalized to unit range) with respect to range (in km)

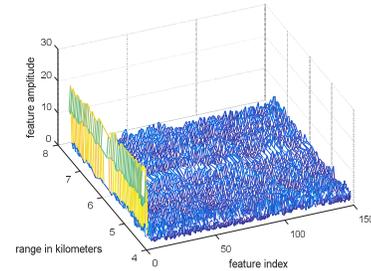


Fig. 3.3: Feature space (normalized to unit variance) with respect to range (in km)

Multilayer perceptrons with one and two hidden layers were tested in order to perform a regression from the acquired spectral and cepstral features (inputs) to the ground-truthed range (labels). With robustly engineered features a multilayer perceptron with a single hidden layer and only 15 neurons was found sufficient to achieve satisfactory results, as described in table 2 and figure 4.

Feature normalization scheme	Linear range scaling	Normalization to unit range	Normalization to unit variance
Linear range			
Mean Error of range estimate in Training set (meters)	293	293	298
Mean Error of range estimate in validation set (meters)	270	363	312
Mean Error of range estimate in Test set (meters)	587	440	341

Table 2 : range prediction on various datasets for different normalizations

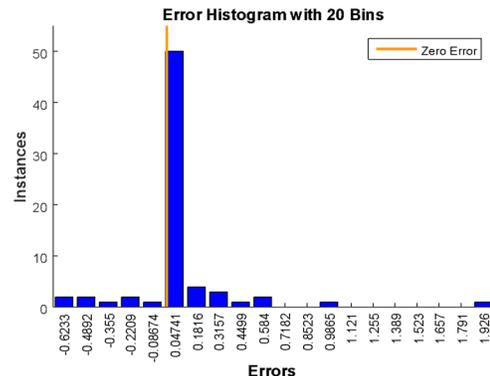


Fig. 4: histogram of range error (in km)

5. CONCLUSIONS AND PERSPECTIVES

The proposed algorithm permits, for each received vocalization, to estimate range with a multi-layer perceptron that produces results in real-time and within a low error margin. The algorithm can be understood as a dynamic calibration system that can be trained on site before a survey so as to learn the environmental propagation characteristics at frequencies of interest by playing back predetermined sound types (humpback whale sounds in this experiment).

Despite the relatively small number of samples (approx. 400), the good convergence of training indicated clear structure in the chosen features which allowed consistent range estimation. Error in the range estimates were mostly encountered below 400 meters using normalization to unit variance and had rare outliers above 1km.

After this early experiment additional deployments have been made which greatly increase the amount of samples available and will be the basis for more extended publications.

The effects of directionality of the calls, not fundamental in this particular experiment, should be investigated. Newly acquired data with playback sounds including real humpback whale sounds as well as sperm whale impulsive clicks, artificial chirps, and artificial broadband white noise, will also allow the assessment of the capacity of generalization of this algorithm to larger groups of tonal and impulsive sounds. . Additional effort is being made to define new features and robustify them so as to further reduce the mean error and variance of range estimates.

ACKNOWLEDGMENTS

This work has been supported by the Spanish government (contract TEC2015-69266-P) as well as by the European Regional Development Fund (ERDF/ FEDER).

BIBLIOGRAPHY

[1] Zimmer, Walter MX. "Range estimation of cetaceans with compact volumetric arrays." *The Journal of the Acoustical Society of America* 134.3 (2013): 2610-2618.

[2] Duan, Rui, et al. "Moving source localization with a single hydrophone using multipath time delays in the deep ocean." *The Journal of the Acoustical Society of America* 136.2 (2014): EL159-EL165.

[3] Zimmer, Walter MX. *Passive acoustic monitoring of cetaceans*. Cambridge University Press, 2011.

[4] Ainslie, Michael A., and James G. McColm. "A simplified formula for viscous and chemical absorption in sea water." *The Journal of the Acoustical Society of America* 103.3 (1998): 1671-1672.

[5] Fisher, F. H., and V. P. Simmons. "Sound absorption in sea water." *The Journal of the Acoustical Society of America* 62.3 (1977): 558-564.

[6] Francois, R. E., and G. R. Garrison. "Sound absorption based on ocean measurements: Part I: Pure water and magnesium sulfate

contributions." *The Journal of the Acoustical Society of America* 72.3 (1982): 896-907.

[7] Francois, R. E., and G. R. Garrison. "Sound absorption based on ocean measurements. Part II: Boric acid contribution and equation for total absorption." *The Journal of the Acoustical Society of America* 72.6 (1982): 1879-1890.

[8] Bonnel, Julien, et al. "Range estimation of bowhead whale (*Balaena mysticetus*) calls in the Arctic using a single hydrophone a." *The Journal of the Acoustical Society of America* 136.1 (2014): 145-155.

[9] Tiemann, Christopher O., et al. "Three-dimensional localization of sperm whales using a single hydrophone a." *The Journal of the Acoustical Society of America* 120.4 (2006): 2355-2365.

[10] Thode, Aaron M. "Source ranging with minimal environmental information using a virtual receiver and waveguide invariant theory." *The Journal of the Acoustical Society of America* 108.4 (2000): 1582-1594.

[11] Houegnigan, Ludwig, et al. "Neural networks for the localization of biological and anthropogenic source at neutrino deep sea telescope." *OCEANS 2015-Genova*.

[12] Houegnigan, Ludwig, et al. "High performance supervised time-delay estimation using neural networks" *IEEE International Conference on Acoustics, Speech and Signal Processing. Proceedings*. 2017.

[13] Nosal, Eva-Marie, and L. Neil Frazer. "Track of a sperm whale from delays between direct and surface-reflected clicks." *Applied Acoustics* 67.11 (2006): 1187-1201.

[14] Shaltaf, Samir. "Neural-network-based time-delay estimation." *Eurasip Journal on Applied Signal Processing* 2004 (2004): 378-385.

[15] Au, Whitlow WL, et al. "Transmission beam pattern and echolocation signals of a harbor porpoise (*Phocoena phocoena*)." *The Journal of the Acoustical Society of America* 106.6 (1999): 3699-3705.

[16] Au, Whitlow WL, et al. "Transmission beam pattern of a false killer whale." *The Journal of the Acoustical Society of America* 93.4 (1993): 2358-2359.

[17] Shaffer, Jessica Ward, et al. "Effective beam pattern of the Blainville's beaked whale (*Mesoplodon densirostris*) and implications for passive acoustic monitoring." *The Journal of the Acoustical Society of America* 133.3 (2013): 1770-1784.

[18] Cerchio, Salvatore, and Dahlheim, Marylin, "Variation in feeding vocalizations of humpback whales *Megaptera novaeangliae* from southeast Alaska." *Bioacoustics* 11.4 (2001): 277-295.

[19] Schmidhuber, Jürgen. "Deep learning in neural networks: An overview." *Neural networks* 61 (2015): 85-117.