Improving Fish School Detection Using Phase Information in Continuous-Wave Underwater Acoustic Signals: A Study Based on Cylindrical Sonar Data

Yohann Gourret[1], Karl Thomas Hjelmervik[2], Tommi Brander[3], and Hector Pena[4]

[1]University of South-Eastern Norway, yohann.gourret@usn.no
[2]University of South-Eastern Norway, karl.t.hjelmervik@usn.no
[3]University of South-Eastern Norway, tommi.brander@usn.no
[4]Institute of Marine Research, hector.pena@hi.no

Abstract: *This paper presents a novel approach to improving fish school detection using underwater acoustics, with a particular focus on the integration of both amplitude and phase information obtained from a cylindrical sonar survey employing Continuous-Wave (CW) pulses. While the inclusion of phase data does not significantly enhance overall classification accuracy, it proves valuable for identifying and discarding potential back lobes. The proposed method leverages Convolutional Neural Networks (CNNs) to analyze the collected sonar data and distinguish between fish schools and non-fish signals. Experimental results demonstrate that incorporating phase information helps reduce false detections associated with back lobes, leading to cleaner classification outputs. Receiver Operating Characteristic (ROC) curves show comparable performance with and without phase data in general, but highlight improved robustness in challenging scenarios where back lobes are present. This study emphasizes the role of phase information in refining underwater acoustic analysis for more reliable fish school detection, particularly in CW-based applications where back lobe artifacts are common.*

Keywords: *Fishery acoustics, detection, Machine Learning.*

1. INTRODUCTION

Accurate identification of fish schools using acoustic methods is crucial for sustainable fisheries management. Traditional methods for fish monitoring rely on labour-intensive and error-prone manual observation and processing. CNNs can automate and enhance the accuracy of fish recognition [1, 2].

Side-looking sonar systems are effective in detecting fish. Nicholas Makris [3] demonstrates large low-frequency systems for detection of entire fish schools over continental-shelf scales. This technology can image vast areas, providing valuable insights into fish population dynamics and behavior. On the other hand, high-frequency sidescan sonar systems detecting individual fish at shorter ranges [4] offer precise detection capabilities and an improved basis for classification of species, but are limited in their range. Sonar systems operating in the low ultrasonic range present a middle ground, offering detection ranges of hundred of meters, and are sufficiently compact for use on smaller vessels, such as unmanned surface vehicles. They balance the need for range and resolution, providing a versatile solution for fisheries management.

We present an approach to fish school recognition using a CNN applied to matched filtered output from a side-looking fishery sonar in the low ultrasonic frequency range. Our method involves conventional detection algorithms on beamformed and matched filtered data, with a time-bearing window of the matched filter output fed into a CNN for binary classification of fish (*Fish School* or *Not Fish School*). We include both amplitude and phase data in the matched filter output, as the phase data helps in recognizing false positives caused by back lobes, a phenomenon resulting from an undersampled array.

The method is demonstrated on 4,760 pings collected by the Institute of marine research with a Simrad SU90 fish finding sonar. The transmission is a 6 ms long continuous wave pulse at 26 kHz. The data used are complex-valued signals sampled at 4 kHz in 64 beams.

2. BACK LOBES

In our data, we can observe back lobes (equivalent to side lobes). Those lobes are around 110° from the main lobe. Looking at the phase, we can observe very interesting patterns, chess patterns (Fig. 1), that make those back lobes easy to identify. We will first present how we visualize the phase and then explain the origin of the patterns.

Visualization of the phase

The original signal has a frequency $f_c = 26$ kHz. We have a complex signal sampled at $f_s = 4$ kHz. We choose to represent the phase information of the complex pressure $p_{i,j}$ at range i and direction j by $|\cos(\arg(p_{i,j}))|$, see Fig. 1.

This representation only works if $\exists n \in \mathbb{N}$ such that $f_c/f_s = \frac{1}{2}n$. In our case $f_c/f_s = 6.5$. To have the same visualization for other frequencies, we can represent the phase information by $|\cos(\arg(p_{i,j}e^{j2\pi f_c t_i}))|$, where t_i is the time at the sample $p_{i,j}$.

Chess pattern

For human eyes, the patterns in the phase information are very easy to identify. Could phase information be beneficial for a neural network in identifying back lobes or, generally, improving classification performance?

Chess patterns in the phase information are due to a variation of the phase measured on the angular axis and on the radial axis. For the angular axis, this originates from the beamformer.

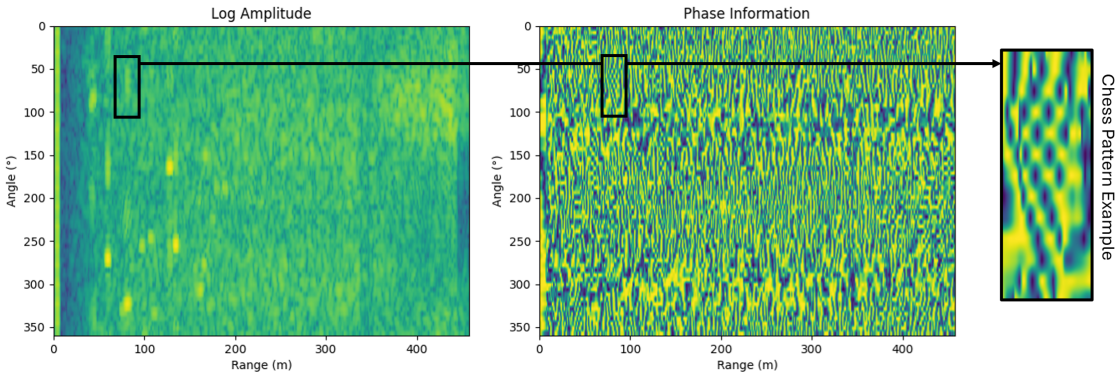

Figure 1: Log Amplitude, Phase Information and Chess Pattern example

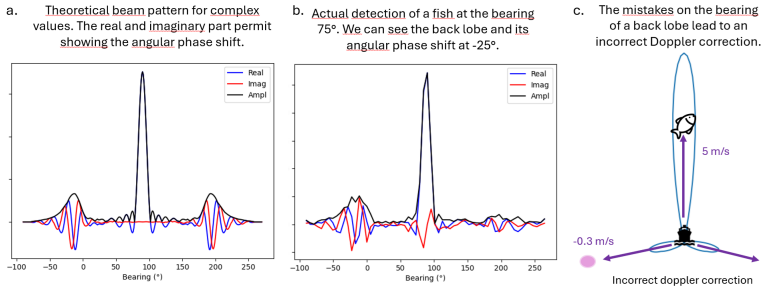

Figure 2: Explanation of the chess pattern effect

The exact beamforming method used by Simrad (Kongsberg Discovery) is unknown, but our guess for the beam pattern is shown in Fig. 2.a. Our analysis shows that the tendency remains similar for different variations of the beamforming method. The figure shows the angular variation of the amplitude and of the phase. We can see that on the back lobes, at around 110°, the phase changes rapidly over angle. The beamforming explains the angular phase shift. Fig. 2.b shows an actual fish school detection and the back lobe. We can see the same type of variation as in the theoretical beam pattern.

For the radial axis, the pattern appears due to the Doppler-corrected matched filtering. As shown in Fig. 2.c, on back lobes, the echo from the fish will be compensated for the wrong Doppler and will shift the phase over time / ”range”. However, for some bearings, for example at 110°/2=55°, the back lobes will be compensated for the correct Doppler by symmetry. Therefore, the pattern over the range is less consistent than the angular pattern, as we observe in Fig. 5 for several detections.

3. METHOD

We use two steps to identify fish schools. The first is to identify a potential fish school by a classical method, and the second is to use a CNN on the potential fish schools to classify them between two classes: *Fish School* or *Not Fish School*.

First step: Classical detection

To detect potential schools of fish, we use a range-dependent threshold. Detection occurs if $p_{i,j} > (G_\sigma * \hat{p})[i] \cdot 10^{T_{\text{dB}}/20}$ with $\hat{p}_i = \frac{1}{64}\sum_{j=1}^{64} p_{i,j}$, G_σ a Gaussian filter with $\sigma = 20$ m and the threshold $T_{\text{dB}} = 12$ dB. There is also a minimum and maximum detection surface and a

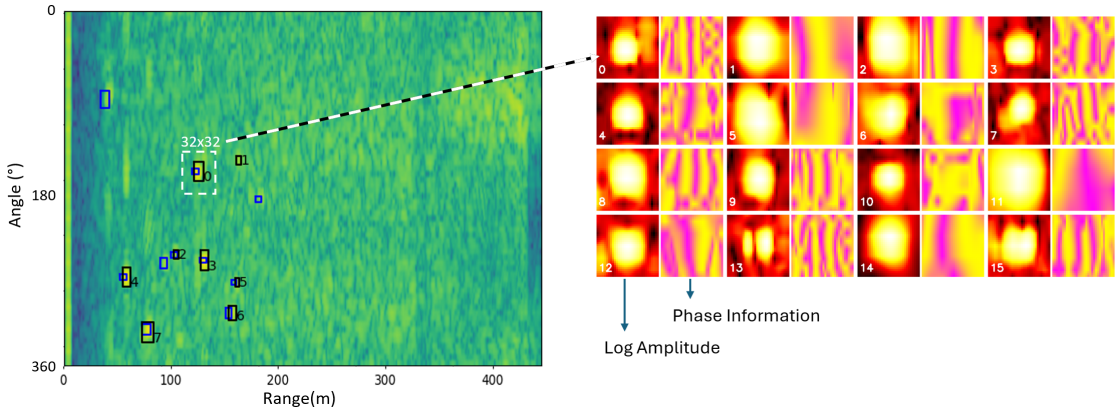

Figure 3: Image extraction from the detections

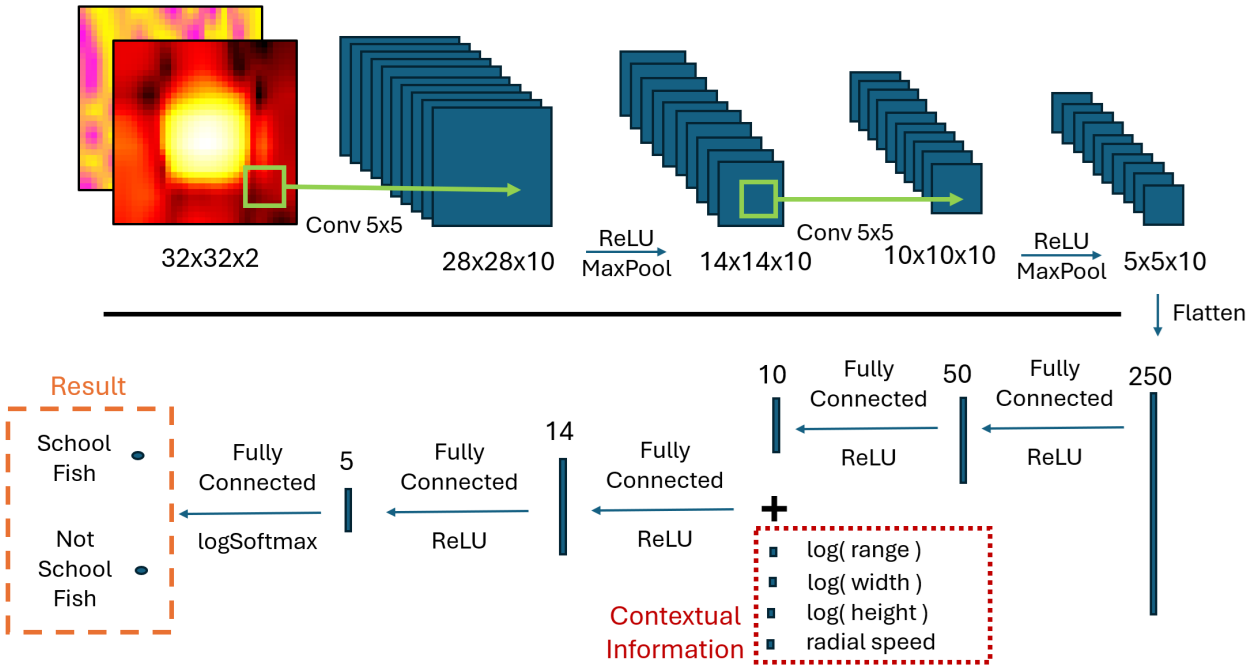

Figure 4: CNN architecture

minimum and maximum range. For each detection box, we generate a 2 channel 32×32 image (Fig. 3), which covers the detection box, is centered on it and has double the side lengths. The first channel of the image is the logarithmic amplitude of the pressure, and the second is the phase information. The resulting image is sent to the CNN for classification.

Second step: CNN

The CNN is presented in Fig. 4. The image of the detection step and four pieces of contextual information are the inputs. Contextual information is introduced later on in the network and contextualizes the image with the logarithm of the range of the detection, the logarithm of the width and the height of the original crop and the radial speed of the boat in relation to the environment. Here, radial speed = boat speed $\times$ cos(bearing).

The dataset is split into a training ($2/3$ of the data) and a testing dataset ($1/3$ of the data).

4. RESULTS AND DISCUSSION

The results of the training for models with and without phase information are shown in Fig. 7. Both models have the same architecture, but for the model without phase information, the phase information is set to 0 at input. The resulting Area Under the Curve on the testing dataset is close to 0.87 for both models. Those results can permit an automatization of the process. However, the phase information does not significantly increase the overall performance.

To identify the relevance of the phase information for the identification of back lobes, we manually identify them to analyze the response of models for those specific cases. This manual work was done for 321 pings at the beginning of the dataset, so they are in the training dataset. This resulted in 207 back lobes, for 3160 detections (6.6%). Some examples of back lobe detections are presented Fig. 5.

Comparison of models by classifying the back lobes between *Fish School* or *Not Fish School* implies choosing a specific threshold. To avoid choosing an arbitrary value, we use the ROC curve of the testing dataset to estimate the confidence of models. Based on the limit threshold needed for the response of the model to be switched between *Fish School* and *Not Fish School*, we can use the ROC curve to estimate the confidence. When this point on the ROC curve is on the left, the Neural Network is confident to classify it as *Fish School*. When it is on the right, the Neural Network is confident to classify it as *Not Fish School*. A visual example of this estimate of the confidence is shown Fig. 6.

Based on this metric, we use the ROC curves of the testing dataset to estimate the confidence of both models, with and without phase information, on the classification of back lobes. The resulting histograms are shown in Fig. 8. Histograms present the confidence of models to classify back lobes as Fish School, but because the back lobes are not fish schools, the closer the confidence is to 0 the better the result is. The average confidence for the model with phase information is 26.4%, which is lower than the 34.1% of the model without the phase information.

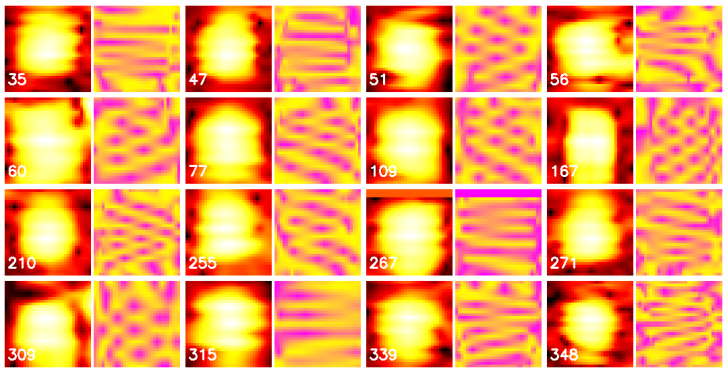

Figure 5: Samples of back lobes detection

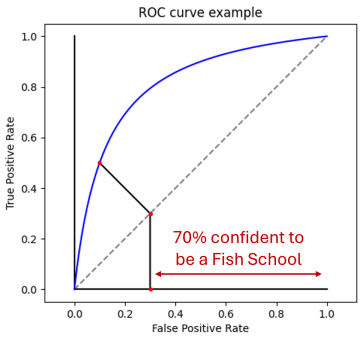

Figure 6: Estimation of the confidence

5. CONCLUSION

The current Area Under of the Curve of 0.87 can permit the automatization of the process, replacing human labor by machine labor. The current algorithm analyzes data ping by ping. Merging information over pings should lead to better results in further works.

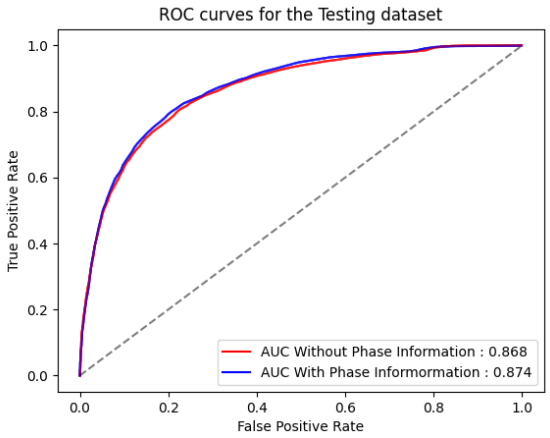

Figure 7: ROC curves for the Testing dataset

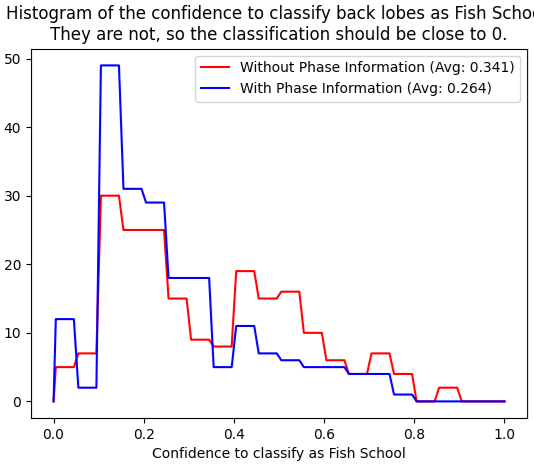

Figure 8: Histogram of confidence

About the phase information, it does significantly increase the performance for back lobes, but it does not significantly increase the performance of the neural network for other detections. Since most back lobes have been filtered away by the threshold of the classical method, the phase information does not increase the overall performance.

ACKNOWLEDGMENTS

This work was supported by the European Union Horizon 2020 research and innovation programme under the SAFARI project [grant number 101147432]. We would like to thank the Institute of Marine Research (IMR) for collecting the data used in this study, and Kongsberg Discovery for providing detailed information on their SU90 sonar system.

REFERENCES

[1] V. Kandimalla, M. Richard, F. Smith, J. Quirion, L. Torgo, and C. Whidden, "Automated detection, classification and counting of fish in fish passages with deep learning," *Frontiers in Marine Science*, vol. 8, Jan. 2022.

[2] W. Shen, M. Liu, Q. Lu, Z. Yin, and J. Zhang, "A fish target identification and counting method based on DIDSON sonar and YOLOv5 model," *Fishes*, vol. 9, p. 346, Sept. 2024.

[3] N. C. Makris, P. Ratilal, D. T. Symonds, S. Jagannathan, S. Lee, and R. W. Nero, "Fish population and behavior revealed by instantaneous continental shelf-scale imaging," *Science*, vol. 311, pp. 660–663, Feb. 2006.

[4] J. L. Ridgway, J. A. Madsen, J. R. Fischer, R. D. Calfee, M. R. Acre, and D. C. Kazyak, "Side-scan sonar as a tool for measuring fish populations: Current state of the science and future directions," *Fisheries*, vol. 49, pp. 454–462, Oct. 2024.