Engineering Proceedings

MDPI

Proceedings

Acoustic Positioning System for 3D Localization of Sound Sources Based on the Time of Arrival of a Signal for a Low-Cost System [†]

Dídac D. Tortosa [1,*,‡], Iván Herrero-Durá [2,‡] and Jorge E. Otero [3,‡]

[1] Institut d'Investigació per a la Gestió Integrada de Zones Costaneres (IGIC), Universitat Politècnica de València (UPV)
[2] KFB Acoustics Sp. z o. o. Mydlana 7, 51-502 Wrocław, Poland; i.herrero-dura@kfb-acoustics.com
[3] Instituto Tecnoacústico de Cantabria (ITC), 39010 Santander, Spain; jotero@ceronoise.com
[*] Correspondence: didieit@upv.es
[†] Presented at the 8th International Electronic Conference on Sensors and Applications, 1–15 November 2021; Available online: https://ecsa-8.sciforum.net.
[‡] These authors contributed equally to this work.

Abstract: The localization of sound sources has received increasing interest over the last decades, given its wide range of applications. The triangulation method using the Time of Arrival (ToA) of a signal has shown to be useful and easy-to-use and, at the same time, provides accurate results. In this work, the acoustic trilateration method is applied in experimental measures to study and demonstrate its precision in air. Firstly, the method is tested in an anechoic chamber (low reverberating environment) demonstrating its functionality and accuracy. The next step has been the application of the method by using a low-cost system to demonstrate how a non-anechoic environment affects the accuracy of the localization. The detection of the received signal is implemented using a cross-correlation method in the time domain for both cases. Furthermore, the influence of the number and positions of the receiver that are used for this process in the accuracy of the results is also studied.

Keywords: acoustic positioning system; source localization; time of arrival; low-cost system

check for updates

Citation: Tortosa, D.D.; Herrero-Durá, I.; Otero, J.E. Acoustic Positioning System for 3D Localization of Sound Sources Based on the Time of Arrival of a Signal for a Low-Cost System. *Eng. Proc.* **2021**, *1*, 0. https://doi.org/

Academic Editor: Dídac Tortosa

Received:
Accepted:
Published:

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 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

Sound localization can be defined as the process of identifying the spatial coordinates of a sound source (emitter) based on the sound signal received by a microphone (receiver) array [1]. Source localization using sensor arrays has been one of the central problems in radar, sonar, navigation, geophysics and acoustic tracking [2]. The localization of a source in space has also received an increasing interest over the last years, as many new applications can obtain substantial benefits from the knowledge of the spatial position of an emitter by means of knowing the characteristics of the signal [1].

Many audio processing applications include animal detection in the wild forest, speech enhancement, tracking of sound sources, maritime applications, localization of brain tumors, teleconferencing and detection of astroparticles, among others. In this context, the development of new technologies allows a higher accuracy in the process related to the time parameters associated to the localization process [3]. Commonly, most algorithms and techniques related to the localization process involve an estimation of Time Difference of Arrival, *TDoA*, at a set of microphones' positions from which one of them can drive information about the spatial position of the source [4]. The processing of this information is usually based on the computation of the Generalized Cross-Correlation, *GCC*, whereby the Time of Arrival, *ToA*, (defined as the time that an emitted signal take to be detected by a receiver) can be obtained [5]. The same procedure can be used for the inverse purpose, this is localizing a receiver by knowing the localization of different emitters [6].

In this work, the localization of an acoustic source in a low reverberating indoor environment by means of a low-cost system is presented. The main peculiarity of the proposed method is that the Time of Emission, *ToE*, (defined in the following Section) must be a known parameter, which implies controlling the emission to be registered. This is the case of an acoustic active system, commonly used to search characterized sources.

The structure of this paper is as follows: Section 2 presents the case of study and describes the methodology used for the definition of the emitted signal, its detection in recorded signals and the localization of the emitter. In Section 3, the set-up used during the experimental measurements, as well as the results obtained by means of these measurements, are exhibited. Finally, the conclusions and important remarks of this work are presented in Section 4.

2. Methodology

The first step in the study is defining the signal that is used to localize the sound source. This signal is emitted by source (emitter), propagates through the medium (which in this case is air) and reaches the receivers. It is important to note that, assuming an homogeneous medium in which the speed propagation of the sound waves, c, is constant, the signal takes a different time to reach each receiver placed in different position due to the different distances to be travelled by the sound waves.

The Time of Flight, *ToF*, of the signal is a crucial parameter involved in localizing the source, as it represents the time that the signal takes to propagate between emitter and receivers. The Time of Emission, *ToE*, is defined as the time instant in which the signal starts being emitted. This parameter can be independent of the data acquisition system, in which case it must be considered as unknown (passive acoustic system). The Time of Arrival, *ToA*, is referred to as the time that the signal takes to be detected by the receiver. From these two last parameters, one can straightforward obtain the *ToF* as the difference between *ToA* and *ToE* ($ToF = ToA - ToE$).

The distance between the emitter and the receiver, d_{ER}, can be defined as a function of the *ToF* and the speed of sound as $d_{ER} = ToF \cdot c$.

In the experiments presented here, the *ToE* is controlled by the signal acquisition system and, consequently, the *ToA* for each receiver is the unknown parameter that must be found.

In this work, the localization of an emitter, E, whose position is theoretically unknown is carried out by means of the signal detected by a receiver, R, placed at different positions. This receiver is moved to different measurement positions in order to simulate an array of receivers. It is also important to note that the method requires the use of a minimum of three receivers' positions for a proper localization of the source.

Different signals were tested in the experimental stage (see Appendix A), and it was checked that sweep and MLS are more suitable for the *GCC* detection [7]. The linear sine sweep signal, W_s, is defined as

$$W_s(t) = \sin\left[2\pi\left(\frac{|f_2 - f_1|}{T} \cdot t + f_1\right)t\right], \tag{1}$$

where T is the duration of the signal, f_1 the initial frequency of the sweep, f_2 the ending frequency of the sweep and t the instant of the time. Note that if the value of f_1 is lower than that of f_2, the resulting sweep signal is ascendant (in the opposite case, it is a descendant sine sweep).

For the low-cost system in reverberating environment case presented in this work, the sine sweep is generated in the frequency range between 2 kHz and 12 kHz and with a duration of 300 ms, both characteristics corresponding to audible signals. A sampling frequency of 44.1 kHz is considered. Details on the emitted, received and correlation signal can be observed in Figure 1.

To test the detection, a simple virtual measure is created: The distance between emitter and receiver, d_{ER}, is 43 cm, and the propagation speed of sound in the surrounding medium

($c = 343.2$ m/s at 20 °C). With this information one can directly obtain the *ToF* of the signal as $ToF = d_{ER}/c$. A *ToE* of 400 ms is considered (as it can be observed in Figure 1c), and the ToA is 401.2573 ms). The Signal-to-Noise Ratio (SNR) in reception is 40 dB and the recording time is 1 s (see Figure 1c,d).

It is important to note that, for the sake of simplicity, in the case presented here the decay of amplitude of the propagating signal with the distance, as well as the variation of the absorption due to the propagation of the sound waves in the medium depending on the frequency, are not considered.

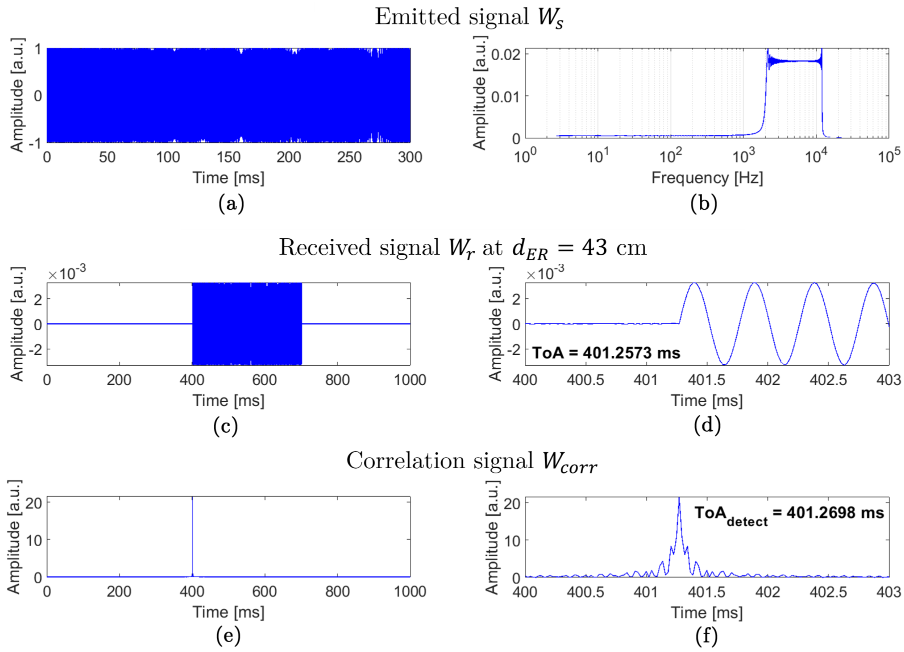


Figure 1. (**a**) Emitted signal in time domain. (**b**) Emitted signal in frequency domain. (**c**) Example of received signal (recorded during 1 s) by a receiver placed at a distance of 43 cm from the emitter with a *ToE* of 400 ms. (**d**) Zoom to the *ToA* of the received signal. (**e**) Resulting correlated signal between emitted and received signals. (**f**) Zoom to the detected *ToA* in the correlated signal.

2.1. Detection of the Signal

In this Section, the detection process of the emitted signal is presented. A high resolution in the *ToA* is crucial for an accurate localization process. Since the emitted signal is a sweep, the *GCC* is an effective method for its automatic. This method is based on correlating the emitted and received signal and finding the *ToA* from the peak in a given instant of time. The results of this process can be observed in Figure 1e,f. As it can be extracted from the difference between the detected *ToA* from the correlation signal (ToA_{detect}, in Figure 1f) and that of the received signal (Figure 1d), there is a delay of 12.5 µs. Considering the propagation speed of sound in the medium, this delay implies a difference of less than 0.5 cm in the detection of the signal, which can be assumed as sufficiently low.

Some other methods can be used with the purpose of finding the *ToA*, such as using a threshold in the time domain for the amplitude received or applying the accumulated frequency of the received signal and detecting harsh changes in the slope [8].

The cross-correlation signal, W_{corr}, obtained by computing the *GCC* between the emitted signal, W_s, and the received signal, W_r, is expressed as a function of the power spectral density $G_{W_sW_r}$ as shown below.

$$W_{corr}^{GCC}(t) = \int_{-\infty}^{+\infty} W_s(f)W_r^*(f)G_{W_sW_r}(f)e^{i2\pi ft}\mathrm{d}f = \varphi^{GCC}(f)G_{W_sW_r}(f)e^{i2\pi ft}\mathrm{d}f, \quad (2)$$

where * indicates a complex conjugated and $\varphi^{GCC}(f)$ is a frequency-dependent weight function. Due to finite observations, it is only possible to obtain an estimation of $G_{W_sW_r}(f)$ [9]. Therefore, to obtain the *TDoA*, the following expression will be used [10]:

$$\hat{W}_{corr}^{GCC1}(t) = \int_{-\infty}^{+\infty} \varphi^{GCC}(f)\hat{G}_{W_sW_r}(f)e^{i2\pi ft}\mathrm{df}, \tag{3}$$

where $\hat{G}_{W_sW_r}(f)$ is the obtained estimation of $G_{W_sW_r}(f)$. For each pair of sensors, the *ToA* is taken as the time delay that maximizes the cross-correlation between the filtered signals of both sensors, that is: $\hat{\tau}_{ij}^{GCC} = \arg\left(\max_t\{\hat{W}_{corr}^{GCC1}(t)\}\right)$.

2.2. Localization of the Emitter

A general model for three-dimensional (3-D) estimation of an emitter using i receivers is developed in this Section. To obtain the location of the source, the first step is knowing the spatial position (x_i,y_i,z_i) in a Cartesian coordinate system of a given number of receivers. Let the position of the emitter to be located be (x_E, y_E, and z_E), the distance between the emitter and the i-th receiver, d_{ER_i}, is defined as:

$$d_{ER_i} = \sqrt{(x_i - x_E)^2 + (y_i - y_E)^2 + (z_i - z_E)^2}. \tag{4}$$

Based on Equation (4), it is possible to create a resolvable nonlinear equation system with 3 unknowns (x_E, y_E and z_E) and i equations. Thus, it is necessary to have a minimum of 3 receivers to solve the system. To solve the system of equations, the method analyzes the difference in the distance between the i-th receiver and the first receiver (d_{i1}), which is given by:

$$\begin{aligned} d_{ER_{i1}} &= d_{ER_i} - d_{ER_1} = \\ &= \sqrt{(x_i - x_E)^2 + (y_i - y_E)^2 + (z_i - z_E)^2} - \sqrt{(x_1 - x_E)^2 + (y_1 - y_E)^2 + (z_1 - z_E)^2}, \end{aligned} \tag{5}$$

where $d_{ER_{i1}}$ is the distance between the first receiver and the emitter.

To obtain the position of the source, the system of equations can be written considering a system of m equations and n unknowns $f_m(x_1, x_2, x_3, ..., x_n) = 0$. This system can be written in vector form as $f(x) = 0$ where f is a vector of m dimensions and x is a vector of n dimensions. To solve this system of equations, it is necessary find out a vector x such the function $f(x)$ equals the null vector. In this case, the problem is solved by means of an algorithm for calculating nonlinear equations systems using the MATLAB tool $fsolve$. This method is based on the Newton-Raphson method for which an initial Position of reference, $Posref$, is proposed to start the calculation process. Since the positions of the receiving microphones are known for all the experimental measurements, the mean value of the receivers' positions (in the middle between them) has been taken as the reference position. To ensure the convergence of the results, the input parameters of the function have been defined considering a maximum of 4000 iterations, a computational error for the tolerance, Inf maximum function evaluations and not using a Jacobian solution.

3. Experimental Framework

3.1. Experimental Set-Up

In the previous Section, the theoretical approach for the localization of a sound source by solving a system of nonlinear equations has been detailed. This method is used to localize a sound source in two different environments: An anechoic chamber and a low reverberating room. This approach in based on the case in which a given emitter, whose position is unknown, wants to be localized by means of the signal received in certain locations that are known. The influence of the number of positions for the receiver in the resulting localization is also studied by comparing the results using a different number of positions (modifying the number of microphones used in the measurements).

A sound source (emitter), controlled by a sound acquisition system and placed at a given position, is considered to be emitting a signal containing the frequencies within the range of study that is generated by a sound card. This signal is registered by a receiver placed at different known positions. It is important to make sure that both, emitter and receiver, have a flat frequency response in the studied bandwidth.

(a) Tests in An Anechoic Chamber

To validate the acoustic positioning system for 3D localization, previously described for a low-cost system, the same method in an ideal environment with low reverberating conditions has been applied. The approach consists of testing different signals and study the results to select the best configuration to be reproduced afterwards with the low-cost system . In this occasion a *Focusrite Scarlett 18i20 v3* audio card was used, connected to 6 microphones *Behringer ECM8000* (array distributed among 3.2 and 4.1 m distance from the emitter) and to a *Genelec 8030A* source. This source is composed of a woofer speaker for low and mid frequencies (below 2 kHz) and a tweeter for high frequencies (above 2 kHz). In the case that the emitted signal contains only high frequency components the emitter, E, to be searched is the center of the tweeter position. If, on the contrary, the signal is in all audible spectrum (e.g., MLS) the E position to be found corresponds to the center position between the tweeter and the woofer.

(b) Tests in a Low-Reverberating Environment: A Low-Cost System

For the low-cost system, the equipment used has been a loudspeaker *Genius SP-U115*, a microphone *Behringer ECM8000*, and a sound card *Focusrite Scarlett Solo*. Note that all the audio systems have an intrinsic latency that must be taken into account. In this experiment, a buffer size of 256 samples is used, since the lower the value of this number of samples, the lower latency in the system. The second channel of the sound card is used as a reference in order to know the latency of each recording, while first channel is used for recording the signal provided by the microphone. After the signal has been registered, the cross-correlation is calculated for both channels. In this configuration, the first channel gives *ToA* and the second channel controls the *ToE*. In the case in which multiple receiver positions are used. they are considered as an array of microphones. This can be demonstrated by the fact that using 4 or more microphones in an anechoic environment does not significantly improve the precision of the detection (see Appendix A).

The goal of this measurements is to localize the emitter with error that, at the most, corresponds to the diameter of the sound source that is used. In this case, this diameter is 2″ (5.08 cm).

The coordinates of the positions of the emitter and the receivers are measured, all of them in a volume of $70\ \text{cm}^3$. It is worth noting here that, even though the position of the emitter is theoretically unknown, the position at which it is placed in the experiment must be known in order to validate the applied model by calculating the difference found between the proposed method and the real position. The positions of the receiver are chosen in order to get a representative sampling in space of the area of study. Given the nature of the emitter (loudspeaker), it has an associated directivity that must be considered in order to avoid measuring in the so-called 'areas of shadow' of the loudspeaker, in which no acoustic pressure is radiated.

3.2. Results

Figure 2 shows the positions of the emitter (red circle) and the receiver (blue circles). The reference positions plotted in the figure is the one considered by the algorithm as a starting point to look for the position of the emitter and is chosen to be the midpoint between all the positions used for the receiver. The sound source is detected by this method in the position marked with the cross.

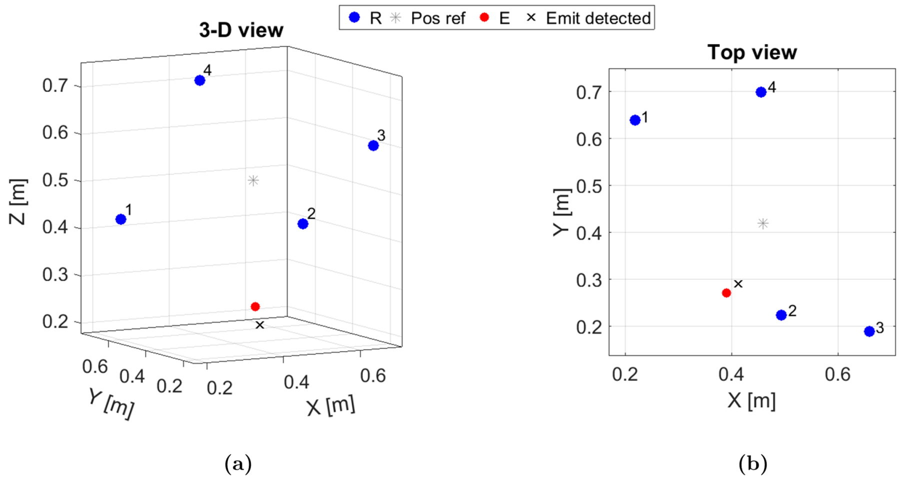


Figure 2. Positions of the emitter (E, in red) and the receivers (R, in blue) in the experimental set-up. (**left**) 3D view. (**right**) Top view

Table 1 shows the quantitative comparison between the coordinates for real position of the emitter and the one obtained by means of the localization with 3 and 4 positions for the receiver, as well as the error in each dimension for these two cases. As it can be observed, the emitter is localized with a maximum error of 4.8 cm by using 3 positions, and 4.4 cm using 4 positions, both in the y-dimension. Consequently, the goal regarding to precision is achieved in both cases, with a better accuracy when 4 positions are used.

Also the value the global distance between the real and the localized positions, defined as $d = \sqrt((x_L - x_R)^2 + (y_L - y_R)^2 + (z_L - z_R)^2)$, gives an idea of the accuracy of the localization. In this case, values of d = 5.5 cm and d = 5.2 cm are obtained with 3 and 4 receiver positions (mics), respectively. This means that, in global terms and as it was already expected, the emitter can be localized in a more accurate way with 4 positions.

Table 1. Comparative between the real and localized position of the emitter.

Coordinates	Δx [cm]	Δy [cm]	Δz [cm]	d [cm]
Real position	-	-	-	
Localized position (3 mics)	2.0	4.8	1.8	5.5
Localized position (4 mics)	0.4	0.7	0.3	0.9

The errors that are observed in Table 1 might be associated mainly to errors in measuring the coordinates of the positions of the receiver and possible reflections in the walls of the room in which the measurements are carried out.

The results corresponding to the calibration of the method that has been carried out in the anechoic chamber can be seen Appendix A.

4. Conclusions

In this article, the localization of an acoustic source has been presented by using a set of microphone's positions and considering the radial properties of the source. In this study, a generalized cross-correlation has been used as a main process for this purpose, due to the fact that the characteristics of the source are known. As a result, a system of equations has been presented in terms of the number the microphone's positions in a Cartesian coordinate system.

In this work, it has also been shown that, although a minimum of three microphone's positions are required to localize the sound source, the results are more accurate when more microphone's positions are considered, as these positions cover different directions of the space.

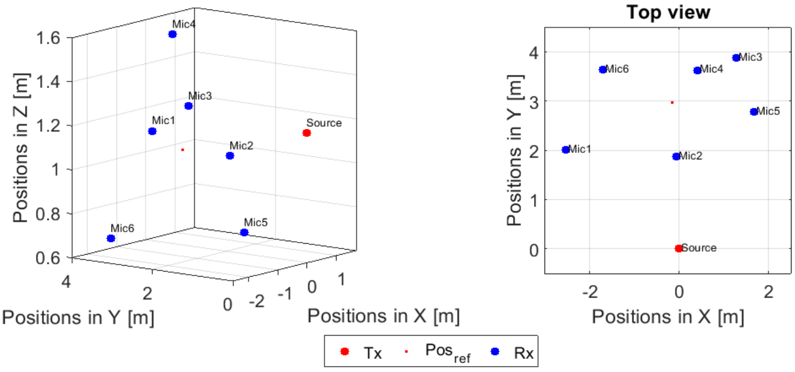
Institutional Review Board Statement:

Informed Consent Statement:

Data Availability Statement:

Appendix A

To calibrate the detection and localization algorithms, it is necessary to test them in different controlled environments. A simulated environment, in which different sensor points (3, 4, 5 and 6) simulations have been carried out for different combinations of sensors and theoretical sources, has been defined. Figure A1 shows a representation of the positions of the receivers together with one of the source points for which their spatial coordinates are known.



	X [m]	Y [m]	Z [m]
Mic 1	-2.53	2.00	1.23
Mic 2	-0.06	1.87	1.04
Mic 3	1.27	3.87	1.17
Mic 4	0.4	3.62	1.53
Mic 5	1.67	2.78	0.61
Mic 6	-1.7	3.63	0.67
Source	0	0	1.19

Figure A1. Positions of the emitter (*E*, in red) and the receivers (*R*, in blue) in the anechoic chamber tests. (**left**) 3D view. (**center**) Top view. (**right**) The Cartesian coordinates

With this, it is possible to create a controlled simulated environment. Thus, to test the results of the localization of a source reconstructed by the localization algorithm, with respect to the real values of its position, an error value was randomly added and 1000 simulations have been performed in each case.

Figure A2 shows an overview of the results provided by the localization algorithm. In all the cases, the abscissa axis shows the difference between the reconstructed position and the real position of the source while the ordinate axis shows the error added to the real position.

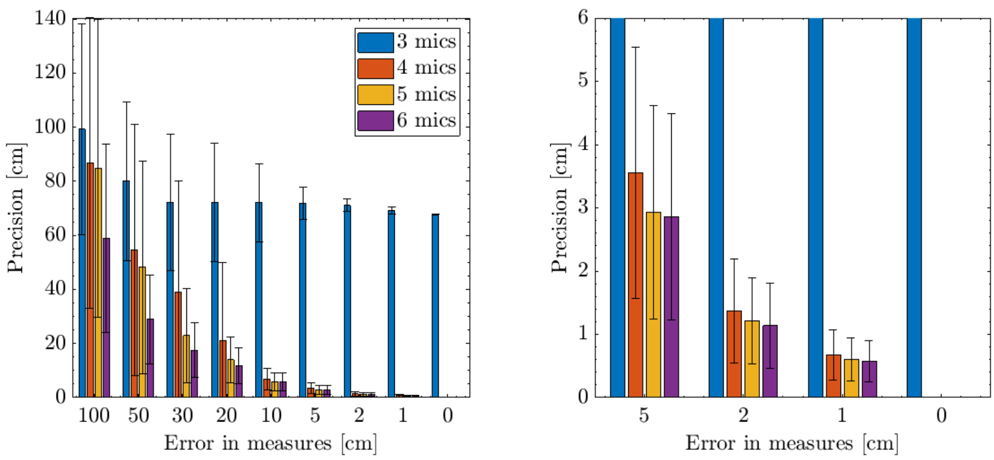


Figure A2. (**left**) Error of localization method expected. (**right**) Zoom in of the figure placed on the left side

On the one hand, as expected, for a larger number of sensors the reconstruction of the source considerably improves. On the other hand, when the error increases, the reconstruction of the source is more affected for a combination of 3 sensors, while for combinations of 4 or more sensors the results are independent of the error. This indicates that for combinations of at least 4 sensors it is sufficient to obtain results with good precision and higher robustness.

After testing the results of the location algorithms, it is necessary to test the detection algorithms. In this sense, We have generated different signals type (sine, MLS and sweep) by using an electro-mechanical loudspeaker with a 4-inch cone inside an anechoic chamber.

Additionally, four signal types were tested in the anechoic experimental set-up using 3, 4, 5, and 6 microphones (see Figure A1): MLS, sweep (10 Hz to 22 kHz), sinus of 500 Hz, and sinus of 4 kHz. These signals were the chosen to localize the source (tweeter, buffer, or the mean between both). Only for the 3 microphones set-up the detection can produce big differences in the precision of the algorithm, despite the fact that a GCC detection method obtain better precision in MLS and sweep signals [7].

In an anechoic environment, the signal detection with 4 (or more) microphones does not require precision on the order of the nanoseconds. Once the detection is assured with a precision less than 10 cm in the $dist_{ER}$ the improvement obtained by increasing the number of microphones is negligible. Thus, it is possible to state that a minimum of 4 in necessary to detect a source with dimensions bigger than 5 cm.

References

1. Pourmohammad, A.; Ahadi, S.M. N-dimensional N-microphone sound source localization. *EURASIP J. Audio Speech Music. Process.* **2013**, *2013* , 1–19.
2. Chen, J.C.; Hudson, R.E.; Yao, K. Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field. *IEEE Trans. Signal Process.* **2002**, *50*, 1843–1854.
3. Kim, S.; Chong, J.W. An efficient TDOA-based localization algorithm without synchronization between base stations. *Int. J. Distrib. Sens. Netw.* **2015**, *11*, 832351.
4. Qu, X.; Xie, L.; Tan, W. Iterative source localization based on TDOA and FDOA measurements. In Proceedings of the 2016 35th Chinese Control Conference (CCC), Chengdu, China, 27–29 July 2016; pp, 436–440.
5. Belloch, J.A.; Cobos, M.; Gonzalez, A.; Quintana-Ortí, E.S. Real-time sound source localization on an embedded GPU using a spherical microphone array. *Procedia Comput. Sci.* **2015**, *51*, 201–210.
6. Viola, S. KM3NeT acoustic positioning and detection system. In Proceedings of the EPJ Web of Conferences, 2019; p. 02006.
7. Adrián-Martínez, S.; Bou-Cabo, M.; Felis, I.; Llorens, C.D.; Martínez-Mora, J.A.; Saldaña, M.; Ardid, M. Acoustic signal detection through the cross-correlation method in experiments with different signal to noise ratio and reverberation conditions. In *International Conference on Ad-Hoc Networks and Wireless*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 66–79.
8. Rossing, T.D.; Rossing, T.D. *Springer Handbook of Acoustics*; Springer: New York, NY, USA, 2007.
9. Proakis, J.G.; Manolakis, D.G. *Digital Signal Processing*; MPC: New York, NY, USA, 1992.
10. Otero, J.; Ardid, M.; Felis, I.; Herrero, A. Acoustic location of Bragg peak for Hadrontherapy Monitoring. *Proceedings* **2018**, *4*, 6.