PROCEEDINGS OF SCIENCE

Identification of Ultra High Energy neutrino induced acoustic pulses

Christos Markou,[a,*] **Konstantinos Paschos,**[a,b] **Dimitris Stavropoulos,**[a,c] **Vasileios Tsourapis**[a,c] **and Georgios Zarpapis**[a,c]

[a] *NCSR Demokritos, Institute of Nuclear and Particle Physics, Ag. Paraskevi Attikis, Athens, 15310 Greece*
[b] *National Technical University of Athens, School of Mechanical Engineering, Zografou Campus, 9, Iroon Polytechniou str, 15780 Zografou, Athens, Greece*
[c] *National Technical University of Athens, School of Applied Mathematical and Physical Sciences, Zografou Campus, 9, Iroon Polytechniou str, 15780 Zografou, Athens, Greece*

E-mail: cmarkou@inp.demokritos.gr

According to the thermo-acoustic model introduced by Askaryan, the instantaneous energy deposition of a particle's shower in a medium produces a sound wave. The concept of acoustic neutrino detection is based on this principle. As the extremely low flux of Ultra-High Energy cosmic neutrinos is a limiting factor for the study of related astrophysical phenomena, large detector configurations have to be built. As sound has a much larger attenuation length than light in water, the development of hydrophone array-based detectors is a feasible solution. The ability to identify Ultra-High Energy neutrino induced acoustic pulses in underwater sound is discussed in this contribution. Acoustic pulses from Ultra-High Energy hadronic showers in the sea water simulated in the EeV regime are added to sea-state 0 noise recordings collected at a water depth of 1600 meters, Southwest of Peloponnese, Greece. Sophisticated signal processing and artificial intelligence techniques are developed and used for signal identification.

38th International Cosmic Ray Conference (ICRC2023)
26 July - 3 August, 2023
Nagoya, Japan

*Speaker

PoS(ICRC2023)1098

1. Introduction

The thermo-acoustic model introduced by Askaryan, describes the production mechanism of an acoustic pulse as a result of the instantaneous energy deposition from a particle shower [1]. The concept of acoustic neutrino detection is based on this principle.

To account for the extremely low flux of neutrinos in the Ultra-High Energy regime, detectors with a large instrumented volume have to be built. Concerning the Cherenkov detectors, the granularity is constrained by the absorption length of the light in seawater or ice. This results in detector configurations with optical modules arranged at a distance of few tens of meters. As sound has a much larger attenuation length than light in seawater, hydrophone-array configurations may be a feasible solution when very large instrumented volume detectors are needed. In such cases, where detectors would consist of hundreds or thousands of hydrophones, the enormous size of underwater acoustic recordings would lead to computational issues. A method to overcome this difficulty is to set a trigger condition according to the acoustic neutrino pulse properties at the hydrophone level.

The ability to identify neutrino induced acoustic pulses in underwater sound hydrophone recordings is discussed in this work. Simulated Ultra-High Energy neutrino induced acoustic pulses are added to underwater sound recordings and an ensemble of analysis techniques are used for their identification. A description of the data and MC simulation used in this work is presented in chapter 2. The data handling as well as the processing through the different frequency analysis steps are mentioned in chapter 3. In chapter 4, the classification algorithm used for the pulse identification is described along with the features used. Finally, a conclusion of this work is found in chapter 5, along with some ideas for the future steps.

2. Data and Monte Carlo simulation

2.1 Data recordings

The data used in this work have been collected at a sea depth of approximately 1600 meters, Southwest of Peloponnese, Greece. Two *Loggerhead DSG-ST* hydrophones [2] were recording continuously the ambient sea noise for about one month, operating at a sampling frequency of 144 kHz. The data collected in a time period of 2 hours correspond on average to ~2.1 GB per hydrophone. A common technique in acoustic data analyses in order to handle this amount of data conveniently, is to separate them into much smaller partitions, hereafter mentioned as *chunks*. The duration of the chunks was chosen to be 5 ms; taking into acount the sampling rate, each chunk corresponds to 720 discrete values with a ~7 μs period each.

The data recordings were calibrated according to the data provided by the manufacturer with respect to the hydrophove sensitivity and the electronics gain, having removed the contribution of the baseline DC offset. As a result, the 720 discrete digitized values that correspond to each chunk are converted into pressure (Pa).

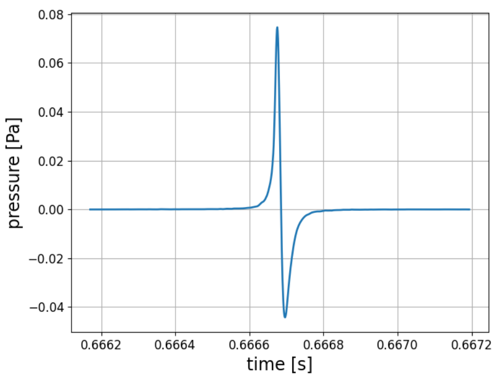
2.2 Simulation of neutrino acoustic pulses

The simulation chain of the acoustic neutrino pulses consists of two parts. The first one concerns the simulation of the particle shower produced by the neutrino interaction, while the second one concerns the calculation of the pulse waveform at a given hydrophone position, by integrating the shower energy deposition and accounting for the attenuation during the propagation of the pulse in the seawater.

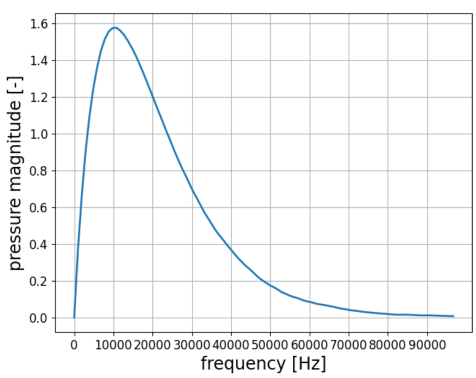
For the shower generation, a special version of the CORSIKA Extensive Air Shower simulation software has been used [3]. This version, the CORSIKA-IW[4], is modified to simulate showers in ice or in seawater. A custom modification has been also implemented and concerns the distribution of the deposited shower energy provided by CORSIKA; it has been modified to provide a two-dimentional histogram, i.e. the deposited shower energy as a function of the shower's axial and radial distance, instead of the standard axial deposited energy distribution provided by CORSIKA. This is essential as a numerical integration is done according to the profile of the shower's deposited energy for the pulse calculation.

The method used for the calculation of the acoustic neutrino pulses is the one developed by the ACoRNE collaboration as described in [5] and [6]. It is a Monte Carlo numerical solution of the wave equation extracted by Askaryan and subsequently the calculation of the pulse waveform accounting for the attenuation. From the technical perspective, this software was re-written from MATLAB to Python with the aim to take advantage of this language's support and compatibility, taking into account its increasing importance in the research community.

A simulation chain was developed using the tools described above. Hadronic showers were generated in the energy range between 10 EeV and 100 EeV, according to an E^{-2} energy spectrum. Dozens of pulses have been calculated for each generated hadronic shower, with the shower to hydrophone distances varying between 50 and 250 meters. Indicatevely, the shape of an acoustic neutrino pulse simulated with the tools described above is presented in Fig. 1a, along with its spectrum at the frequency domain using fourier transform in Fig. 1b. The characteristic narrow shape of the bipolar pulse leads to a broad frequency spectrum.



(a)



(b)

Figure 1: Acoustic neutrino pulse from a 100 EeV shower at a distance of 1000 meters (*left*), and its spectrum at the frequency domain (*right*).

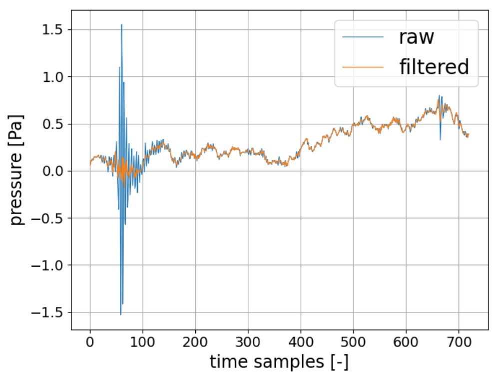
3. Data processing - Frequency analysis techniques

The investigation of the ability to identify acoustic neutrino pulses in underwater data recordings is reduced to a classification problem. The elements of the considered signal and background classes are the chunks, as introduced in Chapter 2. Data chunks with pure underwater noise recordings constitutes the background class. For the signal class, simulated neutrino acoustic pulse waveforms are added in random positions on top of the recorded pressure time-series. The chunks used for both classes have been selected in an alternate way to avoid potential background bias. In more detail, for a given chunk i that is used for the backgound class, the chunks $i+2$, $i+4$, $i+6$ etc. are used for background, while the chunks $i+1$, $i+3$, $i+5$ etc. are used for signal. The processing of the chunks for signal and background classes through frequency analysis techniques is described in the following.

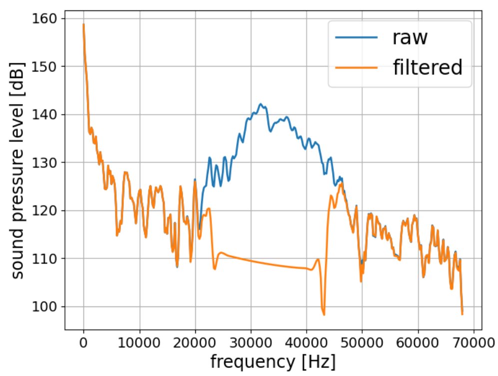
3.1 Subtraction of dolphin clicks

Cetaceans and other underwater mammals produce various sounds, with the so-called *dolphin clicks* to be among them. The waveform as well as the frequency spectrum at which dolphin clicks occure is known [7]. The presence of dolphin clicks in the data recordings was noticed during the analysis; An example is shown in Fig. 2a (*blue*). A criterion was developed in order to spot the chunks containing dolphin clicks, as the subtraction of these chunks is necessary for further analysis.

For the identification of the chunks that contain dolphin clicks, a *Butterworth/bandstop* filter is applied. This filter is applied to each chunk and removes the contribution of the frequencies between 22 kHz and 45 kHz from the waveform. In Fig. 2b, the subtraction of the frequencies is presented using a Sound Pressure Level diagram. The application of this filter on the waveform is illustrated again in Fig. 2a (*orange*). An efficient subtraction of the dolphin click related frequencies is noticed, while the rest of the waveform barely changes. A metric is finally constructed in order to characrerize chunks that contain dolphin clicks, taking into account the difference between the original waveform and the one with the bandstop filter applied. The chunks which are marked to contain dolphin clicks are excluded from further analysis. In a future hydrophone-array acoustic neutrino detector, dolphin clicks could also be rejected by modifying accordingly the higher-level trigger conditions (i.e. correlations between hydrophones).



(a)



(b)

Figure 2: *Left*: Chunk with a dolphin click (blue) and the same chunk after the application of the bandstop filter (orange); *Right*: Sound Pressure Level for the same waveforms.

3.2 Removing the contribution of lower frequencies

An additional filter is applied during the data processing, in order to remove the contribution of the low frequency sources, such as ship engines, seismic activity, etc. The frequencies below 5 kHz are removed with the application of a *Butterworth/High-Pass* filter. In Fig. 3, the waveform of a chunk is shown, before and after the application of the high-pass filter. A neutrino acoustic pulse has been added to this chunk (slightly before 600). With the application of the high-pass filter, an excess is noticed at the position where the neutrino pulse has been added, with respect to the filtered baseline. The impact of this filter has been proven effective as it increases the overall efficiency of the classification, thus this technique has been followed in the processing.

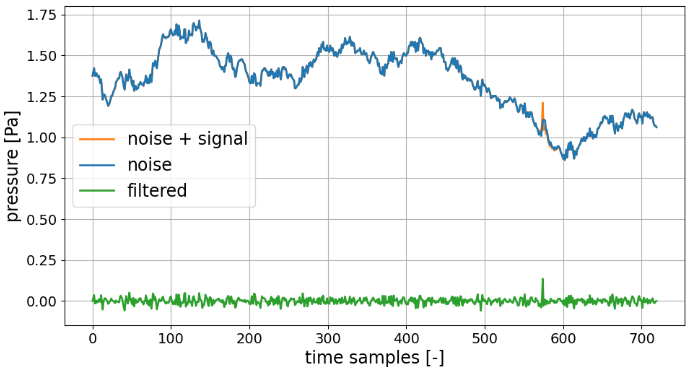
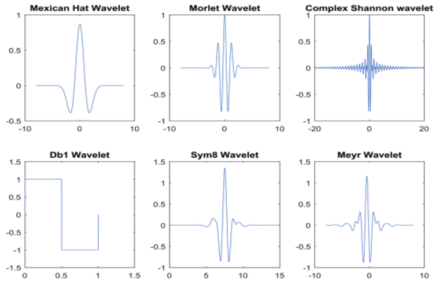


Figure 3: The waveform of a chunk in units of pressure (blue), with the acoustic neutrino pulse added on top (orange), and the latter after the high-pass filtering.

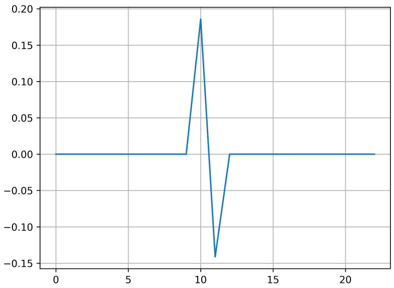
3.3 Application of the wavelet transform

The final step concerns the processing of the chunks through a wavelet transform [8]. The wavelet transform is efficient when deconvoluting a signal which is extremely localized in the time domain. In Fig. 1, it is noticed that the narrow shape of the neutrino pulse at the time domain leads to a broad frequency domain spectrum. This indicates that methods based on the Fourrier transform may not be efficient, as has been also shown in [8]. In contrast to the Fourier Transform which expresses a signal as a linear combination of periodic functions (sine and cosine), the wavelet transform expresses a signal in a basis generated by a mother wavelet. Commonly used mother wavelets are shown in Fig. 4a. The elements of the basis are stretched or shrunk versions of the mother wavelet in the time domain. The mother wavelet that was chosen in order to deconvolute the acoustic neutrino pulse shape from the background is the one shown in Fig. 4b. The basis used for the deconvolution consists of eight wavelets generated by this mother wavelet.

The usage of the wavelet transform in this analysis is illustrated in Fig. 5.At the *top-left* plot, the waveform for a chunk is presented without (blue) and with (green) the high-pass filtering. In addition, the same is shown at the *top-left* plot with a neutrino acoustic pulse added on top slightly before chunk 600 (orange). The bottom plots concern the values of the wavelet transform coefficients, each for the above plot respectively. The first and last four samples have been removed from the chunk after the wavelet transform to reject the strong correlation values because of the discontinuity.



(a)



(b)

Figure 4: *Left*: Common functions that are used as mother wavelets. *Right*: The chosen mother wavelet in this work.

Strong correlation coefficient values are noticed at the position where the acoustic neutrino pulse was added. The final step of the analysis benefits from this, as variables are constructed from the wavelet correlation matrix, with the aim to be used as features in a classification algorithm.

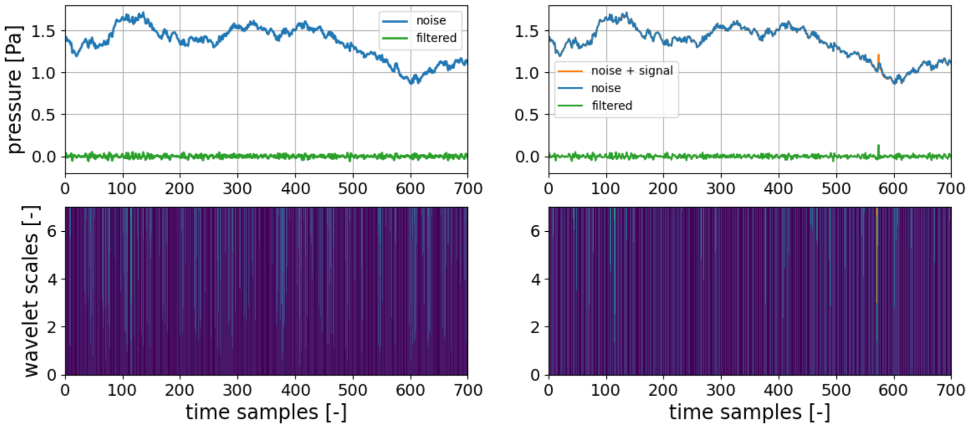


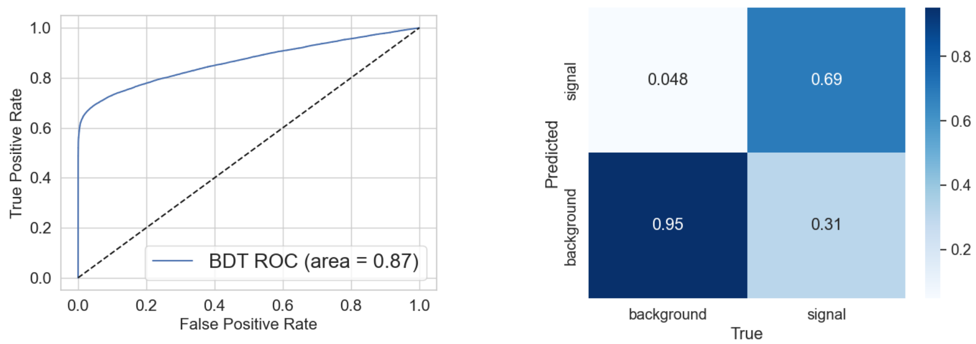
Figure 5: *Left*: The waveform of a chunk before and after the filtering (*top*) along with the wavelet transform coefficients (*bottom*). *Right*: The same as in the left plots with a neutrino acoustic pulse added in the chunk waveform.

4. Classification

4.1 Gradient Boosted Decision Tree

A Boosted Desicion Tree (BDT) with Gradient Boosting has been used for classification using the *scikit-learn* software [9]. Eighteen variables are used as features, all extracted from the wavelet correlation coefficient matrix. The construction of these variables involve operations between the rows of the matrix, ratios between maximum values and the baseline in the rows of the matrix, as well as values of different order matrix norms.

The phase space of the BDT parameters has been also scanned to define optimal values for the number of trees, the boosting parameter and the maximum tree depth. The chunks that have been processed as described in Chapter 3 have been used to train and evaluate the BDT classifier. The resulting ROC curve is shown in Fig. 6a. A signal efficiency of ~ 69% can be achieved with a background rejection of ~ 95%, as shown in Fig. 6b.



(a) **(b)**

Figure 6: *Left*: ROC curve for the evaluation of the BDT classifier. *Right*: Confusion matrix for the same classifier.

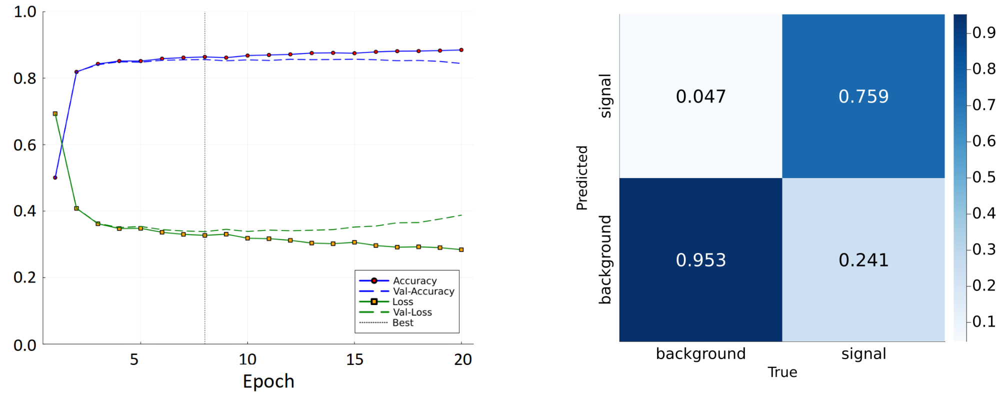
4.2 Convolutional Neural Network

A different approach has been also tried for classification, taking advantage of the state-of-the-art Deep Learning technique of Convolutional Neural Network (CNN). For the implementation of the CNN algorithm in this analysis, the Flux package has been used [10], written in Julia programming language. As the execution of Deep Learning algorithms is computationally intense, high-performance GPUs are used to achieve a significant reduction of the runtime.

A one-dimentional CNN is used for classification. The signal and background classes are as defined earlier, but in the case of the CNN the chunks are only high-pass filtered and they have not processed through the wavelet transform. The input layer consist of 720 nodes, one for each sample of the chunk. The architecture of the model includes Convolutional and MaxPooling layers for pattern identification and dimensionality reduction, as well as Dropout layers, which improve the model's robustness by offering multiple paths to the output layer. The CNN is re-trained several times, the so-called epochs, and the one with the lowest validation loss value is chosen as the optimal as shown in Fig. 7a. The confusion matrix is shown in Fig. 7b, demonstrating a signal efficiency of ~ 76% with a background rejection of ~ 95%.

5. Conclusion

An investigation on the idectification of acoustic neutrino pulses at the hydrophone level has been made in this analysis. A simulation chain has been developed in order to generate Ultra High Energy hadronic showers in the seawater and moreover to simulate the thermo-acoustic model, in order to extract acoustic neutrino pulses. The simulated pulses have been added in



(a) **(b)**

Figure 7: *Left*: Accuracy and loss for the training and validation samples for each epoch. *Right*: Confusion matrix for the CNN with the least validation loss (*8th* epoch).

underwater data recordings, and the data have been processed through frequancy filters and a wavelet transformation. A Machine Learning classifier (BDT) has been developed with features based on the wavelet coefficient matrix; an efficiency of $\sim$ 69% is achieved with a background rejection $\sim$ 95%. The potential to identify acoustic neutrino pulses is explored also with a Deep Learning classifier (CNN), illustrating a signal efficiency of $\sim$ 76% with a background rejection of $\sim$ 95%. Both techniques show promising first results for neutrino acoustic pulse identification at the hydrophone level. With these as starting points, studies for a trigger condition that involve different hydrophones could be implemented. Investigations for such a trigger condition could lead to a feasible data acquisition system for a future acoustic neutrino detector.

Acknowledgements

Work carried out in the context of the program "Intelligent, Specialized Environmental Observatory in Messinia" (ISEO), funded by the Operational Program "Peloponnesos" 2014-2020, of NSRF (2014-2020).

References

[1] J.G. Learned, Phys. Rev. D19 (1979) 3293.
[2] https://www.loggerhead.com/dsgst
[3] D. Heck, J. Knapp, J. Capdevielle, G. Schatz, and T. Thouw, Report FZKA 6019.11 (1998).
[4] https://web.iap.kit.edu/corsika/download/old/v660/, access provided by developers.
[5] S. Bevan et al. Astroparticle Physics 28 (2007) 366–379
[6] S. Bevan et al. Nuclear Instruments and Methods in Physics Research A 607 (2009) 398–411
[7] 10.1016/j.spjpm.2016.08.004
[8] I. Daubechies, IEEE Transactions on Information Theory, vol. 36, no. 5, pp. 961-1005, (1990)
[9] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.
[10] Innes, (2018). Flux: Elegant machine learning with Julia. JOSS, 3(25), 602.