DOI: 10.61782/fa.2025.0897

FORUM ACUSTICUM EURONOISE 2025

TIME SYNCHRONIZATION OF LONG-TERM UNDERWATER ACOUSTIC OBSERVATIONS: THE CHALLENGE OF AUTONOMOUS HYDROPHONES WITHOUT GLOBAL CLOCK BASELINE

Damoon Nazarpour[1] **Pier Francesco Moretti**[2] **Silvano Buogo**[3]
Angela Pomaro[1,4*]
[1] Institute of Marine Sciences, National Research Council of Italy, Venice, Italy
[2] Department of Earth System Sciences and Environmental Technologies, National Research Council of Italy, Rome, Italy
[3] Institute of Marine Engineering, National Research Council of Italy, Rome, Italy
[4] Department of Computer, Control and Management Engineering A. Ruberti, Sapienza University of Rome, Italy

ABSTRACT

The proposed study involves the deployment of three hydrophones in an equilateral triangle configuration alongside an underwater sound source that serves as a reference point for post-hoc synchronization. By applying a band-pass filter centered on the frequency of the emitted sound, the time-frequency spectrum output is used to evaluate performances of three different algorithms. The algorithms use methods of image processing, zero-crossing detection, and convolution with a single square wave to detect the absolute time offset between a pair of hydrophone measurements. The pros and cons of the instrumental set-up and of the algorithms are shown, as well as different options for their use to retrieve the underwater 3D acoustic field. It is illustrated that using convolution with a single square wave delivers the most consistent and reliable results in synchronization.

Keywords: *underwater soundscape, synchronization protocol, three-dimensional acoustic field, AAOT.*

1. INTRODUCTION

The study of the underwater three-dimensional acoustic field enables ocean floor mapping, source localization, monitoring environmental impact of human activities, and identifying indicators for changes in the processes acting in the marine environment. The common practice of deploying an array of hydrophones faces many challenges, such as the uncertainty of the underwater sound profile and the capability to cross reference the measurements of different autonomous hydrophones. Typically, hydrophones are deployed in autonomous mode, in which each unit controls its own power supply, data sampling, and storage. This characteristic makes hydrophones susceptible to clock drifts, as there is no global clock to which they can refer, consequently making ad-hoc synchronization unfeasible [1]. Furthermore, many hydrophones are adhered to data transfer protocols that disrupt the sampling process to write data to storage, resulting in additional recording gaps. Many of the existing synchronization algorithms are based on Dynamic Time Wrapping (DTW) [2] method [3-5], which due to high computational complexity struggle against large and high-frequency audio files. [6] uses signal features like spectral flatness, zero- crossing rate, and signal energy in pairwise

[*] ***Corresponding author***: angela.pomaro@cnr.it.

manner and addresses the problem of large number of recordings rather than large high-frequency audio signals.
This project addresses the requirements for the definition of a hydrophones synchronization protocol and instrumental set-up. This activity aims at potentially connecting a number of distributed autonomous hydrophones, in order to increase ocean observation strategies.
To this end, a 5-minute emission of a Gaussian signal at 9 kHz was selected as a known transmit signal to be received by all hydrophones. Subsequently, the investigation focused on data-driven synchronization methods, which are event-based and rely on the 9 kHz signal as a reference. The measurements obtained from three hydrophones were evaluated and synchronization errors were compared.

EXPERIMENTAL SETUP

As part of the SEAmPhonia Project [7], three hydrophones were deployed in the Adriatic Sea for a series of different campaigns, at various times of the year, under different combinations of hydrophones in terms of devices, sampling frequency, and settings. The synchronization protocol was tested across two distinct campaigns carried out at the Acqua Alta Oceanographic Tower (AAOT), a fixed research infrastructure managed by the National Research Council of Italy in the northern Adriatic Sea, on 16m of depth, 8 nautical miles offshore the coast of the Venice lagoon (Global Positioning System (GPS) coordinates 45°18′51.288″N, 12°30′29.694″E).
For each campaign a set of three different hydrophones was deployed in a known configuration, as shown in Figure 1.
Acoustic signals are synthesized by an electrodynamic underwater sound source installed on the AAOT, whose position is also shown in Figure 1.

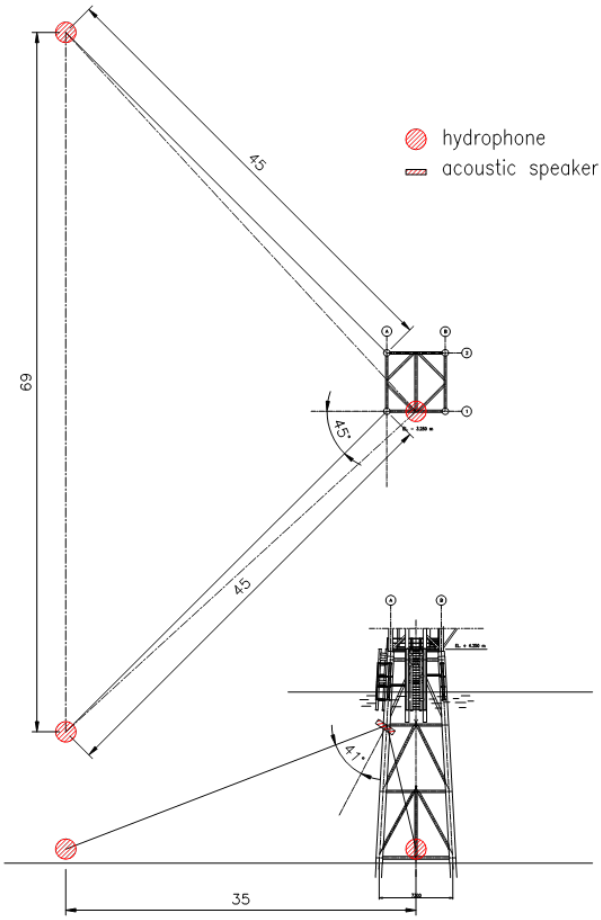

Figure 1. Deployment geometry

The cabled hydrophone, connected to the tower, is designated H0, while the autonomous hydrophones are labelled H1 and H2. The configuration of Figure 1 implies that the time difference between the arrival time of the emitted sound between any pair of hydrophones cannot be obtained by summation of the other pairs time differences.
The present study refers in particular to two campaigns carried out within the project, designated as EXP01 and EXP02. It is noteworthy that each measurement possesses unique characteristics, including the model of the hydrophone and the sampling frequency.

Table 1. First test settings: period 24/01/2024 – 07/02/2024

Hydrophone	# bits	Sampling rate (kHz)	Sampling duration

H0	16	96	1h
H1	16	96	1h
H2	16	96	1h

H0 is represented by a Multi-Purpose Digital Hydrophone (MupHydro) produced by Colmar S.r.l. based on a National Instruments NI sbRIO-9651 circuit, while H1 and H2 are both a Sono Vault Acoustic Recorder, produced by Develogic GmbH whose standard system is equipped with a Reson TC4037-3 hydrophone. Overall, each hydrophone saved 341, 349, and 214 files, respectively. During this campaign, different combinations of synthesized signals were generated by the sound source as a point of reference for synchronization. The tests highlighted the higher sensitivity of the cabled hydrophone H0, compared to H1 and H2. Multiple reference points have allowed segmentation and filtering for the synchronization protocol.

Table 2. Second test settings: period 03/06/2024 – 16/06/2024

Hydrophone	# bits	Sampling rate (kHz)	Sampling duration
H0	16	96	1h
H1	16	128	1h
H2	16	156.250	1h

For this experiment, H0 is the same Multi-Purpose Digital Hydrophone (MupHydro) used for the previous campaigns, while H1 is a RtSys Sylence LP-440-H-P-S underwater recorder equipped with a Colmar GP1516 hydrophone and H2 is another RtSys hydrophone.
As shown in Table 2, for this experiment, each hydrophone is characterized by a different sampling frequency and sensitivity. Overall, 529 mutual recording files have been retrieved between H0 and H1, and 150 mutual recordings have been retrieved between both H0 - H2 and H1 - H2. To synchronize the measurements, both H1 and H2 signals were down-sampled to 96 kHz by using Discrete Fourier Transform (DFT) [8].
The two experiments, hence datasets, have been used to test different synchronization algorithms. During EXP01 campaign, two acoustic signals centered at 2 kHz and 9 kHz respectively have been synthesized by the sound source, while EXP02 campaign offers a long and consecutive recording period with a single 9 kHz acoustic signal. In addition, as shown in Table 1 and Table 2 the sampling rate across all EXP01 hydrophones is 96 kHz, while in EXP02 the sampling frequency of H0, H1, and H2 are set at 96 kHz, 128 kHz, and 156.250 kHz, respectively. The synchronization approach takes into account the distance between hydrophones and the speed of underwater sound, which is assumed to be equal to a nominal value of 1500 m/s, although its precise value depends on salinity, temperature, and pressure. The precise value of the sound speed is not needed as this work focuses on the absolute time offset between acquired signals by the hydrophones.

2. SIGNAL SINCHRONIZATION

In order to achieve the desired synchronization accuracy, three distinct approaches have been studied and rigorously tested. The **first approach** entailed the use of spectrogram images, exclusively derived from computer vision algorithms. The **second approach** involved zero crossing of the second derivative of the signals' integral, with a filtering range extending approximately 9 kHz, and the alignment of their zeros. Finally, the **third approach** entailed the definition of a step array and the subsequent convolution of this array with the signal's 9 kHz band, with the objective of identifying the peak time acoustic signal in both coupled hydrophones. The aforementioned methods have been implemented and evaluated for both EXP01 and EXP02.
The latter approach yielded the most consistent and accurate results, as described in the following sections.

2.1 Image Processing

The underlying principle of this approach involves leveraging the semantic features of spectrogram images generated by SciPy [9][10] to facilitate their alignment. By utilizing rectangular kernels, a neighborhood can be defined over which morphological transformations can be performed. In particular, the difference between dilation and erosion of the image, or morphological gradient can be exploited to facilitate the delineation of object outline. The efficacy of these tools is demonstrated in Figure 2, which illustrates the output of this method, where it can be observed that low frequency noise is not interfered with the 9 kHz, contrary to what has happened to 2 kHz acoustic signal. This method is vulnerable to low-frequency noise, which hinders its automatic execution due to the necessity for bespoke image processing and hard coding. This requires performing a series of additional modifications to the spectrogram images, including adjustments to brightness, contrast, color channels, separations, and image thresholding, among others.

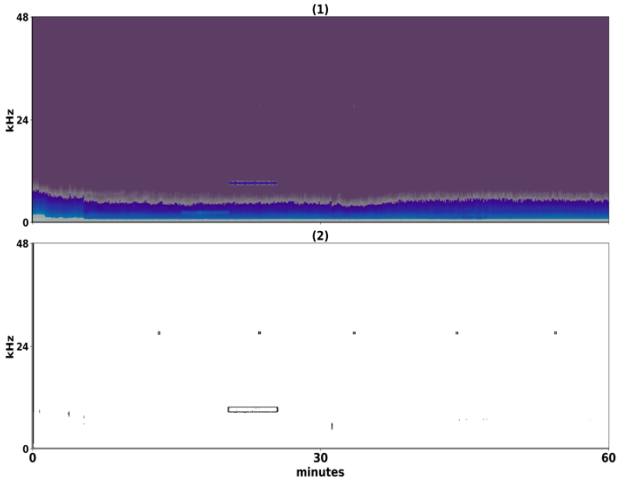

Figure 2. (1) Time-frequency spectrogram of H2 from EXP01 which was recorded on 29/01/2024 at 15:57:30.; (2) final output of morphological transformation.

Another problem is represented by the hollow edges that are delivered by morphological gradients. When images are split into smaller segments to increase the time resolution represented by each pixel, each single pixel displacement can return a big error in time. In a nutshell, this methodology introduces a novel paradigm in signal processing through computer vision; however, it currently lacks consistency and robustness, which can be enhanced through future works.

The process of aligning two images is based on image registration [11], which is achieved by using phase cross correlation [12] to compute relative shift (vertical and horizontal translation) between two images or arrays, being notably resilient to noise. See Figure 3 for the expected outcome of this method.

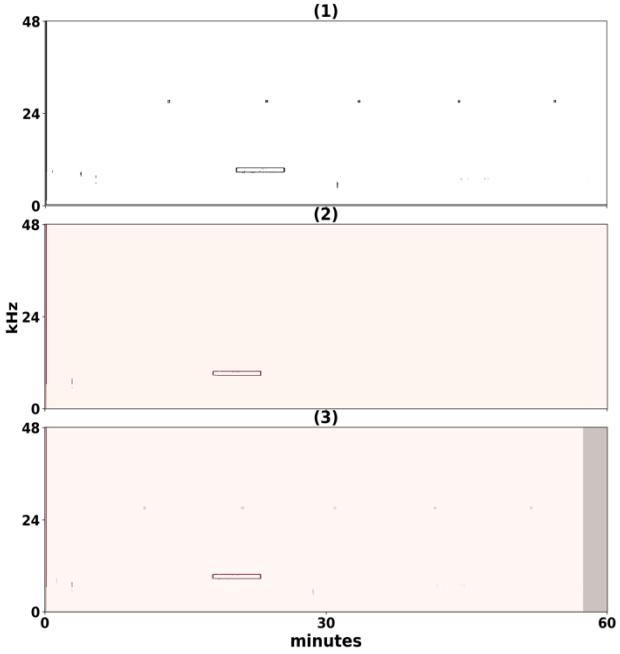

Figure 3. (1) Morphologically transformed image of H1 from EXP01, recorded on 29/01/2024 at 14:35:19; (2) Morphologically transformed image of H2 from EXP01, recorded on 29/01/2024 at 14:57:29; (3) Blended images of H1 and H2 while H1 is shifted for 155.97s towards the left, as it was detected by phase correlation between the two images.

Zero-crossing detection

Inspired by [13], this approach is analogous to the previous, but it refers to zero-crossing of the second derivative, to evaluate the time delay between two measurements. The method uses a band-pass filter with a central frequency of 9 kHz and a ±150 Hz cut-off to the spectrum, which delivers the root mean square (RMS) of the signal, after which the integral is then computed, returning an S-shaped curve from which the first and second derivatives are calculated consecutively, as shown in Figure 4. The difference between the zero-crossings of the second derivatives provides the time delay between the two measurements.

Given that the sound travels for approximately 30 milliseconds from the sound source to H1 and H2, the required time resolution in the synchronization process must be much less than 30 milliseconds. This inherent challenge is reflected in the smoothness of the curve, which decreases with decreasing time resolution. To address this issue, a

Savitzky-Golay [14] smoothing filter has been implemented, although this adversely affects the accuracy of the synchronization algorithm.

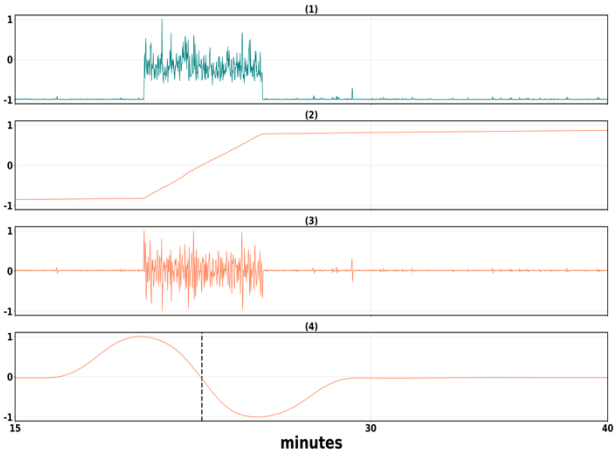

Figure 4. (1) RMS of H1 recording on 29/01/2024 at 14:55:19, after being passed through a band-passed filter with a central frequency of 9 kHz and a ±150 Hz cut-off; (2) Integration of the given RMS; (3) Second derivative of the integration without smoothing; (4) Second derivative of the integration with a Savitzky-Golay smoothing filter and its zero-crossing showed by the dotted black line.

Convolution with a step array

This last method relies on the possibility to convolve a step array of 5 minutes with a filtered signal, when the length of the acoustic signal is known in advance. This operation shall return a curve whose maximum occurs where the step array and the 9 kHz signals are fully aligned. This maximum peak can subsequently be assumed as the reference point for synchronization purposes.

Convolving two 16-bit integer arrays with size of the order 10^6 is time-consuming and can lead to read-only memory issues on common-use devices. To address these issues, a technique has been employed to extract from the measurements only the part containing the 9 kHz signal, while ignoring the rest.

By employing a large sliding window, a time-frequency power spectrum is generated that is characterized by low temporal resolution, while achieving high frequency resolution. A band-pass filter centered at 9 kHz and a ±150 Hz cut-off is set on the time series and subsequently collapsed into a 1D array by averaging. Afterwards, a 5 minutes single square array is convolved with it to approximately detect the region of interest in both hydrophone recordings. Finally, low and high time thresholds are defined such that both recordings contain the 9 kHz signals. The time-frequency domain is then re-examined, with a reduced window size, thereby achieving both high-time and low-frequency resolutions. The subsequent convolution step is executed after applying the aforementioned band-pass filter, truncating the signal between the time thresholds identified in the previous step, and normalizing each recording to the [0,1] range. This technique significantly accelerates the convolution calculation and reduces memory requirements. In order to further assess the practicality and reliability of the algorithm, prior knowledge of the 9 kHz signal was revised. Assuming that only an estimated signal length is known beforehand, a variation of ±5% to ±10% was introduced to the step array length, thereby adding or subtracting 15 to 30 seconds to the length of the step, achieving the results shown in Figure 5.

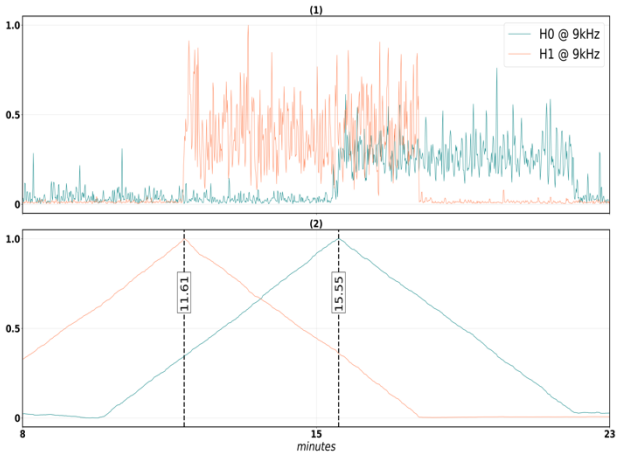

Figure 5. (1) RMS of H0 and H1 from EXP02 centered at 9 kHz which were recorded on 03/06/2024 at 23:57:05 and on 04/06/2024 at 00:00:03 respectively; (2) Results of convolutions that show an absolute time offset of 236.7 s between two signals.

RESULTS

The computer vision approach was tested on a number of files from EXP02 using all three hydrophones, with the process conducted individually on measurement pairs with the challenges outlined in the previous section. The results of this algorithm are presented in Table 3, resuming the

achieved synchronization time in seconds for a subset of analyzed pairs. Precision and accuracy issues arise from this method, where the algorithm returns the same values each iteration with high precision, but without accuracy. Referring to Table 3, it can be noted that while the absolute time difference between H2 and H1 equals almost 150 s across all listed recordings, the gray shaded row highlights an increased delay between H1 and H0, which contradicts the observed decreased delay between the H2 - H0 pair, which logically is not possible.

Table 3. Detected time offsets based on the image processing algorithm (EXP01)

H2-H0	H2-H1
651.525	150.142
659.203	150.129
666.983	150.090
671.267	150.013
-	150.041
685.817	150.047
564.983	150.053
467.529	150.085

Also, the results of the second derivative method were not sufficiently accurate or precise as they drastically change with the slightest change in either smoothness (Savitzky-Golay filter) or integration parameters. This behavior was suppressed by enlarging the Fourier windows size, which consequently results in a decreased resolution in time which makes this method no more useful for our objective to achieve a synchronization accuracy of at least 30 ms.
Finally, the last approach, based on 1D convolution, shows the most promising results. Figure 6 reports the final results, where each dot represents a pair of recordings and the vertical axis reports the detected absolute time delay in seconds.

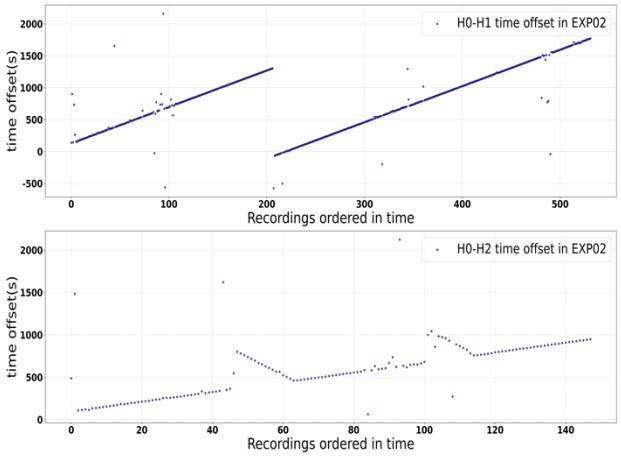

Figure 6. Absolute time offset between all recordings of EXP02 for H0-H1 and H0-H2 pairs. The horizontal axis shows the index of the recording ordered in time, while the vertical axis shows the time offset in seconds for each pair of related recordings.

This algorithm also allowed the detection of interesting patterns. For the case study reported in Figure 6, the time offset between H0 and H1 shows a linear increase in the measurement series (besides a hard reset occurred in the middle). Another pattern is evident in both pairs H0-H2 and H1-H2, where the clock of H2 suddenly jumps up (to increase offset) and then gradually resets back to the main trend line. The reason for such behavior traces back to the hydrophones' electronics or to some internal timeout thresholds.
Based on the observed tendency of H0-H1 time offsets, the objective is to derive a linear fit, which can be used to predict the time offset or to measure the drift between the two hydrophones, with outliers being eliminated in the process.
Due to the high uncertainty associated with these results and to the non-linear pattern observed for H0-H2 and H1-H2 pairs, the algorithm accuracy has been assessed in two ways. One consists in synchronizing two recordings based on two different signals and then evaluating their differences. Another method consists in applying a known delay and noise into both signals before running the synchronization algorithm in order to evaluate whether it successfully catches the superimposed delay.
After the first method, multiple pairs of EXP01 measurements including both 9 kHz and 2 kHz signals have been considered. The algorithm has been run on each pair across time, finding the same returned time offset for both 2 kHz and 9 kHz signals, as shown in Figure 7.

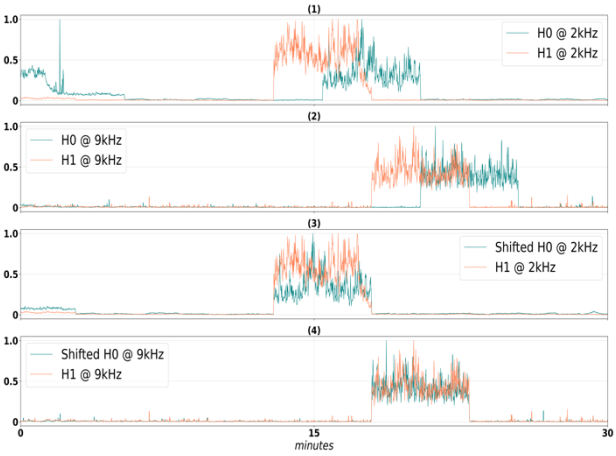

Figure 7. (1) RMS of H0 and H1 from EXP01 centered at 2 kHz which were recorded on 29/01/2024 at 14:35:19 and 14:57:29 respectively; (2) RMS of the same recordings but centered at 9 kHz; (3) result of synchronization based on RMS centered at 2kHz delivered 150.18s; (4) the same amount of time offset (150.18s) was achieved for the RMS centered at 9kHz.

To perform the second evaluation approach, 6 different delays along with 3 levels of Gaussian white noises have been defined and each time, a simulation of 40 iterations has been carried out to reduce the randomness effect of each generated noise. The estimated delays are multiples of the time resolution of the selected spectrogram which corresponds to 0.0013s and noises are added with a naive assumption of having a target SNR value. In other words, the recording is assumed to be noiseless and a Gaussian noise is added to achieve a certain SNR value. Under this assumption the standard deviation of the Gaussian noise can be calculated.
The final results are presented in Table 4 in which σ_1 and σ_2 stand for the Gaussian noise standard deviation that was added to H1 and H2 respectively. Each value represents the average difference between the inserted delay and the detected delay in seconds, and the last row shows the average over all delay values detected for a specific Gaussian noise.

Table 4. Estimated delays are reported for different combinations of introduced known delays and noise levels. σ_1 and $_2$ stand for the Gaussian noise standard deviation that was added to H1 and H2 respectively. The values in each cell represents the average difference between the detected and the actual delay tested on a series of recordings.

Added Delay	Standard deviation of noises		
	σ_1=78.9 σ_2=18.5	σ_1=44.3 σ_2=10.4	σ_1=14.0 σ_2=3.30
13.3 ms	8.4 ms	2.2 ms	1.3 ms
26.6 ms	9.5 ms	3.2 ms	2.0 ms
666.7 ms	12.8 ms	3.8 ms	2.2 ms
1333.3 ms	13.6 ms	2.4 ms	1.7 ms
2000.0 ms	29.7 ms	3.1 ms	2.1 ms
4000.0 ms	11.8 ms	2.4 ms	1.5 ms
AVG	**14.3 ms**	**2.8 ms**	**1.8 ms**

It is evident that as the noise power increases, the synchronization error increases as well, but the algorithm still shows a robust performance and achieves the requirement of a 30ms error in at least two out of three scenarios. It needs to be noted that the actual SNR of the signals is obviously much lower than the target value which proves the high accuracy of the convolution algorithm.
Based on these results it is safe to say that the algorithm based on the step array convolution with the signal filtered at the frequency of the emitted sound delivers the most accurate and robust solution. In addition, RtSys Sylence recorders show a periodic timer (probably due to electronics) which shall be better examined on a longer experiment, to check whether an actual pattern occurs.
Based on the non-linear patterns shown in Figure 6 and on the algorithm's accuracy reported in Table 4 for one-to-one synchronizations, we can infer that the synchronization protocol should be performed on pairs of recordings rather than in sequences. In other words, the synchronization of two signals cannot be done by extrapolating their predecessors offset or by interpolating the offset of their immediate neighbors' files.

CONCLUSIONS

Overall, the study contributed to test and evaluate three different synchronization protocols, which in principle may be extended to different experiments using autonomous or independent hydrophones. Also, the uncertainty in estimating the offset time through fitting a line can be due to the variation in rise time of the sound source mechanical response. A way to compensate for that is to replace the electrodynamic sound source with a piezoelectric transducer which exhibits a better phase response. Moreover, by

increasing the distance between the hydrophone pairs, the time resolution can be set to higher values as the error margin grows accordingly.

3. ACKNOWLEDGMENTS

This work has been supported by the European Union – Next Generation EU PRIN 2022 Project "SEAmPhonia - An innovative enlightening approach to enable the modelling of marine ecosystems by the acoustic 3D field" - grant number 2022AWXT3K.

REFERENCES

[1] K. Hannemann, F. Krüger, and T. Dahm: "Measuring of clock drift rates and static time offsets of ocean bottom stations by means of ambient noise," *Geophysical Journal International*, vol. 196, no. 2, pp. 1034–1042, Feb. 2014.

[2] R. Mullin: *Time warps, string edits, and macromolecules: The theory and practice of sequence comparison*. Edited by D. Sankoff and J. B. Kruskal. Addison-Wesley Publishing Company, Inc., Advanced Book Program, Reading, Mass., Don Mills, Ontario, 1983.

[3] Y. Jiang, Y. Qi, W. K. Wang, B. Bent, R. Avram, J. Olgin, and J. Dunn: "EventDTW: An Improved Dynamic Time Warping Algorithm for Aligning Biomedical Signals of Nonuniform Sampling Frequencies," *Sensors*, vol. 20, no. 9, p. 2700, May 2020.

[4] M. Müller, H. Mattes, and F. Kurth, "An Efficient Multiscale Approach to Audio Synchronization," in *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR)*, Victoria, Canada, pp. 192–197, 2006.

[5] S. Dixon and G. Widmer, "MATCH: A Music Alignment Tool Chest," in *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR)*, London, UK, pp. 492–497, 2005.

[6] J. Kammerl, N. Birkbeck, S. Inguva, D. Kelly, A. J. Crawford, H. Denman, A. Kokaram, and C. Pantofaru, "Temporal synchronization of multiple audio signals," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, pp. 4603–4607, 2014.

[7] P. F. Moretti, A. Pomaro and S. Buogo, "Metrology for Underwater Acoustics: Rethinking Measurement Strategies for the Assessment of the Environmental Status," in *2024 IEEE International Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea)*, Portorose, Slovenia, pp. 546-551, 2024.

[8] J. O. Smith: *Mathematics of the Discrete Fourier Transform (DFT): with Audio Applications*. Palo Alto: W3K Publishing, 2007.

[9] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, CJ Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E.A. Quintero, Charles R Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors: "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, no. 3, pp. 261–272, 2020.

[10] A. V. Oppenheim, R. W. Schafer, and J. R. Buck: *Discrete-Time Signal Processing*. Upper Saddle River, NJ: Prentice Hall, 1999.

[11] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," *Optics Letters*, vol. 33, no. 2, pp. 156–158, Jan. 2008.

[12] C. D. Kuglin and D. C. Hines, "The phase correlation image alignment method," in *Proceedings of the IEEE International Conference on Cybernetics and Society*, New York, USA, pp. 163–165, 1975.

[13] R. M. Haralick, "Digital step edges from zero crossing of second directional derivatives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 1, pp. 58–68, 1984.

[14] A. Savitzky and M. J. E. Golay: "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.