

## Article

# On the Challenges of Acoustic Energy Mapping Using a WASN: Synchronization and Audio Capture

Emiliano Ehecatl García-Unzueta <sup>1,\*</sup>, Paul Erick Mendez-Monroy <sup>2</sup> and Caleb Rascon <sup>3</sup>

- <sup>1</sup> Posgrado de Ingeniería Eléctrica, Universidad Nacional Autónoma de México (UNAM), Universidad 3000, Mexico City 04510, Mexico
- <sup>2</sup> Unidad Académica del IIMAS en el Estado de Yucatán, Universidad Nacional Autónoma de México (UNAM), Parque Científico y Tecnológico de Yucatán, Yucatán 97302, Mexico; erick.mendez@iimas.unam.mx
- <sup>3</sup> Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS), Universidad Nacional Autónoma de México (UNAM), Circuito Escolar 3000, Mexico City 04510, Mexico; caleb@unam.mx
- \* Correspondence: unzueta@ciencias.unam.mx; Tel.: +52-55-4361-6967

**Abstract:** Acoustic energy mapping provides the functionality to obtain characteristics of acoustic sources, as: presence, localization, type and trajectory of sound sources. Several beamforming-based techniques can be used for this purpose. However, they rely on the difference of arrival times of the signal at each capture node (or microphone), so it is of major importance to have synchronized multi-channel recordings. A Wireless Acoustic Sensor Network (WASN) can be very practical to install when used for mapping the acoustic energy of a given acoustic environment. However, they are known for having low synchronization between the recordings from each node. The objective of this paper is to characterize the impact of current popular synchronization methodologies as part of the WASN to capture reliable data to be used for acoustic energy mapping. The two evaluated synchronization protocols are: Network Time Protocol (NTP) y Precision Time Protocol (PTP). Additionally, three different audio capture methodologies were proposed for the WASN to capture the acoustic signal: two of them, recording the data locally and one sending the data through a local wireless network. As a real-life evaluation scenario, a WASN was built using nodes conformed by a Raspberry Pi 4B+ with a single MEMS microphone. Experimental results demonstrate that the most reliable methodology is using the PTP synchronization protocol and audio recording locally.

**Keywords:** wireless acoustic sensor network; synchronization; beamforming; acoustic mapping



**Citation:** García-Unzueta, E.E.; Mendez-Monroy, P.E.; Rascon, C. On the Challenges of Acoustic Energy Mapping Using a WASN: Synchronization and Audio Capture. *Sensors* **2023**, *23*, 4645. <https://doi.org/10.3390/s23104645>

Academic Editors: Lorenzo Ciani and Gabriele Patrizi

Received: 15 April 2023  
Revised: 9 May 2023  
Accepted: 9 May 2023  
Published: 10 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The analysis of the acoustic scene can be used in multiple applications: urban monitoring [1], where acoustic sensors capture information from the environment and compares it with known events to determine urban activity; bio-localization [2], where microphones arrays are deployed to determine the position of sound emitting animals; rescue work in catastrophe situations [3,4], where the sensors are deployed to measure acoustic activity of to-be-rescued entities to determine their position; domestic smart systems [5], where sensors are used to locate the user that is to be serviced inside a domestic environment; etc. In such scenarios, the acoustic information of the environment (sources of interest, noise, interference and reverberation) is obtained by processing acoustic signals captured by a monitoring system [6,7] as shown in Figure 1. The problem of detecting of acoustic events for the mentioned applications has been approached by implementing a network of sensors that capture information from the environment and processing it to determine the type of acoustic event and its position.

With this configuration, different mathematical models can be used to generate what is known as an ‘acoustic energy map’, which provides information that could be used to localize sound sources, to establish how ‘loud’ they are, etc. To carry this out, an acoustic sensor network is usually used to sample the acoustic scene [8]. However, approaching the

problem like this requires a vast amount of nodes (or microphones) in the acoustic network. Thus, we propose to employ mathematical models that rely on the concept of beamforming which do not require as many nodes in the acoustic network to provide an acoustic energy map, but heavily depend on the difference in time of arrival of an acoustic source to each of the nodes of the acoustic network. This acoustic mapping model is described in Figure 2, where  $x_i(t)$  is the acoustic signal captured for the  $i$ th node of the network, and an acoustic source is assumed to exist in a point  $(x, y)$ . The signal is formed by applying a beamformer with which it is possible to determine the energy in that point. Then, a grid of candidate points (or proof points) are proposed and for each one its energy is calculated and plotted to obtain the acoustic energy map.

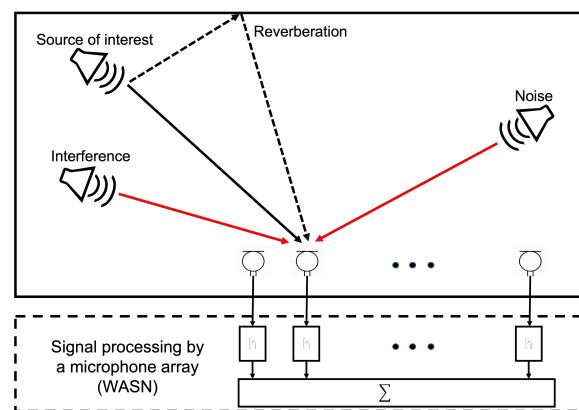


Figure 1. Acoustic signal analysis.

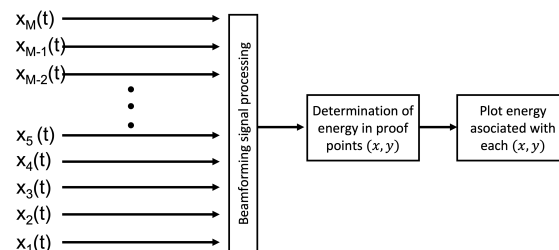


Figure 2. Methodology for acoustic energy mapping.

Additionally, it is of interest to employ a Wireless Acoustic Sensor Network (WASN) to capture the multi-channel data from which the acoustic energy map is estimated, because of their ease of installation and modification due to its nodes not being tethered by any physical cabling. However, the recordings may be affected by synchronization issues which, in turn, may affect the performance of the aforementioned beamforming-based acoustic energy mapping technique.

The main goal of the paper is to characterize how different synchronization/recording methodologies impact beamforming-based acoustic energy mapping. The methodologies employed are:

- Local recording with Network Time Protocol (NTP) synchronization. The WASN is synchronized with NTP and each node records its captured acoustic data locally.
- Local recording with Precision Time Protocol (PTP) synchronization. The WASN is synchronized with PTP and each node records its captured acoustic data locally.
- Remote recording with PTP synchronization. The WASN is synchronized with PTP and each node transmits its captured acoustic data to a centralized server.

Then, the shape of the acoustic energy map, as well as the statistical values on the synchronization error in the node timestamps, are employed to characterize these methodologies.

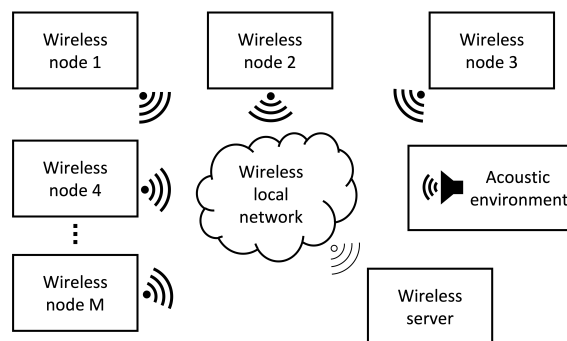
In summary, the overall process presented and characterized in this work of generating an acoustic energy map using a WASN consists on:

- The design of the WASN: obtained by analyzing the characteristics of different processing units and microphones and selecting the ones that suited the most for the desired application.
- The implementation of the synchronization protocols: it consisted on the selection of available synchronization protocols for a WASN (NTP and PTP), and their characterization in the creation of an acoustic energy map.
- The proposal of several acoustic signal recording methodologies: it took in consideration the advantages and disadvantages that each methodology presented and their evaluation in a controlled experiment.
- The generation of an acoustic energy map through beamforming techniques: there was a stage of research for beamformers techniques and the characterization of the selected ones in the application of acoustic energy mapping.

The paper is structured as follows: Section 2 describes the methodologies used for synchronization; Section 3 describes the acoustic energy mapping methodologies; Section 4 presents the results of several beamforming-based acoustic energy mapping techniques using signals from a simulated WASN and simulated sound sources, artificially inserting synchronization errors; Section 5 presents the results using signals recorded with a real-life WASN, while employing the aforementioned synchronization methodologies; Section 6 discusses the advantages, disadvantages and what are the feasible applications for each synchronization methodology; finally Section 7 presents some concluding remarks.

## 2. Wireless Acoustic Sensor Network and Synchronization

A Wireless Acoustic Sensor Network (WASN) consists of a set of sensing nodes distributed on a physical space that captures data from the acoustic environment. Each node consists of a processing device and an acoustic sensor (microphone) [9–12]. All nodes are connected through a local wireless network to receive and send data, with the final actionable device being a multi-channel audio recording WASN (each channel being provided by a node in the WASN). The basic design of a WASN with  $M$  nodes is presented on Figure 3.



**Figure 3.** Wireless Acoustic Sensor Network.

As mentioned in the introduction, beamforming-based techniques [13] are meant to be applied towards the generation of an acoustic energy map. Beamformers use the difference in arrival times of the wavefront of the acoustic source to each node to reconstruct the signal at a given position.

Thus, the multi-channel audio recording must be synchronized for them to work properly, since errors in synchronization will affect their performance, the generated acoustic energy map will not represent the actual acoustic environment [14,15].

### 2.1. WASN Design

The WASN that will be used to capture signals to generate the acoustic energy map consisted of a set of wirelessly-interconnected capture nodes, each composed by a processing unit connected to an acoustic sensor (e.g., microphone). Based on previous experimentation with different capture node configurations, a Raspberry Pi 4B+ with a MEMS (micro electret

microphone) Microphone Breakout Board was chosen as the processing unit and acoustic sensor, respectively. The processing unit was selected over other models of Raspberry and Arduinos because its features allow the network to capture, send and receive data, while bearing a relatively low cost. The MEMS microphone was selected because of its low cost, its ease of installation and its high sensibility and SNR. The characteristics of both are presented in Tables 1 and 2.

**Table 1.** Processing device characteristics.

Model	Processor	RAM	Conectivity
Raspberry Pi 4B+	Broadcom BCM2711, Quad core 64 bit @ 1.5 GHz	1 GB	2.4 GHz and 5.0 GHz IEEE 802.11ac wireless, Bluetooth 5.0

**Table 2.** Sensor characteristics.

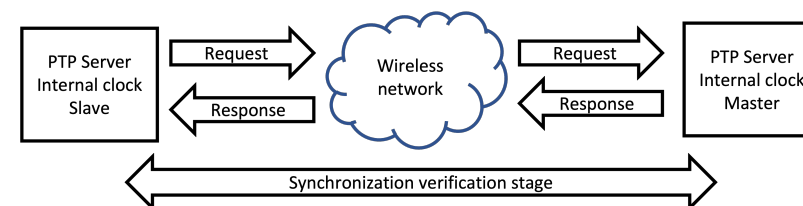
Model of Microphone	Sensitivity	Signal to Noise Ratio	Output Interface
Knowles, I2S Output Digital Microphone: SPH0645LM4H-B	−26 dBFS	65 dB(A)	I2S

## 2.2. Synchronization Protocols

Each node in the WASN has its own processing device with its own internal clock which can be affected by physical interactions with the environment, such as: temperature, pressure, vibration, etc. When the acoustic signal is captured, the timestamps of each audio file have to be synchronized. The synchronization represents a problem that can be solved by using extra hardware (like an external clock) or by applying synchronization protocols to synchronize all of the node's internal clocks [16,17]. There are different strategies that can be implemented to achieve this, most of which use a server to exchange messages with their respective timestamps, which then are used to adjust the internal clocks of the nodes of the WASN. The most commonly used are: Network Time Protocol (NTP) (Figure 4) and Precision Time Protocol (PTP) (Figure 5) [18,19]. Both protocols exchange messages through a network between a master-server and a client-slave. The client-slave sends a request to the server-master and when it responds, the client-slave adjusts its internal clock. The difference between NTP and PTP is that for the latter, there is an extra stage of verification for the synchronization, which results in higher precision.



**Figure 4.** NTP.



**Figure 5.** PTP.

A synchronization implementation is proposed as follows: one of the nodes of the WASN is the time server for the network, the internal clock of which is the one that all the

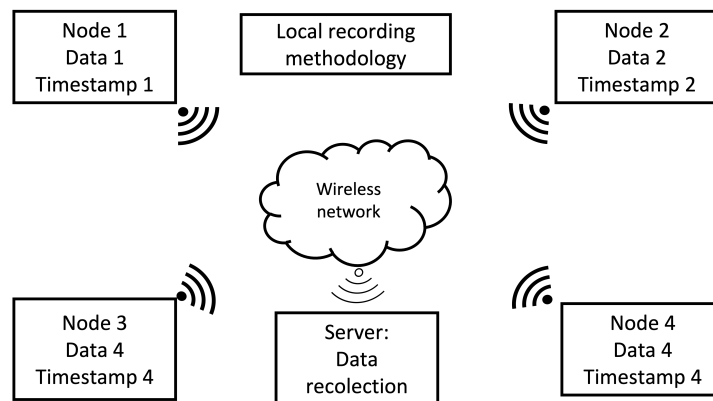
other nodes will adjust their time to. There is also the possibility for the time server node to synchronize its internal clock with a publicly-available time server.

Although only the the NTP and PTP protocols were implemented, it is left for future work to explore other strategies for synchronization and protocols.

### 2.3. Recording Methodologies

#### 2.3.1. Local Recording Methodology

In this methodology, each node is responsible for capturing acoustic data from the environment, store it, to later send such data to a central processing unit to generate the acoustic energy map. Each node generates its own timestamp according to its internal clock. This methodology is presented in Figure 6.



**Figure 6.** Local recording methodology.

Different synchronization protocols can be used with this recording methodology, thus, we propose the following:

- Local recording with NTP. Each client node synchronizes its internal clock with the server node on a background process, and the server node synchronizes its internal clock with a time server service on the internet.
- Local recording with PTP. It operates on a similar fashion to the previous methodology, but the slave nodes synchronize their internal clocks with the clock from the master.

In this type of recording methodology, each node is responsible for monitoring the acoustic environment and generating and storing audio and timestamps files locally. The data processing for this type of methodology requires the adjustment of the timestamps of each node to a general frame of time. After the adjustment, the beamforming process is carried out to generate the acoustic energy map.

#### 2.3.2. Remote Recording Methodology

This methodology generates only one timestamp for all the audio signals captured by the nodes in the WASN, shown in Figure 7. Hence, instead of storing the recorded data locally, the data is sent through the local network to a central processing unit for storage. The advantage of this methodology is that there is no need for post-processing multiple timestamps. Each node of the WASN only needs to capture the acoustic data and transmit it, reducing the load of all the processing units.

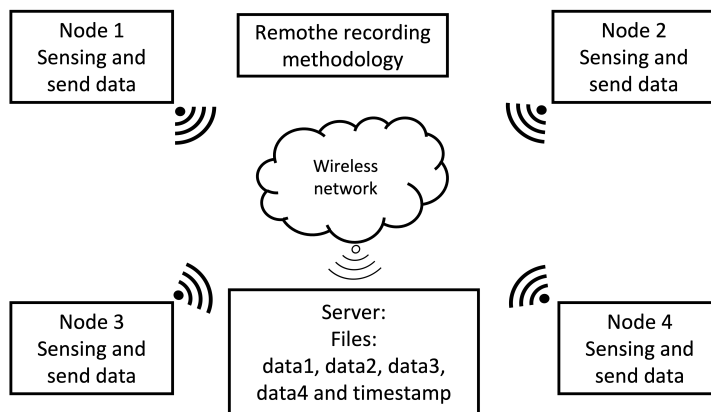


Figure 7. Remote recording methodology.

### 3. Beamforming-Based Acoustic Energy Mapping

A beamformer usually employs a series of weights (or steering vector) to estimate the radiation pattern related to a given steered position. These weights are calculated based on the respective time difference of arrival, for the given steered position, for each node in the WANS. To this effect, a collection of points  $(x, y)$  is proposed, referred here as “proof points”. A beamformer and its respective set of weights are applied as if there was an acoustic energy source at each proof point. The signal associated to that point is reconstructed, and its energy calculated and recorded for its respective proof point. The result is an acoustic energy map which was “sampled” at each proof point.

The aforementioned time differences depend on the geometry of the WASN and the distance between the acoustic source and each node, as exemplified in Figure 8.

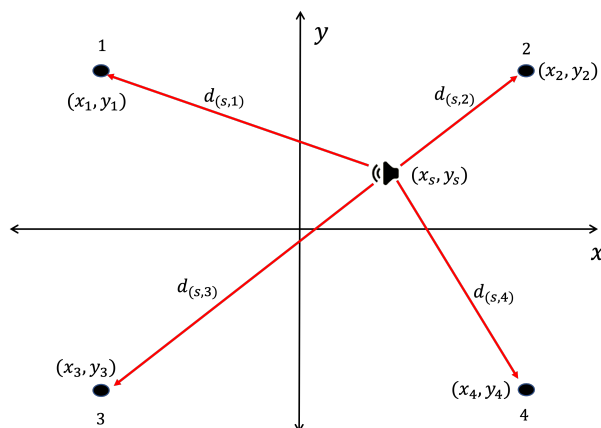


Figure 8. Distance between acoustic source and the nodes of a 4 nodes WASN process.

Let the amount of nodes inside the WASN be  $M$ , thus, the Cartesian coordinates of each can be written as  $(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_M, y_M)$ . Let the frequency-domain signal captured at each node  $i$ th node be  $X_i$ . Additionally, let the acoustic source position be  $(x_s, y_s)$  (which in this case, is the location of a given proof point). The time of arrival of the wavefront of the acoustic source to each  $i$ th node ( $t_i$ ) can be calculated by:

$$t_i = \frac{\sqrt{(x_s - x_i)^2 + (y_s - y_i)^2}}{c} \tag{1}$$

where  $c = 343 \frac{m}{s}$  is the speed of sound in the air. Consequently, the vector of arrival times ( $T = [t_1, t_2, \dots, t_M]$ ) can be expressed as:

$$T = \left[ \frac{\sqrt{(x_s - x_1)^2 + (y_s - y_1)^2}}{c} \quad \frac{\sqrt{(x_s - x_2)^2 + (y_s - y_2)^2}}{c} \quad \dots \quad \frac{\sqrt{(x_s - x_M)^2 + (y_s - y_M)^2}}{c} \right] \tag{2}$$

To reconstruct a signal  $S(x, y)$  at a given proof point located at  $(x, y)$ , a series of weights  $w_i(x, y)$  are applied to the input signals  $X_i$ , namely:

$$S(x, y) = \sum_{i=1}^M w_i(x, y)^H X_i \quad (3)$$

where  $w_i(x, y)^H$  is the complex conjugate factor. As mentioned earlier, the weights are based on the time differences of arrival at each node, for each possible proof point. Their calculation differ between beamforming methodologies. The beamformers proposed for the generation of the acoustic energy map are:

- Delay and Sum (DAS).
- Minimum Variance Distortionless Response (MVDR).
- Steered-Response Power with Phase Transform (SRP-PHAT).
- Phase-based Binary Masking (PBM).

Once the signal is reconstructed for a particular proof point, its energy is proportional to the signal amplitude squared.

$$E(x, y) \propto \|S(x, y)\|^2 \quad (4)$$

The acoustic energy map is then obtained by proposing a grid of proof points and plotting the total energy associated of the signal in a period of time to each of them.

In the rest of these sections, the beamformer techniques are briefly explained, noticing their differences, advantages and disadvantages.

### 3.1. Delay and Sum (DAS)

This is the simplest version of the beamforming techniques, since it relies in shifting the input signals such that the information that is received from a given steered position is aligned across all the input signals [20]. Thus, if summed, the signal in the given steered position is amplified. To this effect, the weights for this beamformer are simply a series of delay operators that counter the occurring time differences of arrival. That is to say, they are calculated as:

$$w_i(x, y) = \begin{bmatrix} e^{-i2\pi\omega_1 t_1} & e^{-i2\pi\omega_2 t_1} & \dots & e^{-i2\pi\omega_N t_1} \\ e^{-i2\pi\omega_1 t_2} & e^{-i2\pi\omega_2 t_2} & \dots & e^{-i2\pi\omega_N t_2} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-i2\pi\omega_1 t_M} & e^{-i2\pi\omega_2 t_M} & \dots & e^{-i2\pi\omega_N t_M} \end{bmatrix} \quad (5)$$

where  $\omega_l$  is the  $l$ th frequency bin. Usually,  $t_1$  is set to 0, and all the other  $t_i$ 's are set relative to that time, such that  $t_i = t_1 - t_i$ . It is important to note that this set of weights is known as the 'steering vector', and is re-used in other beamformer techniques.

### 3.2. Minimum Variance Distortionless Response (MVDR)

This technique approaches the problem of reconstructing the signal by aiming to optimize the weights such that the energy that is not coming from a steered position is minimized [21]. It does so by first expressing the output energy ( $|S(x, y)|^2$ ) in terms of the co-variance ( $R = XX^H$ ) between the input signals ( $X = [X_1, X_2, \dots, X_M]$ ). Due to Equation (3), and  $W(x, y) = [w_1(x, y), w_2(x, y), \dots, w_M(x, y)]$ , it is derived that:

$$|S(x, y)|^2 = |W(x, y)^H X|^2 = (W(x, y)^H X)(X^H W(x, y)) = W(x, y)^H R W(x, y) \quad (6)$$

It carries out a minimization of such energy to obtain its optimized weights  $W_{op}(x, y)$ , with the following restriction:

$$W_{das}(x, y)^H W_{op}(x, y) = 1 \quad (7)$$

where  $W_{das}$  is the steering vector as calculated in Equation (5). This minimization has the following generalized solution:

$$W_{op}(x, y) = \frac{R^{-1}W_{das}(x, y)}{W_{das}(x, y)^H R^{-1}W_{das}(x, y)} \quad (8)$$

### 3.3. Steered-Response Power with Phase Transform (SRP-PHAT)

This technique tries to maximize the output signal of a DAS beamformer but in a determined direction, and directly provides the power of the signal that arrives from such direction [22,23].

The power of the output signal  $S(x, y)$ , as described in Equation (3), is given by:

$$P(x, y) = \int_{-\infty}^{\infty} |S(x, y, \omega)|^2 d\omega \quad (9)$$

where  $S(x, y, \omega)$  is the  $\omega$  frequency of  $S(x, y)$ .

This integral is called the steered-response power and its discrete form can be written as:

$$P(x, y) = \sum_{n=0}^{M-1} \sum_{m=n+1}^{M-1} R_{n,m}(\tau_{n,m}(x_s, y_s)) \quad (10)$$

where  $\tau_{n,m}$  is the time difference of arrival between nodes  $n$  and  $m$  of source located at  $(x_s, y_s)$ , and can be calculated similarly as in Equation (1), meaning:

$$\tau_{n,m}(x_s, y_s) = \frac{\sqrt{(x_s - x_n)^2 + (y_s - y_n)^2} - \sqrt{(x_s - x_m)^2 + (y_s - y_m)^2}}{c} \quad (11)$$

Furthermore, the factor  $R_{n,m}(\tau_{n,m}(x_s, y_s))$  is the cross-correlation between the  $n$ th and the  $m$ th microphone of the network and it is calculated by:

$$R_{n,m}(\tau_{n,m}(x_s, y_s)) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Psi_{n,m}^{PHAT}(\omega) X_n(\omega) X_m^*(\omega) e^{i\omega\tau} d\omega \quad (12)$$

where  $\Psi_{n,m}^{PHAT}$  is the weighting function in the frequency domain and the phase transform is:

$$\Psi_{n,m}^{PHAT}(\omega) = \frac{1}{|X_n(\omega) X_m^*(\omega)|} \quad (13)$$

Written in its discrete form,  $R_{n,m}(\tau)$  is:

$$R_{n,m}(\tau_{n,m}(x_s, y_s)) = \sum_{l=1}^L \frac{X_n(\omega) X_m^*(\omega)}{|X_n(\omega) X_m^*(\omega)|} e^{i\omega\tau_{n,m}(x_s, y_s)} \quad (14)$$

It is important to note that the output of this technique is not a reconstructed signal, but the power of the signal at a given point  $(x, y)$ . When carrying out this technique over different points, a 2-dimensional steered power response is provided, which be considered as the acoustic energy map for our purposes.

### 3.4. Phase-Based Binary Masking (PBM)

Phase-based Binary Masking (PBM) [24,25] is not an actual beamformer, but it has a similar application for the acoustic mapping. This technique employs the first stage of the DAS beamformer, such that the information arriving from a given steered position is aligned in all nodes. Then, the average phase difference  $|\varphi|_{\omega}$  is calculated between all of the possible pairs of  $m$ th and  $n$ th nodes in the network, for all  $\omega$  of the input signal.

$$|\varphi|_{\omega} = \frac{2}{M(M-1)} \sum_{m=1}^{M-1} \sum_{j=n+1}^M |\varphi_{m,\omega} - \varphi_{n,\omega}| \quad (15)$$



where  $\varphi_{m,\omega}$  is the phase of the signal captured at node  $m$  at frequency  $\omega$  after it being aligned by multiplying it by the weights in Equation (5).

Consequently, a phase difference threshold ( $\varphi_{max}$ ) is then used to establish if  $\omega$  is a frequency component of the reconstructed signal or not, effectively creating the following binary mask:

$$B(x, y, \omega) = \begin{cases} 1, & \text{if } |\varphi|_{\omega} \leq \varphi_{max} \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

The output signal is then calculated as:

$$S(x, y, \omega) = B(x, y, \omega) * X_1(\omega) \quad (17)$$

#### 4. Simulation Results and Analysis

In this section, the impact of synchronization errors in the generation of the acoustic energy map are studied to determine the required precision levels. The hypothesis presented is that if there is an error in synchronization, the shape of the acoustic energy map will be modified which will result in an error of the position of the acoustic sources.

For this analysis, an acoustic source is simulated on a known fixed point  $(x, y)$ , as well as a WASN with four nodes on a rectangular distribution with the source located in the same plane. For simplicity, the position of the source is set at  $(x, y) = (0, 0)$ , so the distance between each microphone and the source is the same.

Then, the four previously described beamformers are used to reconstruct a signal at each given proof point. The obtained acoustic map can be represented in a 3D plot where the XY-plane is the physical plane of the WASN and the Z-axis represents the acoustic energy associated to each point. Another way to present the acoustic map is as a heat-map, where each proof points is presented with a determined color associated to its energy.

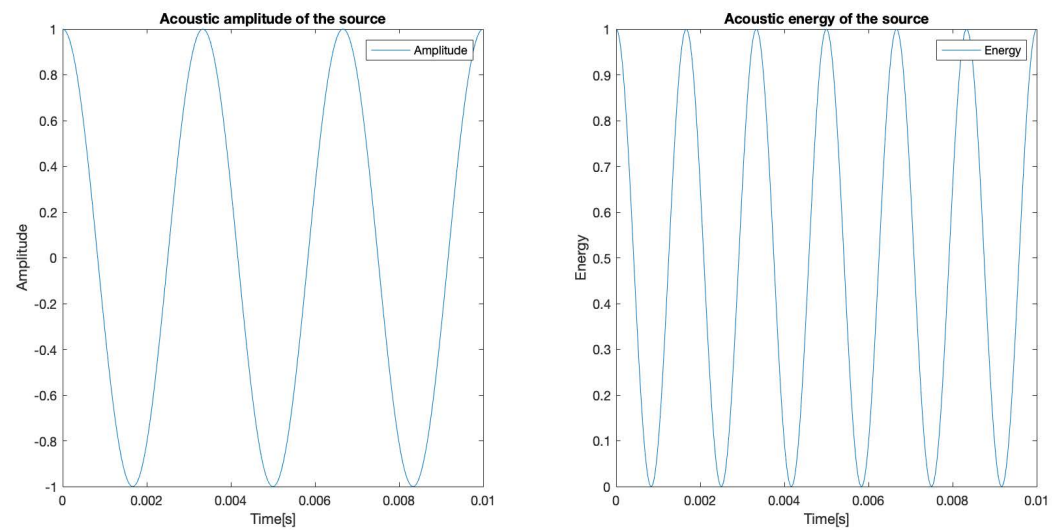
The characteristics of the simulated acoustic environment are:

- Sampling rate of the capturing nodes:  $f_s = 48,000$  Hz
- Speed of sound in air:  $c = 343 \frac{\text{m}}{\text{s}}$
- Amplitude signal of the source =  $\cos(2 * \pi * \omega * t)$
- Frequency of the source:  $\omega = 300$  Hz
- Time between samples:  $\tau_s = \frac{1}{f_s} = 0.0208$  ms
- Coordinates of the source:  $(x, y) = (0, 0)$
- Coordinates of the nodes:
  - $(x_1, y_1) = (-0.32, 0.32)$
  - $(x_2, y_2) = (0.32, 0.32)$
  - $(x_3, y_3) = (-0.32, -0.32)$
  - $(x_4, y_4) = (0.32, -0.32)$

As it was mentioned, the time-dependent energy function is proportional to the squared amplitude of the time-dependent signal function, for simplicity and without losing generality, the proportionality constant is proposed to be 1, so that the energy function is:

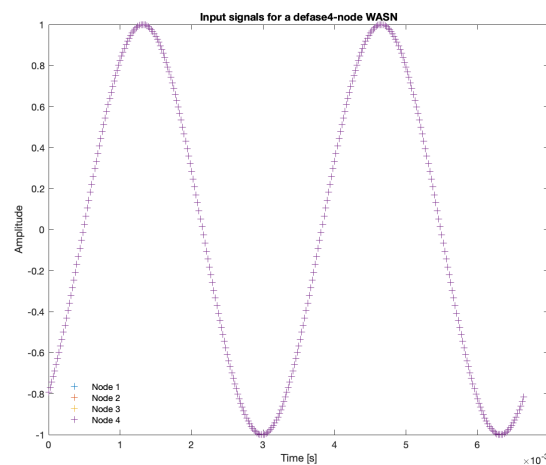
$$E(x, y, t) = \|S(x, y, t)\|^2 \quad (18)$$

The original amplitude and energy for the source are represented in Figure 9:



**Figure 9.** Original amplitude and energy signal of the source.

The signals captured by the four nodes of the simulated WASN are shown in Figure 10. In this image, all of the signals are overlapped since all of them have the same arrival time. In the synchronization analysis, one of this nodes will be artificially de-synchronized so that the signal is delayed a certain amount of samples.



**Figure 10.** Input signals of the four nodes of the WASN.

The acoustic maps obtained by applying each of the beamformers to the simulated input signals are shown in Figure 11, and their respective heatmaps are shown in Figure 12.

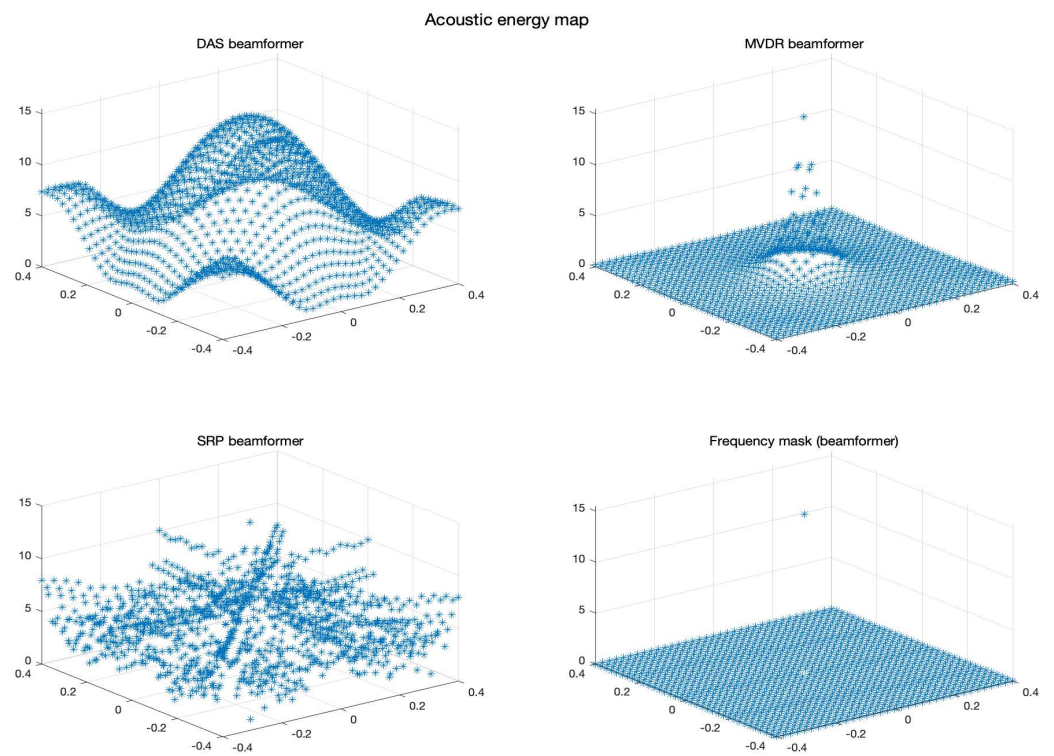


Figure 11. Acoustic energy map with: DAS, MVDR, SRP-PHAT and PBM beamformers.

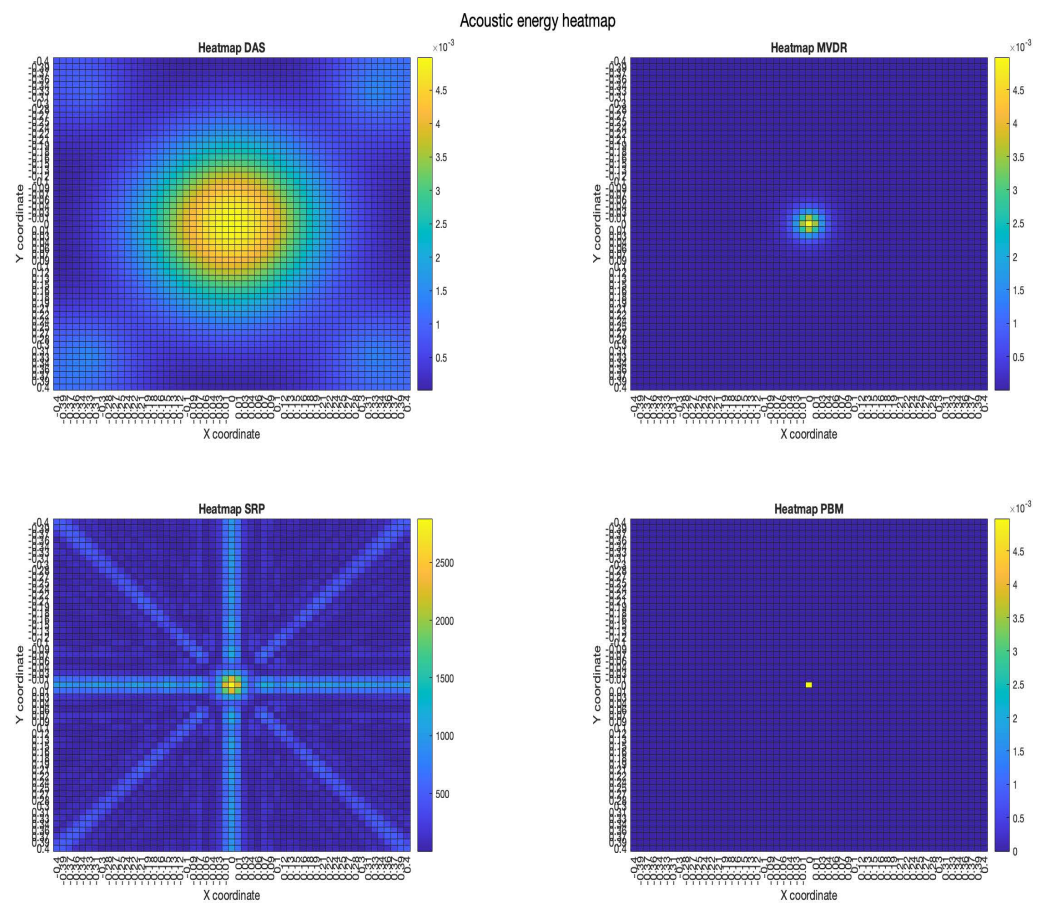
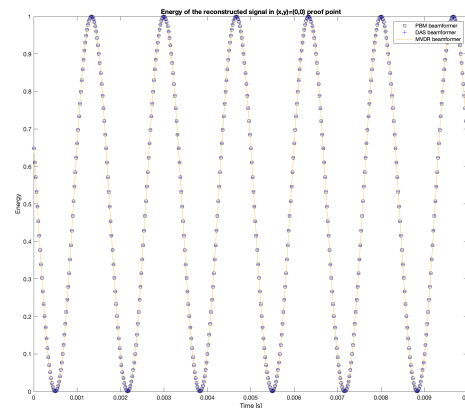


Figure 12. Heatmap of acoustic energy for beamformers: DAS, MVDR, SRP-PHAT and PBM.

The proof point with the largest energy is  $(x, y) = (0, 0)$ . The energy obtained with the reconstructed signal in this proof point with the DAS, MVDR and PBM beamformers are shown in Figure 13. Additionally, the 2D steered power response provided by SRP-PHAT is also shown.



**Figure 13.** Energy of the reconstructed input signal at  $(x, y) = (0, 0)$ .

To analyze the performance on the generation of the acoustic energy map, the total energy in a period of time will be calculated. In the simulated signal, the acoustic energy in a period of time can be calculated by the integral and approximated by the trapezoidal method with the following expression:

$$\int_{t=0}^{t=t_f} S(x, y, t) dt \approx \frac{t_f}{2N} \sum_{n=1}^N (S(x, y, t_n) + S(x, y, t_{n+1})) := \sum_{n=1}^N S(x, y, t) \quad (19)$$

where  $S(x, y, t)$  is the simulated signal of the source,  $t_f$  is the time length of the signal, and  $N$  is the size of the signal in number of samples. The calculated sum for the simulated signal is:

$$\frac{t_f}{2N} \sum_{n=1}^N (\cos^2(2 * \pi * \omega * t_n) + \cos^2(2 * \pi * \omega * t_{n+1})) = 0.0049791 \quad (20)$$

While the calculated sum for the reconstructed energy in proof point  $(0, 0)$  is:

$$\sum_{n=1}^N S_{DAS}(0, 0, t) = \sum_{n=1}^N S_{MVDR}(0, 0, t) = \sum_{n=1}^N S_{PBM}(0, 0, t) = 0.0049861 \quad (21)$$

This shows that the energy shown in the acoustic map is quite close to the energy of the original simulated signal.

Furthermore, the results show that the beamformers generate an acoustic energy map with an important amount of energy centrally located around the position where the simulated signal originated. However, there are some considerations that need to be addressed:

- There are differences in the shape of the acoustic energy map obtained with each beamformer. In this regard, some are more robust against noise and reverberation (such as MVDR), while others are more efficient in terms of processing time and simplicity on their implementations (such as DAS and PBM). The main objective of this analysis is in terms of performance in the presence of errors in synchronization, thus, the analysis of these other elements are left for future work.
- The value of the calculated energy from the reconstructed signal in the proof point  $(x, y) = (0, 0)$  for DAS, MVDR and PBM has a small 0.14% error in comparison to the energy of the original input signal. This can be considered as evidence of good

- performance of these beamformers when it comes to reconstructing the source of interest in the proof point that matches the simulated position of the source.
- DAS, MVDR and SRP-PHAT provide moderately high energy values in proof points that are not at the position of the source. The PBM beamformer finds energy only in the simulated position of the source.
  - The PBM beamformer seems to be the most accurate for locating the position of the source. However, this technique requires a parameter to be set *a-priori*, thus, there is an extra stage of calibration for the implementation of this technique.
  - The SRP-PHAT beamformer, like DAS and MVDR, finds energy in points where there is not acoustic energy from the source, but also encounters a pattern that is not representing an expected effect of the beamformer.

### Simulation of Synchronization Error

The acoustic energy map model for the four proposed beamformers results in the correct localization of the simulated acoustic source. Now, their performance in the presence of error in synchronization will be analyzed.

It is important to remember that the beamformers rely on the proper synchronization of node network, which is not always the case with WASNs. To simulate this, a delay (established as a determined number of samples) is artificially inserted in the first simulated node, located in  $(x_1, y_1) = (-0.32, 0.32)$ , in the upper left position of the array, referred here as “sync-error node”.

A synchronization error implies that before the signal arrives to the delayed node, there is no acoustic signal, as shown in Figure 14: the delayed input signal bears a value of zero for the simulated delay time. The beamformers may have problems with this discontinuity in the time domain, which may be solved by applying a window (such as Hann or Hamming) to reduce its effects, but this analysis is left for future work.

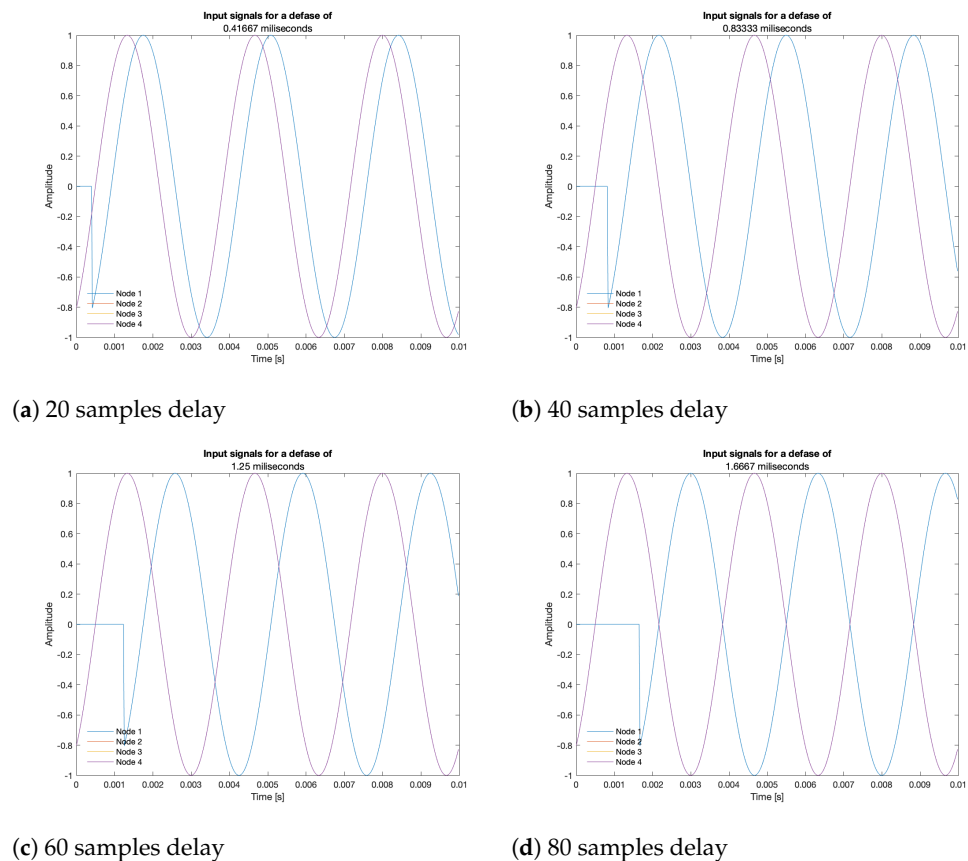
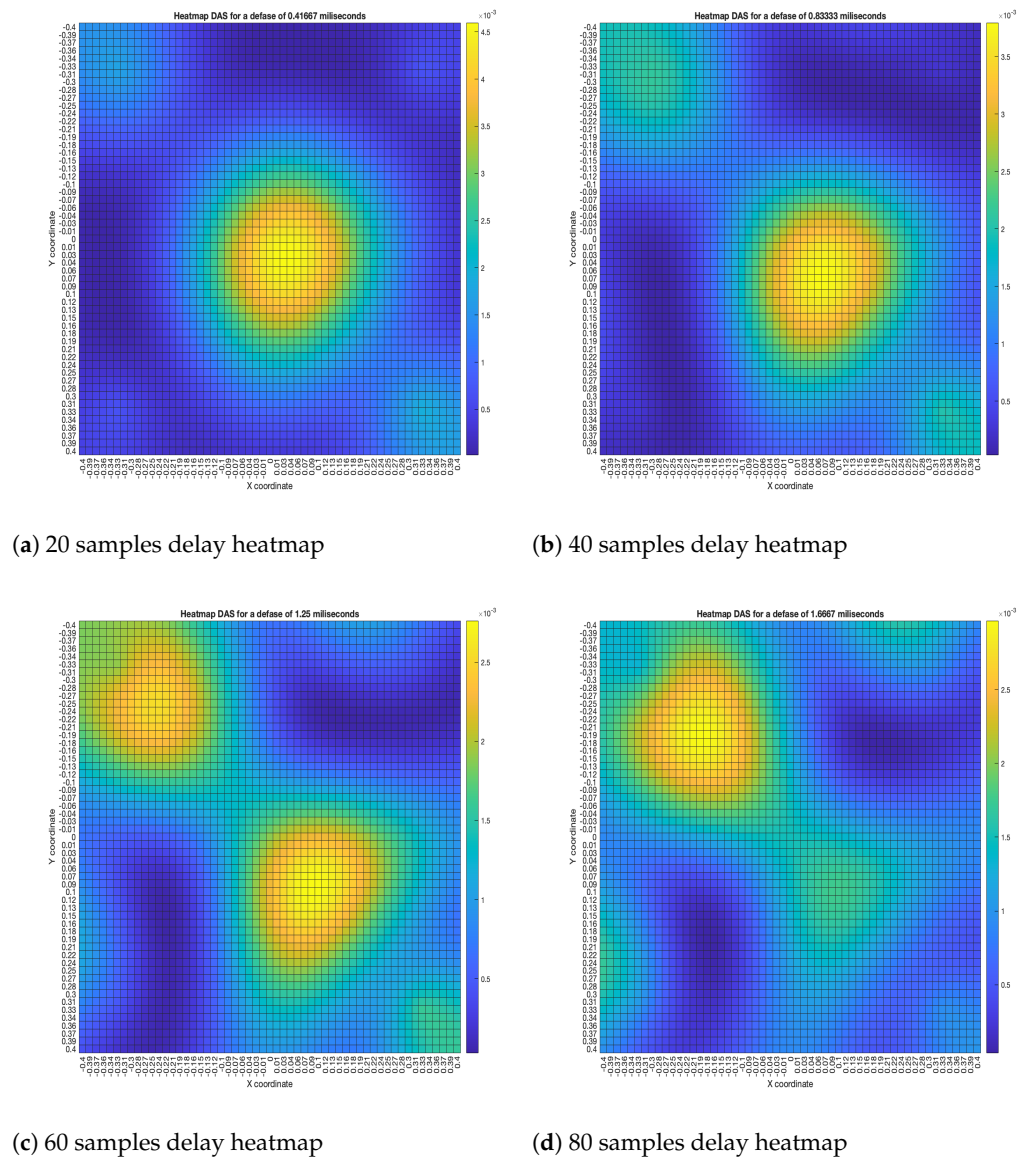


Figure 14. Input signals delayed by 20 samples intervals.

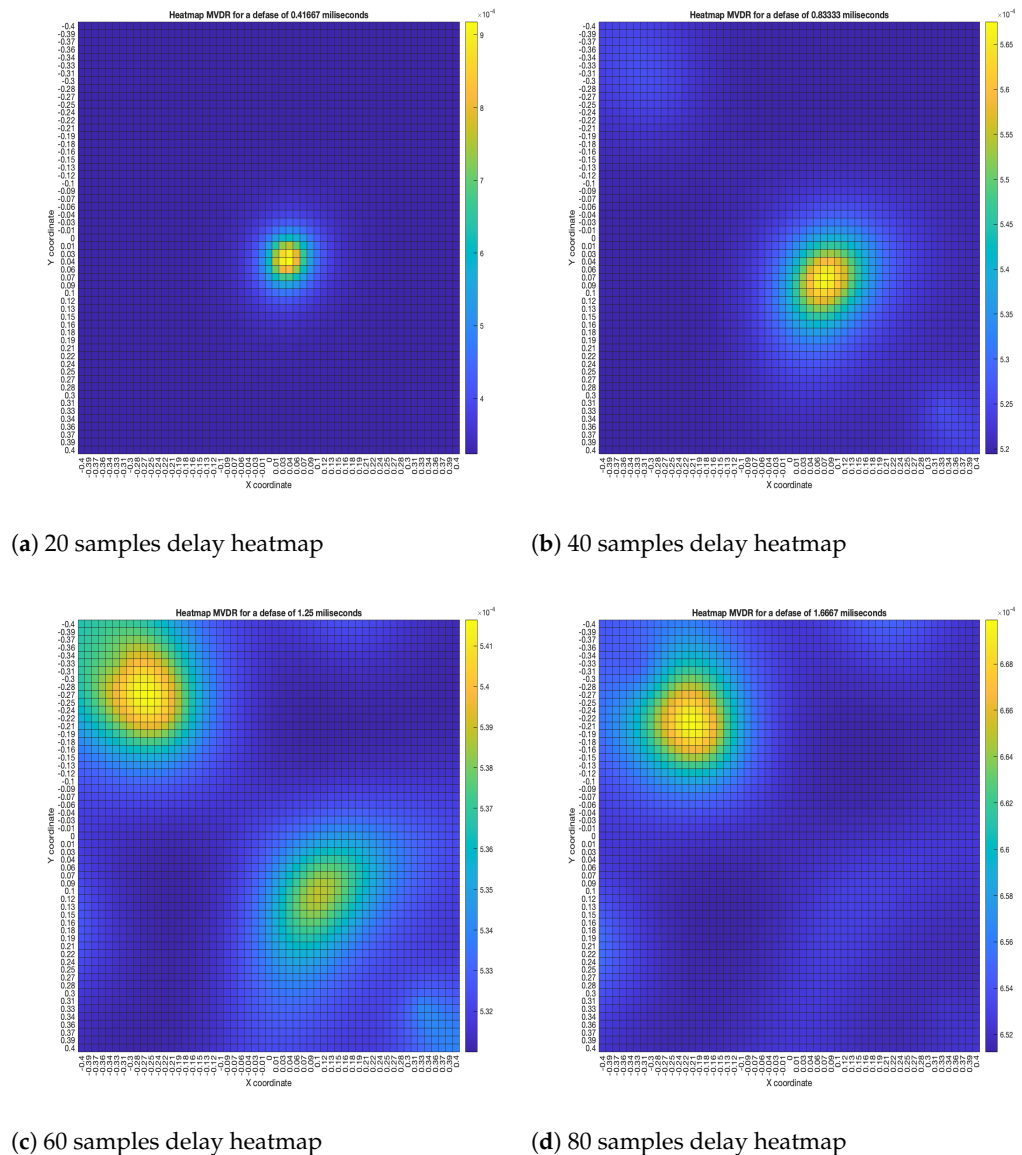
This simulated synchronization error is then fed to the beamformers, and its impact is observed in the following figures, and quantified in the change in the shape of the acoustic energy map, in the position of the source and in the value of its reconstructed energy.

The DAS beamformer energy maps are shown in Figure 15. As it can be seen, the position of the maximum value of energy (and its surrounding “hill” or lobe) gets reallocated to a different position as the samples delay increases, shifting away from the sync-error node. In Figure 15a,b, this behavior is consistent up until the 60-sample delay; from then on, another lobe appears close to the sync-error node.



**Figure 15.** DAS energy heatmap for the simulated delay in capture.

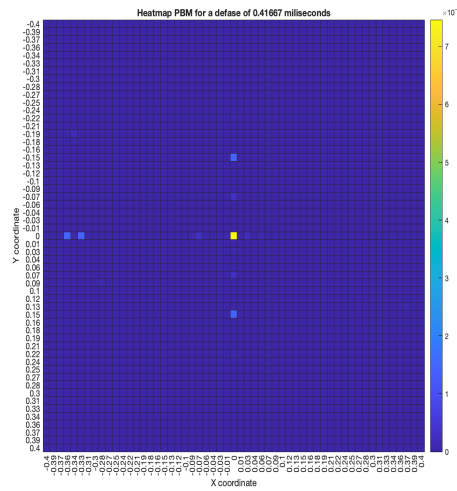
The “moving” trends of the DAS beamformer are also present in the MVDR beamformer, as it is shown in Figure 16. However, an important difference is that the size of the lobes are far smaller, providing better precision.



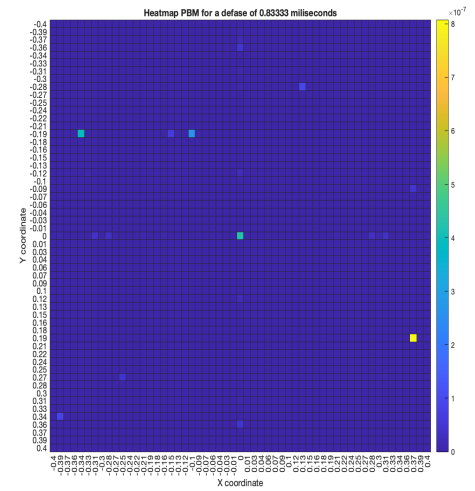
**Figure 16.** MVDR energy heatmap for the simulated delay in capture.

The reason why these trends occur is many-fold: these beamformers rely on the time-difference-of-arrival between nodes, and a delay adds to this, resulting in the source “moving”; they also compensate these time differences by artificially shifting the input signals, which bears a cyclical nature, causing the “return” to the sync-error node. In any case, it can be seen that for DAS and MVDR the shape of the acoustic energy map changes if the nodes of the network aren’t synchronized. This results in a change of the estimated position of the source, determined by locating the maximum value of energy in the map.

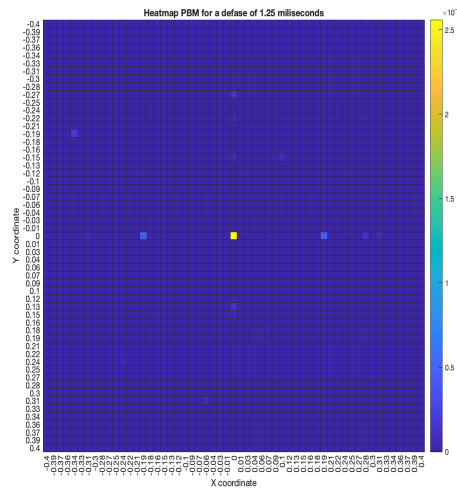
As for the PBM beamformer, as it can be seen in Figure 17, with most simulated delays, the maximum energy value is in the  $(x, y) = (0, 0)$  position. However, there are some simulated delays in which this is not the case.



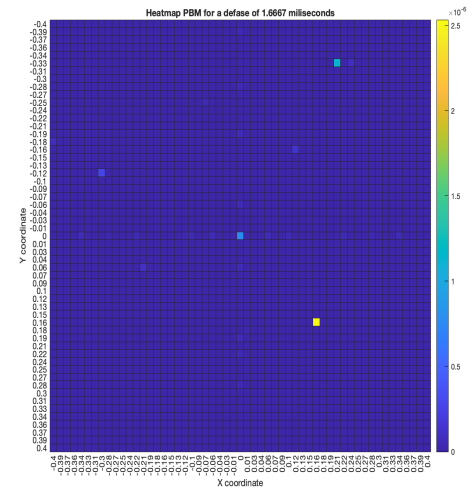
(a) 20 samples delay heatmap



(b) 40 samples delay heatmap



(c) 60 samples delay heatmap

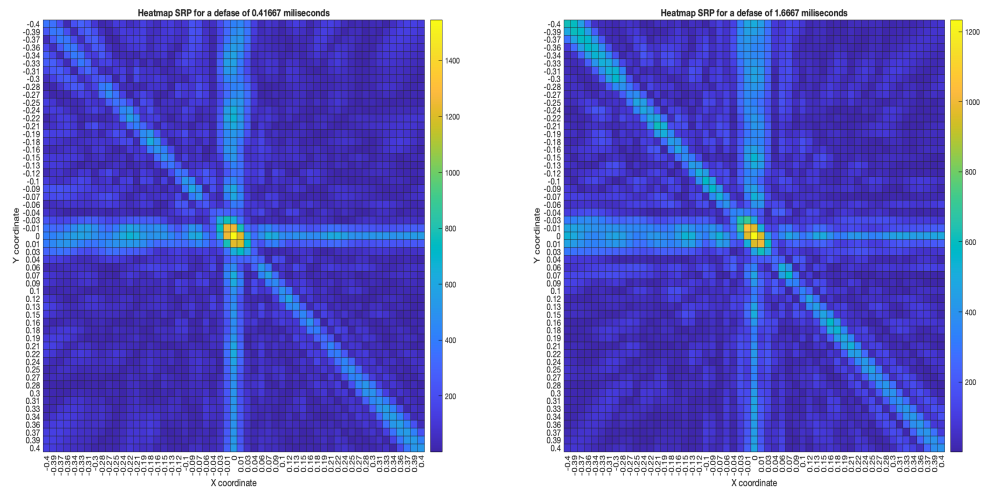


(d) 80 samples delay heatmap

**Figure 17.** PBM energy heatmap for the simulated delay in capture.

As for the SRP-PHAT, since it does not calculate the actual energy (only the steered power), the heatmaps shown in Figure 18 aren't technically acoustic energy maps. However, they can still be used for locating acoustic sources. It can be seen that the shape of the steered power response map doesn't change much, even between 20 and 80 delay samples (reason which no other heatmaps is shown in Figure 18). Meaning, this technique is mainly used to only determine the most probable proof point where a single source may be located.





(a) 20 samples delay heatmap

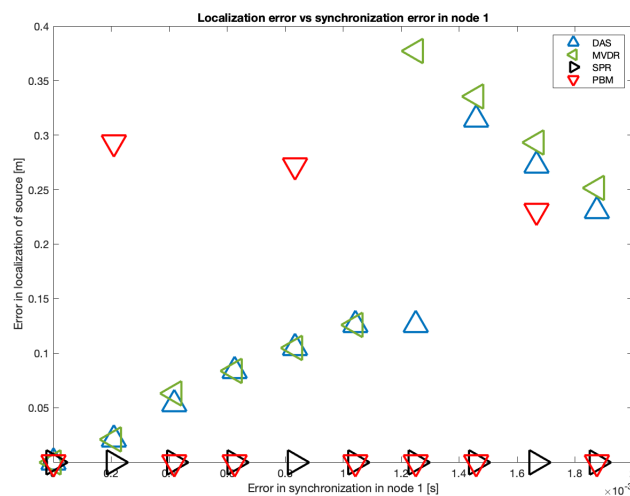
(b) 80 samples delay heatmap

**Figure 18.** SPR-PHAT energy heatmap for the simulated delay in capture.

With these results it is possible to find the position of the acoustic source by locating the maximum value of energy. Thus, the next part of the analysis is to determine, if there is any, a relationship between the errors in synchronization and the position shift of the acoustic source. After finding the position of the acoustic source, the difference between the known position of the source  $r_s$  and the obtained position of the source  $r_m$  was calculated:

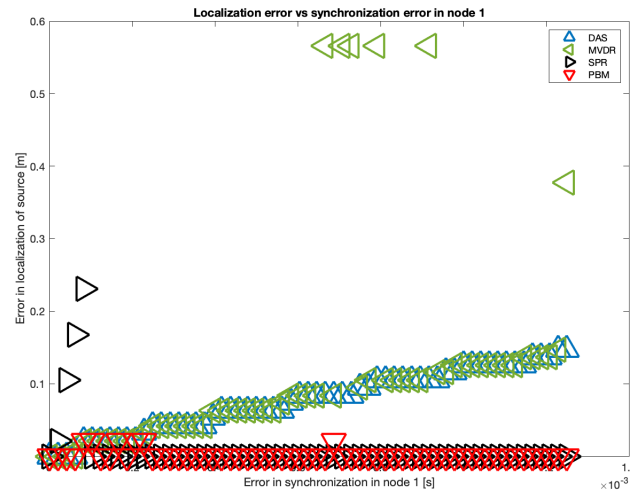
$$\Delta \vec{r} = |\vec{r}_s - \vec{r}_m| \tag{22}$$

The results of the simulation in 10 samples intervals are shown in Figure 19. It can be seen that there is a tendency in the DAS and MVDR beamformers to increase the localization error as the error in synchronization increases. The PBM and the SRP-PHAT beamformer present the lowest errors in localization, but the PBM beamformer generates errors in localization for particular values of delay (not with the tendency observed in DAS and MVDR). The SRP-PHAT beamformer has the most consistent performance, but given the noisiness of the shape of its steered response, it should only be used for acoustic source localization, not for the generation of an acoustic energy map.



**Figure 19.** Localization errors for a source in  $(x, y) = (0, 0)$  and errors in synchronization in node 1.

As it can also be seen in Figure 19, there is a considerable “jump” in localization error over the 60-sample simulated delay. Thus, this is the upper delay value we used to establish a relationship between localization error and synchronization error. In Figure 20, this relationship is shown using all possible delays in the range of 1 to 60 samples:



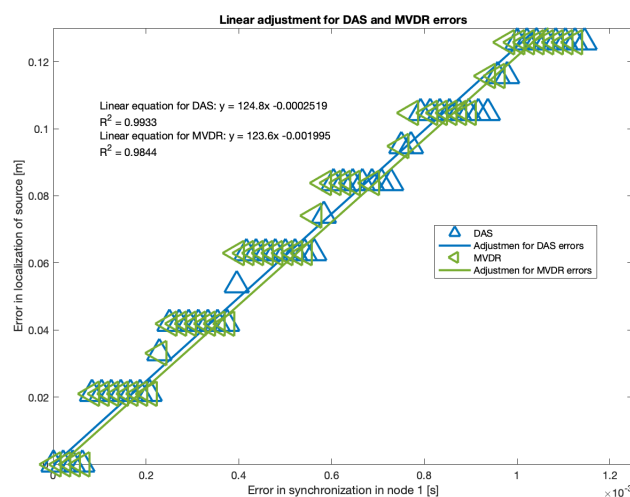
**Figure 20.** Errors in localization for all beamformers.

As it can be seen, there is a near-linear relationship between localization error and synchronization for DAS and MVDR. Applying basic statistic fitting, these are the models that best describe this relationship:

- $\Delta\vec{r}_{DAS} = 124.8\Delta t - 0.0002519$ , with  $R^2 = 0.9933$
- $\Delta\vec{r}_{MVDR} = 123.6\Delta t - 0.0019950$ , with  $R^2 = 0.9844$

where  $\Delta t$  is the synchronization error in node 1.

It is important to note that MVDR has some values for  $\Delta t$  where  $\Delta\vec{r}$  is much larger than the rest, thus, were treated as outliers for the fitting process. The resulting location-synchronization models for DAS and MVDR are shown in Figure 21:



**Figure 21.** Fitting for the DAS and MVDR beamformer errors in localization.

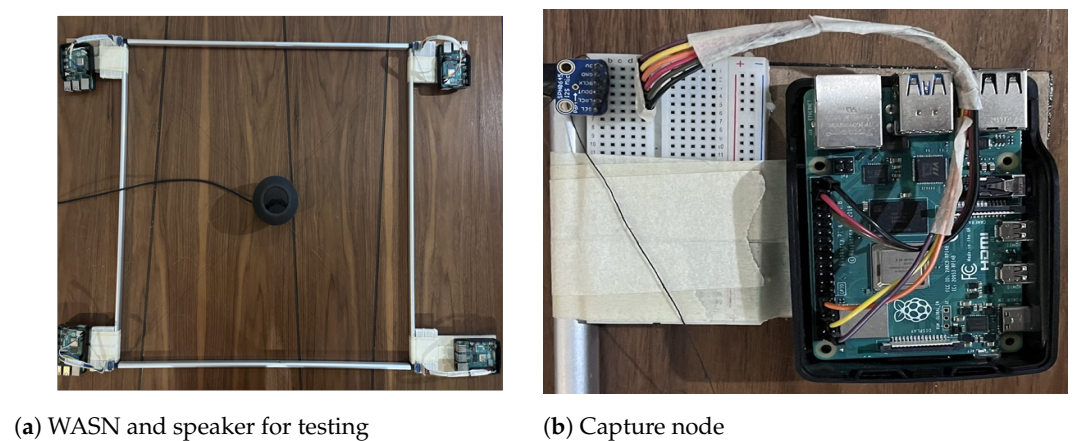
There is a “step” behavior in Figure 21, which can be explained by the grid-type search that the acoustic energy map is based on. Additional simulations were carried out with a larger number of proof points, and it was observed that these “steps” were closer to each other, i.e., the grid is refined. The maximum values for synchronization errors didn’t

change, but the processing time did increased considerably. Thus, there is a compromise to be made: a refined “step” behavior for a larger processing time.

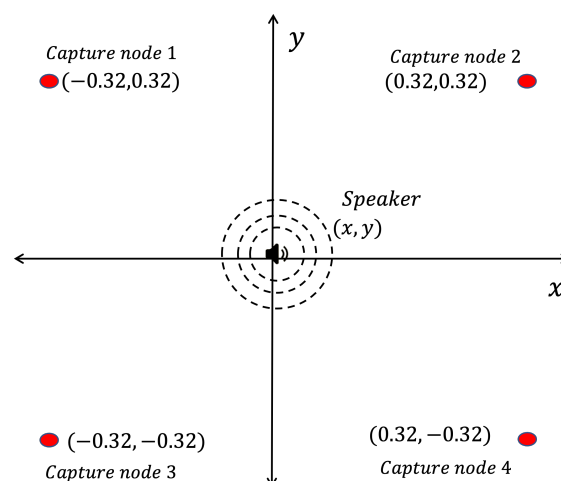
As for the PBM and the SRP-PHAT beamformers, they tend to locate the acoustic source exactly in the position where it was simulated in a near-consistent manner across the simulated time-delays, having an error near to zero for the same limit of 60 samples of delay. Thus, no relationship model was calculated for these beamformers.

## 5. Experimental Results and Analysis

In this section, the recording methodologies proposed for the capture of acoustic signal and its application on the generation of the acoustic energy map will be analyzed. A WASN composed by the devices described in Section 2 was used to capture an acoustic source (a speaker emitting an acoustic sinusoidal wave of 300 Hz) placed in the center of the WASN. The network was built into a rectangular array with the dimensions presented in Section 4. The WASN and the capture node are shown in Figure 22a,b. The schematic representation of the WASN and the source is shown in Figure 23.



**Figure 22.** Experimental design for testing.

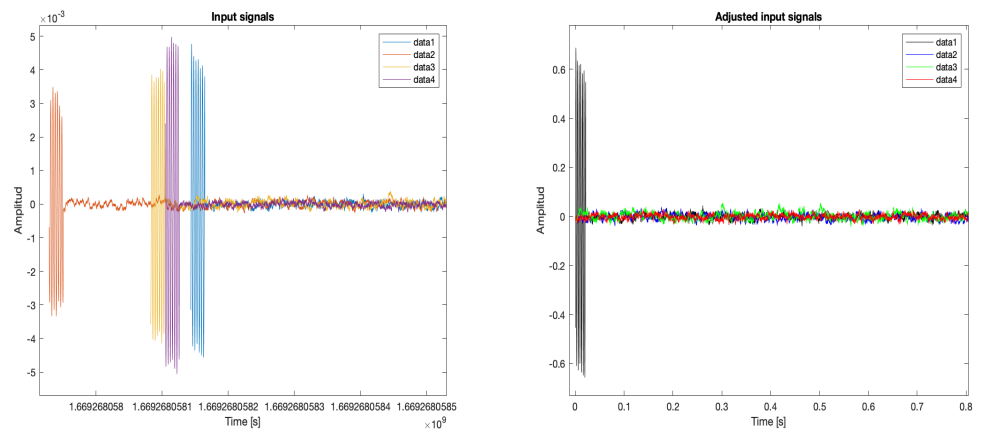


**Figure 23.** Experimental array for the evaluation of the methodologies.

Each of the nodes captured a signal with the MEMS microphones. Then, a post-processing adjustment was applied for the generation of the acoustic energy model, as follows:

1. **Retrieve data and timestamps adjustment.** Each node generates audio and time data files for the local recording methodologies: an audio file (in WAV format) for the captured signal and a text file with the timestamps generated during the recording.

Afterwards, these files are retrieved by a processing unit. The four captured signals are shown in Figure 24a, without any post-processing, with their own timestamps and amplitude values. As it will be seen later, when the local methodologies for signal capture are implemented, each node begins recording at a different time. Thus, the timestamps of each of the recordings of the nodes are adjusted to begin at the time where all the timestamps are closest. Each recording has associated the time of the local clock, so it is also necessary to establish a global frame of reference for the time. All signals are normalized and adjusted to begin at  $t = 0$  as Figure 24b shows.

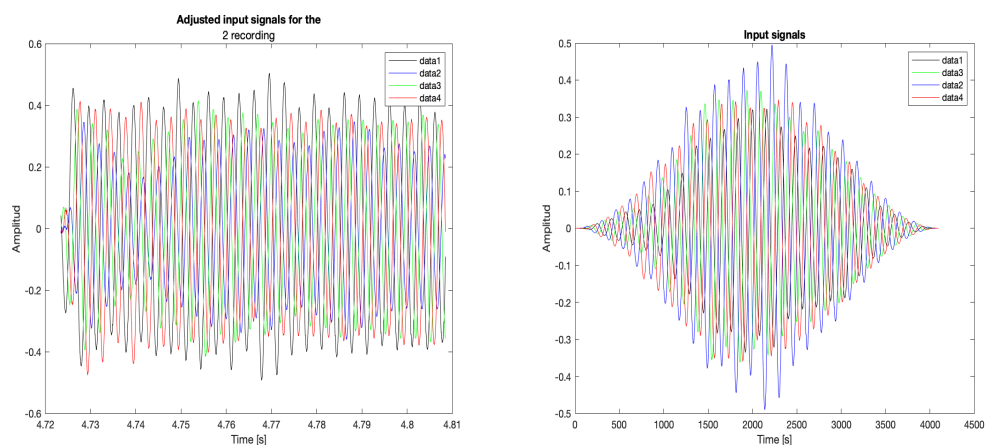


(a) Input signals without post-processing

(b) Input signals with adjusted timestamps

**Figure 24.** Retrieving and adjusting of audio data signals.

2. **Sample selection.** With the adjusted timestamps, a segment of the full signal was selected for the generation of the acoustic energy map. A Hann window was applied to this segment to avoid frequency bleed-over effects when the signals are processed by all the proposed beamformers shown in Figure 25a,b.



(a) Sample of the acoustic signal

(b) Signal with a Hann window applied

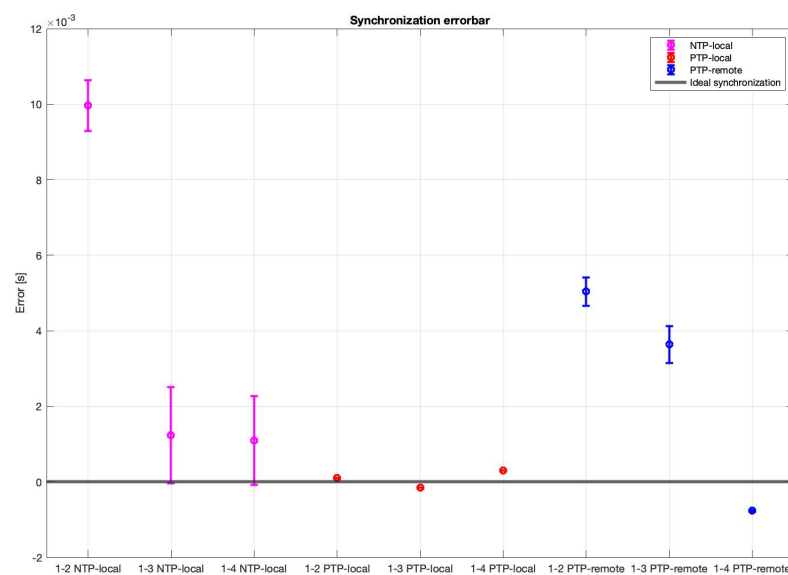
**Figure 25.** Post-processing of the acoustic input signals.

3. **Acoustic energy map generation.** The post-processed captured signals were processed by the beamformers to generate the acoustic energy map with the techniques described in Section 3. In this stage of the processing it is also possible to determine the position of the acoustic source by finding the maximum value on the acoustic energy map.

### 5.1. Synchronization Analysis

Given the location of the acoustic source, its signal should arrive to the each of the nodes of the WASN at approximately the same time. Thus, the analysis of synchronization can be based on the difference of the arrival times of the acoustic signal of the source to each node of the WASN. Additionally, it can also be based on the shape of the acoustic energy map generated by the signals captured by the WASN, compared to the ones obtained in the simulation without simulated synchronization errors, as well as to each other.

For the difference in arrival times analysis for each synchronization methodologies, the time of arrival of the signal of the source to each node was calculated and, by taking node 1 as reference, the differences of arrival times were calculated. In a perfectly synchronized WASN, with a perfectly placed acoustic source, the difference in arrival times between the reference node and all the other in the WASN should be zero. It is important to mention that synchronization errors may happen because of wireless interferences or mishaps in the TCP/IP hardware/software stack; the geometry of the array has no effect on the synchronization errors. Thus, each methodology will perform differently solely based on its characteristics and implementation. The results are shown in Figure 26.



**Figure 26.** Synchronization errors, as boxplots, for each node and each methodology.

These results show that the PTP-Local methodology has the lowest errors in synchronization, while the NTP-Local methodology has the largest errors in synchronization.

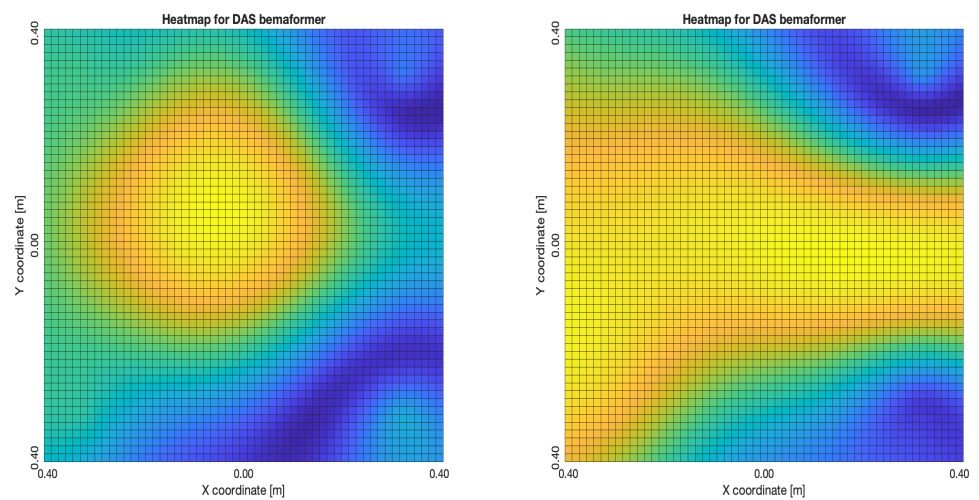
As for the map-shape analysis, in the following sub-sections (one for each synchronization methodology), two maps for each beamformer will be presented. The map on the left is representative of the best maps the beamformer provided given the synchronization methodology, and the map in the right representing the worst maps. The best and worst energy maps were obtained by capturing the signal of an acoustic energy source with a WASN as shown in Figures 22 and 23. The evaluation process involved running each methodology to capture 30 samples of 60 s each. With these signals, in post-processing, the beamformers were applied to generate an acoustic energy map from each sample. The map that was the most similar to the real acoustic environment is presented here as the best map. The worst map is the one that presented the most discrepancy with the real acoustic environment.

#### 5.1.1. NTP-Local Methodology

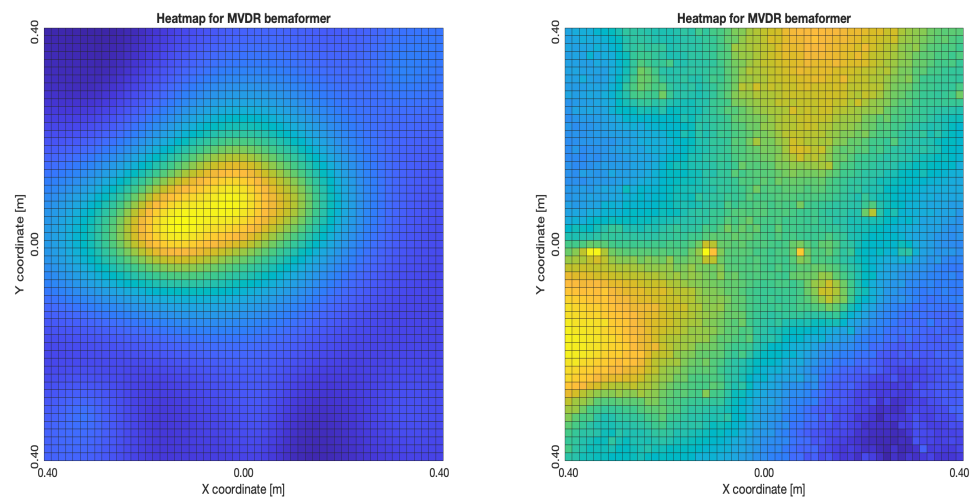
The NTP-Local methodology generated the worst acoustic energy maps, which is expected since it obtained the largest amount of synchronization errors. Thus, in the worst cases, the beamformers did not generate an acoustic energy map that represented the nature

of the acoustic environment. However, in the best cases where synchronization errors were low, the generated acoustic energy maps were able to locate a maximum of energy in the shape of an acoustic source.

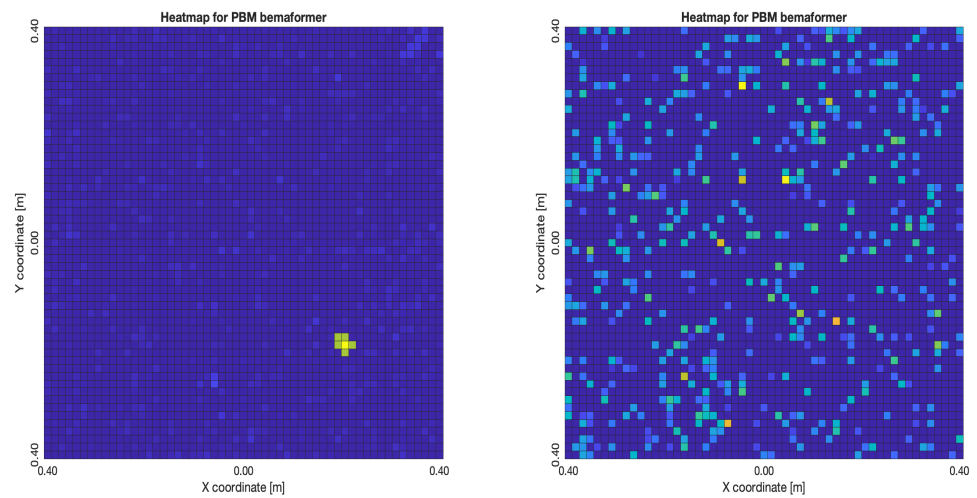
For DAS, Figure 27a,b, the shape of the acoustic energy map doesn't present a maximum value as a point source, instead it is distributed in a region (or lobe). For MVDR, Figure 28a,b, the energy map presents two regions where there seems to exist two acoustic sources, neither of which correspond to the real position of the acoustic source. For PBM, Figure 29a,b, the best case presented an almost point acoustic source, but not in the position of the acoustic source; in the worst case, several proof points presented a considerable amount of acoustic energy, none of which represented the acoustic source's position. Finally for SPR-PHAT, Figure 30a,b, in the best map, the most likely point for the localization of the source is close to the real position; in the worst map, there are many possible points where the source can be located.



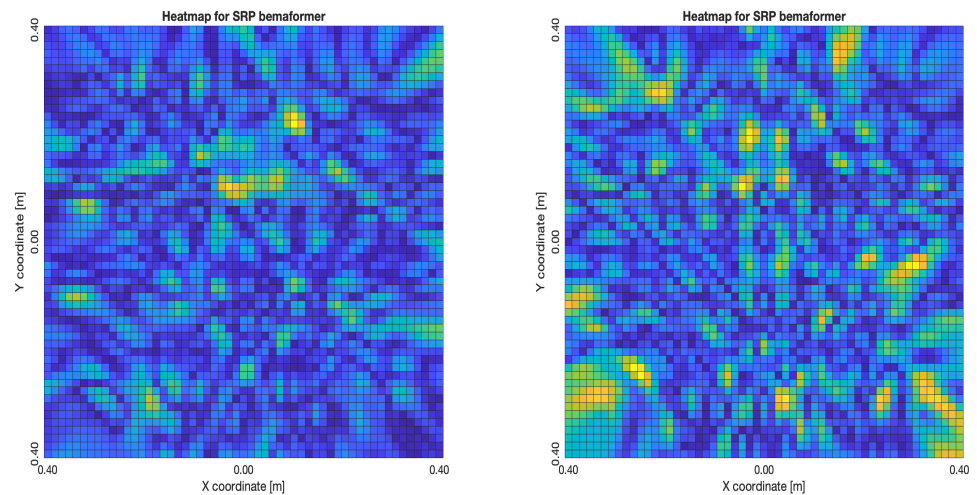
**Figure 27.** Acoustic energy maps obtained with DAS (left: best, right: worst).



**Figure 28.** Acoustic energy maps obtained with MVDR (left: best, right: worst).



**Figure 29.** Acoustic energy maps obtained with PBM (left: best, right: worst).



**Figure 30.** Acoustic energy maps obtained with SRP-PHAT (left: best, right: worst).

### 5.1.2. PTP-Local Methodology

The PTP-Local methodology shows an improvement over the NTP-Local methodology in terms of the shape and position of the lobes in the acoustic energy map. It was the most consistent of the methodologies, for all of the beamformers, locating the acoustic source near its real position.

For DAS, Figure 31a,b, even in the worst map, the acoustic energy map always presented a lobe with a localized maximum acoustic energy. For MVDR, Figure 32a,b, it presents smaller lobes (which are desirable), centered around a location that is close to the real location of the source. For PBM, Figure 33a,b, it presents the same problem as it did with the NTP-Local methodology: when there is a synchronization error, the beamformer finds energy in many proof points of the acoustic energy map. For SPR-PHAT, Figure 34a,b, it also performs similarly as with the NTP-Local methodology.

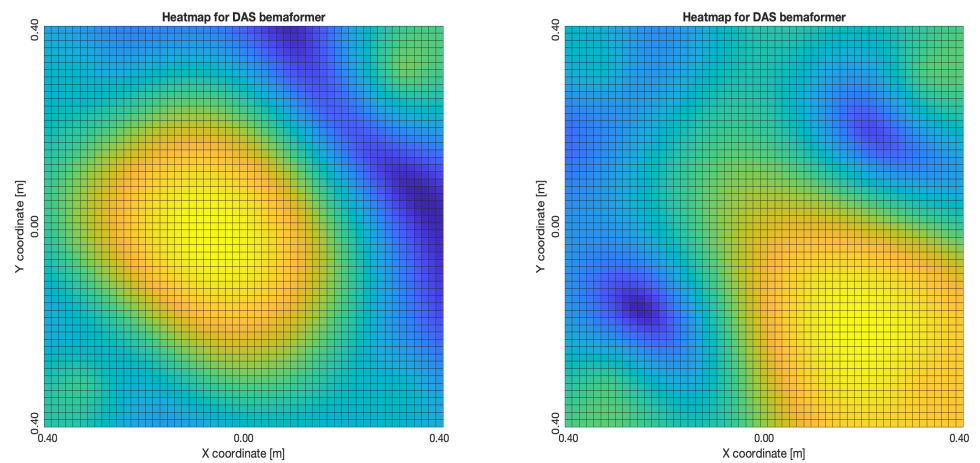


Figure 31. Acoustic energy maps obtained with DAS (left: best, right: worst).

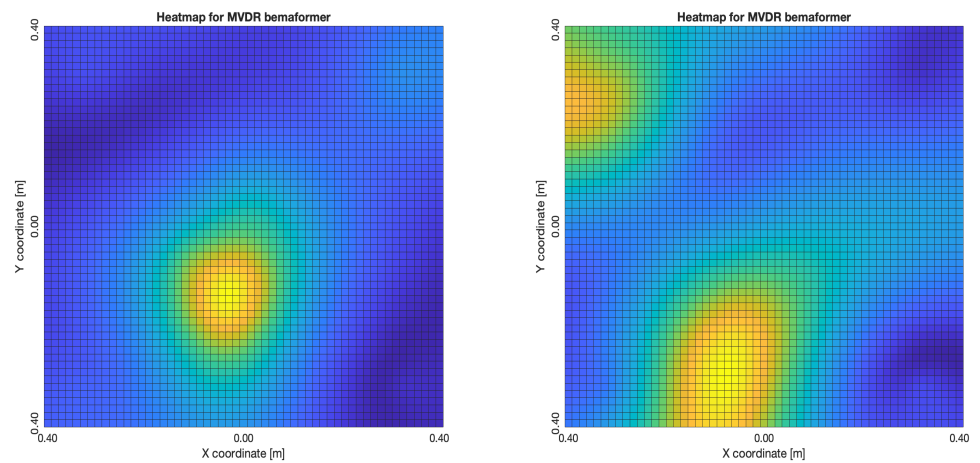


Figure 32. Acoustic energy maps obtained with MVDR (left: best, right: worst).

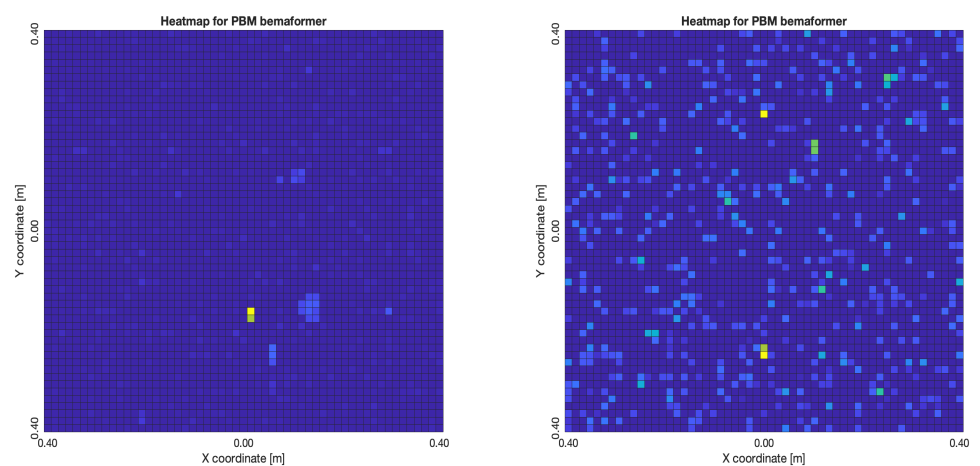
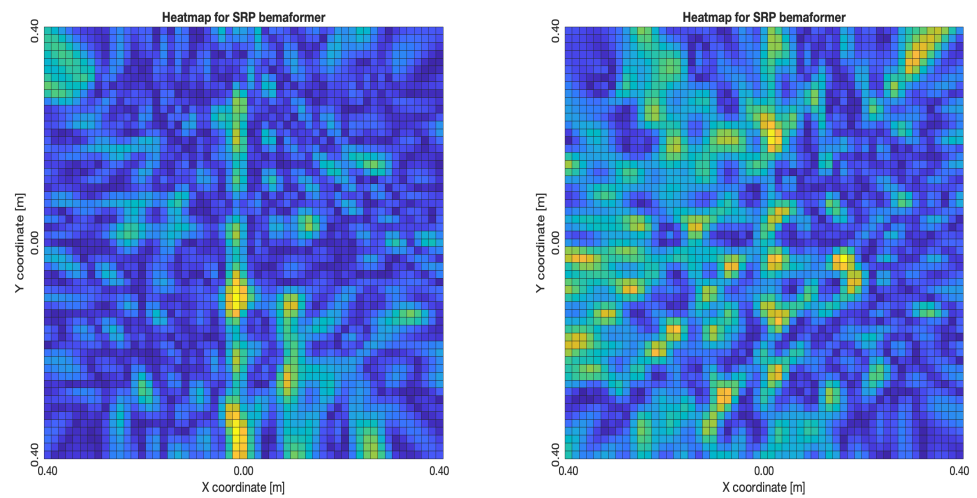


Figure 33. Acoustic energy maps obtained with PBM (left: best, right: worst).



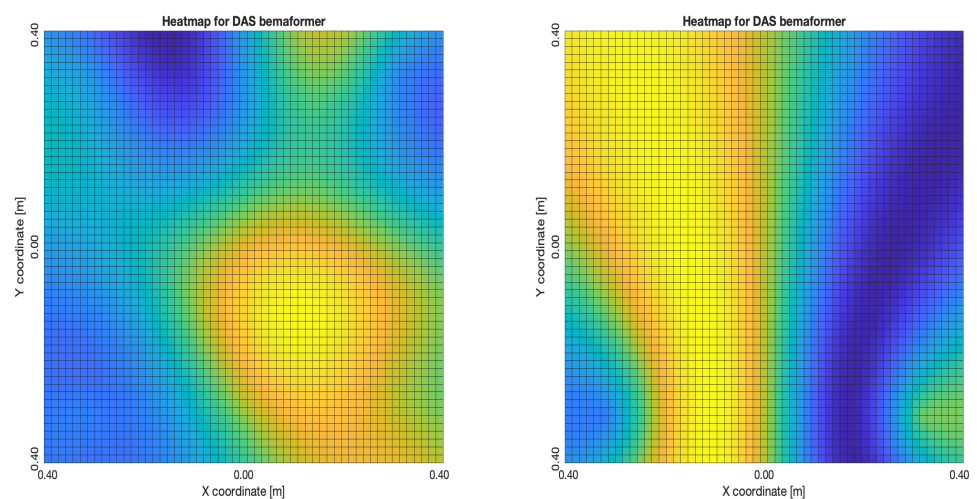


**Figure 34.** Acoustic energy maps obtained with SRP-PHAT (left: best, right: worst).

### 5.1.3. PTP-Remote Methodology

The PTP-Remote methodology shows that when the WASN is correctly synchronized, the acoustic energy maps obtained are close in precision to the ones obtained with the PTP-Local methodology. However, when the WASN presents synchronization errors, the maps obtained have a performance with the same problems as the ones seen in the NTP-Local methodology.

It is important to mention that, although the worst maps are similar to the worst maps obtained using NTP-Local methodology for DAS, Figure 35a,b, in the case of MVDR, Figure 36a,b and PBM, Figure 37a,b, their best maps are similarly localized compared to their respective best maps obtained using the PTP-Local methodology, except for SRP-PHAT, Figure 38a,b. Additionally, acquiring the audio data is more practical using the PTP-Remote methodology compared to the other two methodologies (which require manual extraction from each node).



**Figure 35.** Acoustic energy maps obtained with DAS (left: best, right: worst).

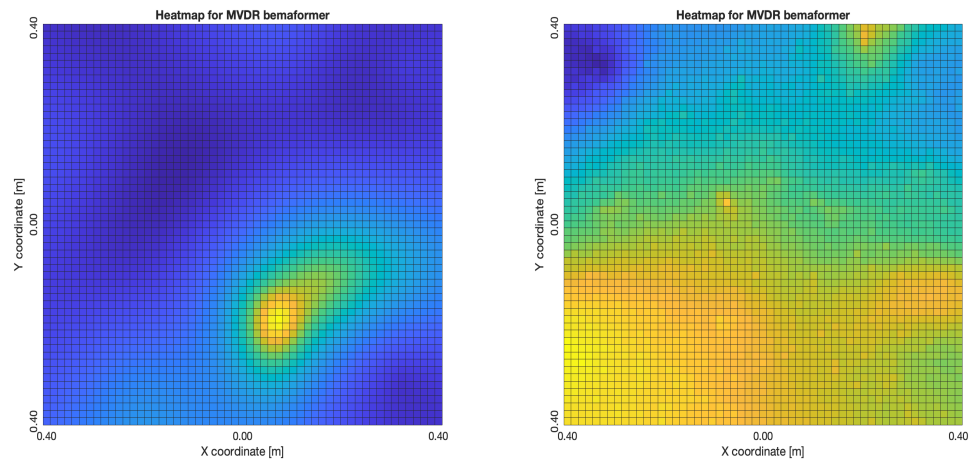


Figure 36. Acoustic energy maps obtained with MVDR (left: best, right: worst).

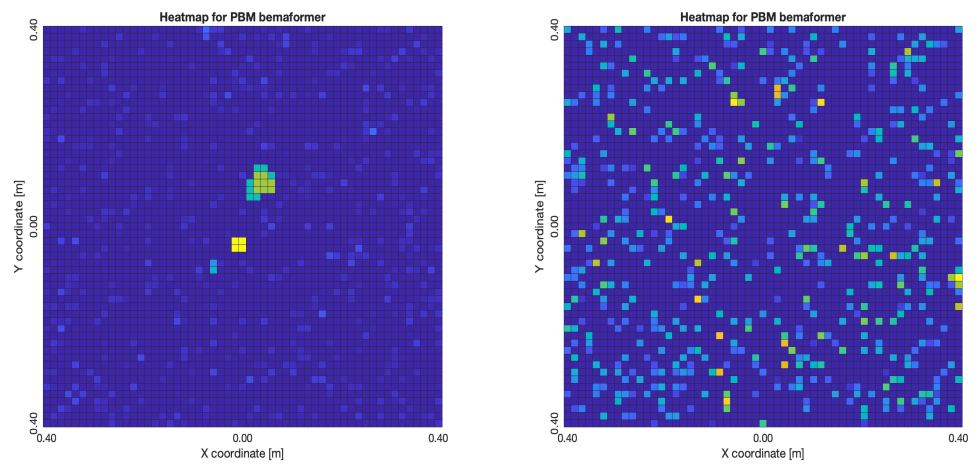


Figure 37. Acoustic energy maps obtained with PBM (left: best, right: worst).

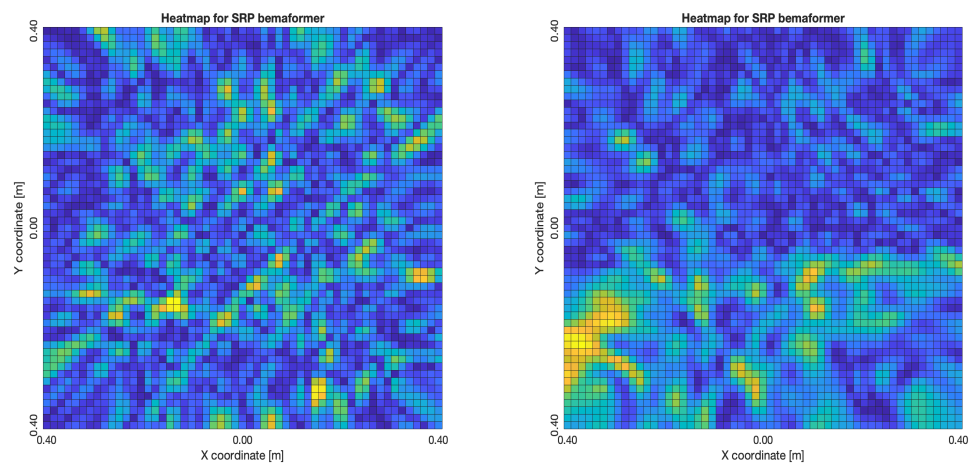


Figure 38. Acoustic energy maps obtained with SRP-PHAT (left: best, right: worst).

## 6. Discussion

In terms of precision in the localization of a simulated acoustic source, PBM tends to find high energy values only in the simulated position of the source, which is the ideal expected behavior. However, it is important to note that PBM requires a parameter (the phase difference threshold) to be established prior to the generation of the acoustic

energy map; the rest of the beamformers do not require any *a-priori* calibration. As for SRP-PHAT, although it does not technically create an acoustic energy map, the location with the highest steered power matches the location of the simulated acoustic source even when synchronization error are inserted. And as for DAS and MVDR, although they present larger localization errors when synchronization errors are simulated, the relationship between these errors is near-linear. This insight can be very valuable since this relationship can be used to establish a type of precision threshold given an expected level of synchronization errors which, in turn, can be estimated from the nature of a given application scenario. Thus, from the findings in the experimental analysis, a user can know beforehand what type of localization errors are bound to happen using a WASN to create an acoustic energy map.

Interestingly, DAS and MVDR provide very similarly shaped acoustic energy maps, and share similar tendencies of shifts of the location of the lobe above the simulated acoustic source, displacing it away from the sync-error node up to a simulated synchronization error of 60 samples. This can be explained by the geometrical characteristics of the array and its relationship to the wavelength of the signal. The analysis of the nature of the beamformers and its implications to the acoustic energy map generation are left for further studies.

The experimental analysis show that an acoustic energy map can be generated using any of the described synchronization methodologies. The NTP-Local and the PTP-Remote methodologies provide the largest synchronization errors, while the PTP-Local methodology presented the lowest synchronization errors, as well as the lowest variance. Additionally, when an acoustic source was located in the center of the array, where the difference in arrival times between nodes should close to 0, the PTP-Local methodology provided the lowest values and variance. In the simulation analysis, 60 samples was found to be a type of upper threshold of synchronization error; with higher values, a big “jump” in localization errors was observed with DAS and MVDR. A synchronization error of 60 samples, sampled at 48 kHz, is equivalent to 1.25 ms. Connecting this with the observed mean difference of arrival times to the reference capture node ( $mean_{1-2} = 0.0958$  ms,  $mean_{1-3} = -0.1555$  ms, and  $mean_{1-4} = 0.3035$  ms) when using the PTP-Local methodology, it can be deduced that it provides synchronization errors well below the aforementioned threshold and, thus, well suited to calculate acoustic energy maps with any of the described beamformers.

Having calculated an acoustic energy map using all the described beamformers along with the described synchronization methodologies, the position of their maximum value was found and considered as the estimated position of the acoustic source. The localization errors in the experimental scenario are shown in Table 3. The statistical measures presented are the mean and the Median Absolute Deviation (MAD) [26], which is a distribution-free measure of dispersion. These quantities are used as a comparison factor between methodologies.

**Table 3.** Localization error analysis for each methodology and each beamformer.

Methodology	Localization Error [m]	DAS	MVDR	SRP-PHAT	PBM
NTP-Local	Mean	0.3429	0.3224	0.2814	0.3198
	MAD	0.1013	0.1182	0.0823	0.1359
PTP-Local	Mean	0.2585	0.2533	0.3046	0.2151
	MAD	0.1069	0.1126	0.0949	0.1247
PTP-Remote	Mean	0.2461	0.2922	0.3067	0.2803
	MAD	0.0930	0.1005	0.1015	0.1021

As it can be seen, these localization errors are considerably larger than the ones found in the simulation analysis. This is expected since a real acoustic scenario has more variables that were not simulated (reverberation, other interferences, etc.). Additionally, the simulated synchronization errors were inserted into just one node (the sync-error node), while in the experiment scenario all nodes are affected by such issues. Finally, the speaker

used in the experiments is not a point acoustic source, as was the case in the simulation. However, given all these factors, it is important to note that since these acoustic maps were generated from signals captured using a WASN, the placement of the nodes are not limited to any tethering, thus, are easily positioned anywhere in the acoustic scenario, such as in a living room. Considering this, the observed localization errors are reasonably low, with PBM using the PTP-Local methodology providing the lowest mean error. Interestingly, SRP-PHAT provides the lowest variance (also using the PTP-Local methodology), however, it provides the largest mean error, thus, it may not be as suitable as PBM.

## 7. Conclusions

In this paper, three different synchronization methodologies were used to generate an acoustic energy map applying four popular beamforming techniques: Delay-and-Sum (DAS), Minimum Variance Distortionless Response (MVDR), Steered-Response Power with Phase Transform (SRP-PHAT), and Phase-based Binary Masking (PBM).

The synchronization methodologies were conformed by: (1) a WASN capture protocol (either local or remote) used to capture a multi-channel audio recording from the environment; and (2) a synchronization protocol, which could be either Network Time Protocol (NTP) or Precision Time Protocol (PTP). The beamforming techniques were then applied to the captured signals, to measure the energy coming from a given series of proof points (arranged in a grid) to generate the acoustic energy map.

A simulation was carried to, first, validate this proof of concept, and to characterize the behavior each beamforming technique against a simulated synchronization error.

Furthermore, an experimental scenario was used to test the synchronization methodologies. The devices used for capture nodes in the WASN were conformed by a Raspberry 4B+ as a processing unit and a MEMS microphone breakout-board. This configuration was chosen because of its low cost, accessibility and straightforward implementation.

With the results obtained by simulation and experiments, the following conclusions can be stated:

- In the simulation, SRP-PHAT and PBM showed higher robustness against synchronization errors. DAS and MVDR were more sensitive, but a near-linear relationship between localization errors and synchronization errors was found which can be used for ease of expectation of the user.
- In the experimental scenario, from a subjective point of view, MVDR and PBM provided an acoustic energy map close to what was expected, with either a lobe or point over the acoustic source location; SRP-PHAT and DAS provided unexpected maps. While PBM requires a parameter to be calibrated beforehand, it generates a more precise acoustic energy map. And, while MVDR generates a less precise acoustic energy map, it is robust against environmental elements (such as noise and interferences) and does not require any a-priori parameter calibration.
- The PTP-Local and PTP-Remote methodologies are the most suitable for acoustic energy mapping using a WASN. The PTP-Local methodology has the best performance in terms of synchronization errors, but the process to capture, retrieve and process the data collected is more tedious than the PTP-Remote methodology, since it requires the user to manually extract the recording from the capture node. The PTP-Remote methodology is less reliable in terms of synchronization (with a higher mean error value). However, the captured signals can still be used to provide a close-to-precise acoustic energy map (with either PBM or MVDR), while being easier to implement (since no synchronization agent is required to run locally in the capture nodes), and the recordings are directly streamed to the server (no manual extraction necessary).

For future work, other synchronization protocols are to be explored. Additionally, other characterizations are to be carried out, such as: robustness against noise and reverberation, processing time, implementation complexity, etc. Finally, the effect of different windows (other than Hann, used in this work) to reduce time-domain discontinuities is to be characterized.

**Author Contributions:** Conceptualization, E.E.G.-U. and C.R.; methodology, E.E.G.-U., C.R. and P.E.M.-M.; Investigation, E.E.G.-U. and P.E.M.-M.; resources, P.E.M.-M. and C.R.; supervision, P.E.M.-M.; writing—original draft, E.E.G.-U. and C.R.; writing—review and editing, E.E.G.-U. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research and the APC for this article was funded by PAPIIT-UNAM IA100222 and was supported by UNAM-PAPIIT IN105623.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Acoustic mapping code for simulated signals and real signals captured by a WASN of 4 nodes: <https://github.com/emilianounzueta/acoustic-energy-mapping>.

**Acknowledgments:** The authors would like to thank CONACYT for providing support through its National Scholarship Program.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

WASN	Wireless Acoustic Sensor Network
NTP	Network Time Protocol
PTP	Precision Time Protocol
MEMS	Micro-electret microphone
SNR	Signal-to-noise ratio
DAS	Delay and Sum
MVDR	Minimum Variance Distortionless Response
SRP-PHAT	Steered-Response Power Phase Transform
PBM	Phase-based Binary Masking

## References

- Arce, P.; Salvo, D.; Piñero, G.; Gonzalez, A. FIWARE based low-cost wireless acoustic sensor network for monitoring and classification of urban soundscape. *Comput. Netw.* **2021**, *196*, 108199. [[CrossRef](#)]
- Verreycken, E.; Simon, R.; Quirk-Royal, B.; Daems, W.; Barber, J.; Steckel, J. Bio-acoustic tracking and localization using heterogeneous, scalable microphone arrays. *Commun. Biol.* **2021**, *4*, 1275. [[CrossRef](#)] [[PubMed](#)]
- Xiong, H.; Agcayazi, T.; Latif, T.; Bozkurt, A.; Sichertiu, M.L. Towards acoustic localization for biobotic sensor networks. In Proceedings of the 2017 IEEE SENSORS, Glasgow, UK, 29 October–1 November 2017; pp. 1–3.
- Whitmire, E.; Latif, T.; Bozkurt, A. Acoustic sensors for biobotic search and rescue. In Proceedings of the SENSORS, 2014 IEEE, Valencia, Spain, 2–5 November 2014; pp. 2195–2198.
- Hollosi, D.; Schröder, J.; Goetze, S.; Appell, J.-E. Voice activity detection driven acoustic event classification for monitoring in smart homes. In Proceedings of the 2010 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL 2010), Rome, Italy, 7–10 November 2010; pp. 1–5.
- Selmic, R.; Phoha, V.; Serwadda, A. *Wireless Sensor Networks. Security, Coverage, and Localization*; Springer: New York, NY, USA, 2016; pp. 1–16, 37–53.
- Ginovart-Panisello, G.J.; Vidaña-Vila, E.; Caro-Via, S.; Martínez-Suquía, C.; Freixes, M.; Alsina-Pagès, R.M. Low-Cost WASN for Real-Time Soundmap Generation. *Eng. Proc.* **2021**, *6*, 57. [[CrossRef](#)]
- Pleshkova, S.; Panchev, K.; Bekyarski, A. Developing a Functional Scheme of an IoT Based Module to an Acoustic Sensor Network. In Proceedings of the 2021 IV International Conference on High Technology for Sustainable Development (HiTech), Sofia, Bulgaria, 7–8 October 2021; pp. 1–4.
- Cobos, M.; Antonacci, F.; Alexandridis, A.; Mouchtaris, A.; Lee, B. A Survey of Sound Source Localization Methods in Wireless Acoustic Sensor Networks. *Wirel. Commun. Mob. Comput.* **2017**, *2017*, 3956282. [[CrossRef](#)]
- Liu, Z.; Zhang, Z.; He, L.; Chou, P. Energy-based sound source localization and gain normalization for ad hoc microphone arrays. In Proceedings of the 2007 IEEE International Conference on Acoustics, Speech and Signal Processing—ICASSP '07, Honolulu, HI, USA, 15–20 April 2007.
- de Groot, T.H. Localization and Classification Using an Acoustic Sensor Network. Master's Thesis, Delft University of Technology, Delft, The Netherlands, 2010.
- Griffin, A.; Alexandridis, A.; Pavlidis, D.; Mastorakis, Y.; Mouchtaris, A. Localizing multiple audio sources in a wireless acoustic sensor network. *Signal Process.* **2015**, *107*, 54–67. [[CrossRef](#)]

13. Wang, R.; Wang, Y.; Han, C.; Gong, Y.; Wang, L. Robust Adaptive Beamforming Based on Interference Covariance Matrix Reconstruction and Steering Vector Estimation. In Proceedings of the 2021 IEEE International Conference on Signal Processing, Communications and Computing, Xi'an, China, 17–19 August 2021 .
14. Unnikrishna, S. *Array Signal Processing*; Springer: New York, NY, USA, 1989; pp. 8–20.
15. Benesty, J.; Chen, J.; Huang, Y. *Microphone Array Signal Processing*; Springer Topics in Signal Processing: Berlin, Germany, 2008; pp. 39–65.
16. Yang, Y. Time Synchronization in Wireless Sensor Networks: A Survey. Bachelor's Thesis, Mittuniversitetet, Sundsvall, Sweden, 2012.
17. Jingchao, W.; Ruohan, Z.; Weiwen, G. Time Synchronization in Networks: A Survey. In Proceedings of the 2nd International Conference on Control and Computer Vision, Jeju, Republic of Korea, 15–18 June 2019; pp. 121–126.
18. Neagoie, T.; Cristea, V.; Banica, L. NTP versus PTP in Computer Networks Clock Synchronization. In Proceedings of the 2006 IEEE International Symposium on Industrial Electronics, Montreal, QC, Canada, 9–13 July 2006; pp. 317–362.
19. Ferrari, P.; Flammini, A.; Rinaldi, S.; Bondavalli, A.; Brancati, F. Experimental Characterization of Uncertainty Sources in a Software-Only Synchronization System. *IEEE Trans. Instrum. Meas.* **2012**, *61*, 1512–1521. [[CrossRef](#)]
20. Perrot, V.; Polichetti, M.; Varray, F.; Garcia, D. So you think you can DAS? A viewpoint on delay-and-sum beamforming. *Ultrasonics* **2021**, *111*, 106309. [[CrossRef](#)] [[PubMed](#)]
21. Capon, J. High-resolution frequency-wavenumber spectrum analysis. *Proc. IEEE* **1969**, *57*, 1408–1418. [[CrossRef](#)]
22. Diaz-Guerra, D.; Beltran, J.R. Direction of Arrival Estimation with Microphone Arrays Using SRP-PHAT and Neural Networks. In Proceedings of the 2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM), Sheffield, UK, 8–11 July 2018; pp. 617–621.
23. Zhuo, D.-B.; Cao, H. Fast Sound Source Localization Based on SRP-PHAT Using Density Peaks Clustering. *Appl. Sci.* **2021**, *11*, 445. [[CrossRef](#)]
24. He, L.; Zhou, Y.; Liu, H. Phase Time-Frequency Masking Based Speech Enhancement Algorithm Using Circular Microphone Array. In Proceedings of the 2019 IEEE International Conference on Multimedia and Expo (ICME), Shanghai, China, 8–12 July 2019; pp. 808–813.
25. Rascon, C. A Corpus-Based Evaluation of Beamforming Techniques and Phase-Based Frequency Masking. *Sensors* **2021**, *21*, 5005. [[CrossRef](#)]
26. Sachs, L. *Applied Statistics: A Handbook of Techniques*; Springer: New York, NY, USA, 1984; p. 253.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.