DOI: 10.61782/fa.2025.0791

FORUM ACUSTICUM
EURONOISE 2025

SOUND SOURCE LOCALIZATION IN SMALL-VOLUME INDUSTRIAL REVERBERANT CAVITIES THROUGH THE STEERED RESPONSE POWER PHAT-β

Konstantinos Rousou[1,2*] **Dionysios Panagiotopoulos**[1,2] **Dimitrios Chronopoulos**[1]
Elke Deckers[1,2]
[1] Department of Mechanical Engineering, KU Leuven, Belgium
[2] Flanders Make@KU Leuven, Belgium

ABSTRACT

Sound source localization in reverberant environments is hindered by strong sound reflections, which might lead to overlapping signals and thus to false detection. Although the Steered Response Power (SRP) method with a Phase Transformation (PHAT) has demonstrated high performance in noisy and large-volume reverberant environments, its applicability in smaller volumes has not been investigated. In that context, this study evaluates the performance of SRP-PHAT in small-scale industrial reverberant cavities, starting from the academic KU Leuven Soundbox and then transitions to a Laser Powder Bed Fusion (LPBF) 3D printing cavity, assuming a single broadband source and a randomly distributed microphone array. In the initial investigation, due to the small spacing distances between microphones combined with the sampling rate used, a microphone selection strategy based on their signal correlation was employed to determine the subset of microphones to be used within the SRP-PHAT. In both cases, it was demonstrated that the SRP-PHAT method can successfully locate the source, validating its applicability in smaller reverberant volumes.

Keywords: *Sound Source Localization, Steered Response Power, Reverberant cavities*

***Corresponding author**: konstantinos.rousou@kuleuven.be.

1. INTRODUCTION

Sound source localization (SSL) techniques aim to identify the positions of one or more sound sources by examining the captured signals by a microphone array. However, their application is hindered in small-scale industrial reverberant cavities, where increased reflections and shorter reverberation times create complex acoustic environments [1]. Existing SSL algorithms can be divided into three categories [2, 3]: (a) spectral estimation-based algorithms, (b) time difference of arrival (TDOA)-based algorithms, and (c) steered response power (SRP)-based beamforming algorithms. Spectral estimation algorithms use the covariance matrix obtained from the received signals. Although these algorithms are designed for narrowband signals, they can also be applied to broadband signals. On the other hand, TDOA-based algorithms begin by calculating the TDOA for each microphone pair. Using these values along with the microphone array geometry, the source position can be determined. However, both types of algorithms lack robustness in reverberant environments, resulting in a significant drop in performance and making them unsuitable for such conditions.

In contrast, algorithms based on steered beamforming involve filtering the received signals and estimating the source position by calculating the steered response power (SRP) at various candidate locations, selecting the position with the highest SRP value. A notable enhancement to this approach is the Steered Response Power-Phase Transformation (SRP-PHAT) algorithm proposed by Dibiase [2] which applies a phase transformation filter to the SRP method. This combination improves localization ac-

curacy by leveraging the robustness of the SRP method and the phase transformation's effectiveness in time delay estimation, making SRP-PHAT particularly effective in noisy and reverberant environments.

Several modifications to this method have been proposed in the literature. Donohue et al. [4] introduced the PHAT-β transformation, a variant of PHAT that adjusts the degree of spectral magnitude information incorporated through a single parameter β. They initially examined the optimal range of β through simulations and later validated their findings experimentally [5]. The same author proposed a time segmentation strategy to improve the efficiency of the SRP method, which was tested both computationally and experimentally in a mildly reverberant environment [6]. Other improvements include an enhanced SRP-PHAT algorithm using principal eigenvectors to improve localization performance [3] and the ODB-SRP-PHAT method, which integrates an offline database for real-time source localization [7]. A comprehensive survey of the SRP method and its variants, with a focus on SRP-PHAT, is provided in [8].

A common characteristic of the aforementioned studies is the application of SRP-PHAT to large acoustic volumes, where a balance between reflections and attenuation is considered. Additionally, most approaches rely on high sampling frequencies, which reduce the occurrence of near-zero time lags and mitigate quantization errors in time delay estimation. However, sound source localization in smaller industrial reverberant cavities presents additional challenges, as reflections can be as strong as the direct signal, creating spatial aliasing effects that hinder accurate localization. These conditions significantly increase the difficulty of detecting the source position within such environments.

To address these challenges, this study evaluates the performance of SRP-PHAT-β in small-scale industrial reverberation cavities. The first experimental setup, based on the academic KU Leuven Soundbox, employs a low sampling frequency, introducing the issue of near-zero time lags in certain microphone pairs. To counteract this, a microphone selection algorithm is developed to improve time delay estimation by removing redundant or problematic microphone pairs. The second experimental setup, based on a Laser Powder Bed Fusion (LPBF) 3D printing cavity, employs a higher sampling rate, which naturally mitigates the zero time lag issue without requiring additional microphone selection. In both cases, the method is tested using an irregular microphone array to detect a single broadband sound source. To reduce the complexity and focus on 2D localization performance, the source height is assumed to be known a priori.

The structure of this paper is as follows. Section 2 introduces the Steered Response Power PHAT-β method, while Section 3 describes the microphone selection algorithm developed. Section 4 presents the experimental setups and methodology, followed by Section 5, which evaluates the performance of the approach in both experimental environments. Finally, Section 6 concludes the paper with a summary of the findings.

2. STEERED RESPONSE POWER FORMULATION

2.1 Signal Model

Consider a 3D microphone array consisting of M microphones and Q sound sources located at positions $\mathbf{c}_q$, where $q = 1, \ldots, Q$. The recorded time-domain signal at the m^{th} microphone positioned at $\mathbf{r}_m$, generated by the q^{th} source, is given by

$$x_m(t) = h_{m,q}(t, \mathbf{r}_m, \mathbf{c}_q) * u_q(t, \mathbf{c}_q) + n_m(t), \quad (1)$$

where $h_{m,q}(t, \mathbf{r}_m, \mathbf{c}_q)$ represents the impulse response of the propagation path from $\mathbf{c}_q$ to $\mathbf{r}_m$, $u_q(t, \mathbf{c}_q)$ is the amplitude from the q^{th} source, and $n_m(t)$ represents the signal coming from noise sources.The time signal is segmented into windows of N samples and can be expressed as,

$$\mathbf{v}_m[n] = \Big[x_m[n]w[0],\, x_m[n+1]w[1], \ldots, x_m[n+N-1]w[N-1]\Big]^T, \quad (2)$$

where n is the first sample, $w[\cdot]$ are the values of the applied window function, and ^{T} denotes the transpose operation. In general, a Hann window $w[\cdot]$ is applied to signal $\mathbf{v}_m[n]$ before further processing.

2.2 Steered Response Power PHAT-β

The formulations and equations for the steered response coherent power (SRCP) calculation are based on the work presented in [6]. First, the PHAT-β operation is applied, defined as,

$$\bar{\mathbf{v}}_m[n] = \text{IFFT}\left(\left[\frac{\hat{V}_n[0]}{|\hat{V}_n[0]|^\beta}, \frac{\hat{V}_n[1]}{|\hat{V}_n[1]|^\beta}, \ldots, \frac{\hat{V}_n[2N-1]}{|\hat{V}_n[2N-1]|^\beta}\right]^T\right), \quad (3)$$

where $\hat{V}_n[k]$ is the k^{th} FFT coefficient of $\mathbf{v}_m[n]$ (a scalar), IFFT denotes the inverse Fourier transform, and β controls partial whitening of the signal. The process begins by the Fourier transform of the windowed signal, producing frequency-domain coefficients $\hat{V}_n[k]$. The magnitude of each coefficient is then normalized using the PHAT-β operation, as defined in equation (3), where the exponent β controls the degree of spectral whitening. This normalization ensures a balanced contribution of different frequency components in the coherent power calculation. Finally, the inverse Fourier transform (IFFT) reconstructs the processed time-domain signal $\bar{\mathbf{v}}_m[n]$. Then, based on a delay and weighting operation, the processed signal is aligned as follows,

$$\mathbf{s}_{m,p} = \bar{w}_{m,p} D_{m,p}[\bar{\mathbf{v}}_m[n]], \tag{4}$$

where $\mathbf{s}_{m,p}$ is the aligned signal for the m^{th} microphone and the p^{th} scanning point, $\bar{w}_{m,p}$ is the weighting function, and $D_{m,p}[\cdot]$ is the shift operator based on the relative propagation delay between the m^{th} microphone and the p^{th} grid point at position $\mathbf{g}_p$, $p = 1, \ldots, P$, which discretizes the volume or plane of interest into a finite set of P candidate points at a predefined grid distance between them. The number of samples by which the delay operator shifts the signal is given by

$$n_{m,p} = \text{round}\left(f_s \frac{(|\mathbf{g}_p - \mathbf{r}_m| - |\mathbf{g}_p - \mathbf{r}_{\text{ref}}|)}{c} \right), \tag{5}$$

where round$(\cdot)$ denotes rounding to the nearest integer, f_s is the sampling rate, and $\mathbf{r}_{\text{ref}}$ is the position of the farthest microphone from the p^{th} scanning point. The weight is calculated as,

$$w_{m,p} = \frac{1}{|\mathbf{g}_p - \mathbf{r}_m|}, \quad \bar{w}_{m,p} = \frac{w_{m,p}}{\sum_{m=1}^{M} w_{m,p}}. \tag{6}$$

The SRCP for each grid point is computed by removing the self-power from the SRP value as,

$$S_p = \left(\sum_{m=1}^{M} \mathbf{s}_{m,p} \right)^T \left(\sum_{m=1}^{M} \mathbf{s}_{m,p} \right) - \sum_{m=1}^{M} \left(\mathbf{s}_{m,p}^T \mathbf{s}_{m,p} \right), \tag{7}$$

resulting in the SRCP spectrum map, which visualizes the SRCP values over all grid points. The SRCP values are often normalized, allowing them to be scaled relative to the highest peak in the spectrum. By locating the maximum values, the potential source points can be identified. Removing the self-power of each microphone signal, which is always aligned with itself, helps to focus on inter-microphone relationships, enhancing the analysis of sound sources and spatial information.

3. MICROPHONE SELECTION ALGORITHM

3.1 Problem Formulation

The amplitude of signals recorded by microphones is influenced not only by their relative position to the sound source, but also by their proximity to reflective surfaces, such as corners. Microphones near reflective surfaces capture both direct and reflected waves, leading to increased amplitude variations due to the multiple propagation paths present.

Additionally, the sampling rate plays a critical role in delay estimation. Higher sampling rates provide finer time resolution, reducing quantization errors. Low sampling rates combined with small microphone spacing can introduce inaccuracies in delay estimation due to limited time discretization and resolution.

To address these challenges, a microphone selection algorithm is developed, considering both amplitude differences and the time lag between captured signals. The aim is to enhance localization accuracy by eliminating signals that are highly affected by reflections or exhibit inconsistencies in delay estimation.

3.2 Amplitude-Based Signal Selection

To account for positional and reflective influences, the peak amplitude of each time-segmented signal $\mathbf{v}_m$ is computed and stored in an array $\mathbf{A}_{\text{peak}}$, where each element is given by

$$\mathbf{A}_{\text{peak}}(m) = \max(|\mathbf{v}_m|). \tag{8}$$

To filter out signals with unusually high amplitudes, an automatic thresholding technique based on the standard deviation of peak amplitudes is applied. The threshold is given by

$$\text{Thld}_{\text{std}} = \mu(\mathbf{A}_{\text{peak}}) + \sigma(\mathbf{A}_{\text{peak}}), \tag{9}$$

where μ is the mean of $\mathbf{A}_{\text{peak}}$, and σ is the standard deviation. The computed threshold serves as a reference for identifying signals exhibiting excessive amplitude deviations.

3.3 Time Lag Estimation

In certain cases, false localization is observed, prompting further analysis of the time-domain microphone signals. The estimated time lag between two microphones is directly related to their spatial separation. If two microphones exhibit a near-zero time lag, it indicates that they are positioned such that their distances to the sound source are nearly identical. This scenario can occur when microphones are physically close together or when the sound wavefront reaches them at almost the same time due to their alignment relative to the source.

The time lag between signals captured by different microphones is estimated using cross-correlation,

$$R_{ij}(\tau) = \sum_{n} \mathbf{v}_i[n]\mathbf{v}_j[n+\tau], \tag{10}$$

and the estimated time delay is given by

$$\tau_{\text{est}} = \arg\max_{\tau} R_{ij}(\tau). \tag{11}$$

Due to the low sampling rate and small microphone spacing, the resolution of τ_{est} is limited, introducing quantization errors. These errors propagate into the SRP-PHAT algorithm, affecting the computed spatial power distribution and leading to incorrect localization peaks. Microphones that exhibit near-zero time lags may not provide new independent information, leading to redundant data and potential localization inaccuracies. Therefore, time lag estimation serves as a key criterion for identifying redundant or poorly spaced microphones.

3.4 Proposed Microphone Selection Algorithm

To improve localization accuracy, a microphone selection algorithm is developed, integrating both amplitude-based filtering and time lag estimation. For the latter, a small threshold ϵ is introduced to identify pairs with near-zero time delay, compensating for inaccuracies caused by time discretization. Since the measured signal is time-segmented, the threshold is set to one sample ($\epsilon = 1$) to differentiate actual signal delays from potential measurement errors. The procedure for microphone selection is described in Algorithm 1.

3.5 Microphone Selection Criteria

The microphone selection criteria are established based on both amplitude variations and time lag estimates, ensuring that redundant or suboptimal microphones are removed to improve localization accuracy. The spatial arrangement of

Algorithm 1 Microphone Selection Based on Signal Amplitude and Time Lag

Input:
- Time-segmented and windowed signals from N microphones: $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_M\}$.
- Microphone coordinates: $\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M\}$, where $\mathbf{r}_m = (x_m, y_m, z_m)$.

Step 1: Compute Peak Amplitudes
for each microphone m **do**
  Compute $\mathbf{A}_{\text{peak}}(m)$ using (1).
end for
Step 2: Remove High-Amplitude Signals
Define threshold using (2).
for each microphone m **do**
  if $\mathbf{A}_{\text{peak}}(m) > \text{Thld}_{\text{std}}$ **then**
    Remove microphone m.
  end if
end for
Step 3: Compute Time Lag via Cross-Correlation
for each microphone pair (i, j) **do**
  Compute $R_{ij}(\tau)$ using (4).
  Estimate time lag τ_{est} using (5).
end for
Step 4: Identify Near-Zero Time Lag Pairs
Identify microphone pairs satisfying $|\tau_{\text{est}}| \leq \epsilon$.
Step 5: Microphone Selection Rules
if three microphones share a near-zero time lag across two pairs **then**
  Compute Euclidean pairwise distances.
  Remove the two microphones with the largest Euclidean pairwise distance.
else if two microphones have near-zero time lag **then**
  Compute the average distance of each microphone to all others.
  Remove the microphone with the smallest average Euclidean pairwise distance.
end if
Output: Selected microphone subset.

the microphones plays a crucial role in the selection process, particularly when analyzing time lag characteristics.

3.5.1 Amplitude Thresholding Criteria

Signals satisfying

$$\mathbf{A}_{\text{peak}}(m) > \text{Thld}_{\text{std}} \tag{12}$$

are excluded from the localization process. This step isolates the most relevant signals while eliminating those affected by strong reflections, which could distort localization accuracy.

3.5.2 Time Lag-Based Selection

When three microphones form two near-zero time lag pairs, it suggests an overly clustered arrangement. In this case, removing two of the three microphones with the largest Euclidean pairwise distance ensures a more balanced spatial distribution. Conversely, when only two microphones exhibit a near-zero time lag, the microphone with the smallest average Euclidean pairwise distance to the rest of the array is removed to avoid excessive clustering. These rules prevent redundant microphone placements, improving localization robustness by preserving a well-spaced microphone array.

4. EXPERIMENTAL INVESTIGATION

4.1 Experimental Setup 1: Soundbox

The KU Leuven Soundbox is selected for the SRP method implementation due to its irregular geometry and small volume, which leads to stronger reflections and complicates the sound source identification within the chamber. An irregular array of nine PCB 378B20 microphones is used, which are ideal for diffuse field applications. For acoustic excitation, a Qsource with a nozzle diameter of 30 mm is employed due to its controllability, ensuring consistent acoustic excitation in a controlled setting. A broadband signal up to 2000 Hz is defined as the output, and the sampling rate is set to 4086 Hz. Figure 1a illustrates the interior of the chamber with the microphones and source installed.

4.2 Experimental Setup 2: LPBF Cavity

The LPBF cavity is selected as a second test environment (Figure 2a) due to its structural complexity and the presence of multiple reflective surfaces, which significantly affect sound propagation. Unlike the KU Leuven Soundbox, which has an irregular geometry but is a controlled setup, the LPBF cavity introduces additional modeling uncertainties arising from metallic surfaces with varying absorption characteristics, the presence of cables and mechanical components, strong reverberation effects, and background noise from operational systems. These factors create overlapping reflections, non-uniform propagation paths, and increased ambient noise levels, further complicating the accurate localization of sound sources.

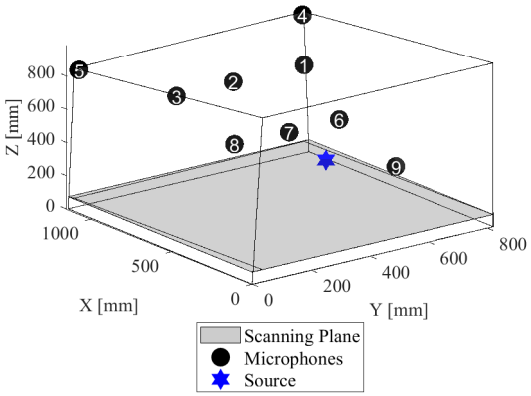
To mitigate these challenges, an irregular array of eight microphones of the same type as in experimental setup 1 is carefully positioned throughout the cavity, ensuring broad spatial coverage while avoiding areas with dominant noise sources, such as the active gas pump. An additional microphone is placed in front of the gas pump to investigate its spectrum, but it is not included in the localization process. For acoustic excitation, a 30 mm diameter mini wireless speaker is used to generate a broadband signal up to 10 kHz, created in MATLAB. This speaker is chosen due to space constraints for cable routing and nozzle placement, as well as to better simulate real operating conditions. The sampling rate is set to 82 kHz, improving time resolution and minimizing quantization errors caused by lower sampling rates.

4.3 2D Source Localization and Parameter Settings

This paper focuses on 2D source localization, assuming that the height of the source is known beforehand. As shown in Figure 1b and Figure 2b, a scanning plane is defined at this height and divided into grid points spaced 5 mm apart. In the KU Leuven Soundbox, the scanning plane extends across the entire chamber. Conversely, in the LPBF case, a significant part of the internal space is occupied by the module itself, along with cables and mechanical components. Therefore, the main acoustic volume of interest is defined as the region above the module (Figure 2a), excluding internal components such as cables, support structures, and mechanical elements.

For the SRP PHAT-β algorithm and the spectrum map calculation, a 0.2 s time window is used for both experimental setups. While different window lengths gave good results, this duration was chosen for consistency in the analysis. The time window is selected by identifying the maximum peak across all time signals, assumed to correspond to the direct signal, and centering the window around it. The β value is set to 0.5, which, based on literature for broadband sources, typically ranges between 0.5 and 0.8, often resulting in significant performance improvements [5].

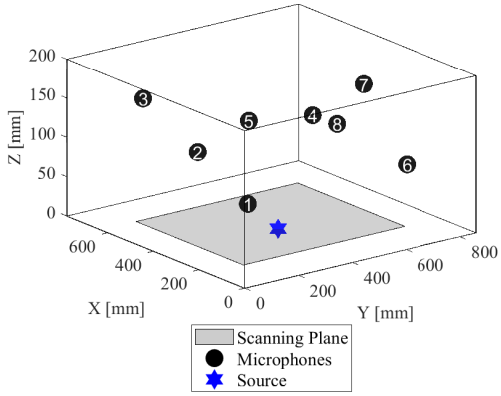
(a)



(b)

Figure 1: Soundbox (a) interior cavity and (b) representation

(a)



(b)

Figure 2: LPBF (a) interior cavity and (b) representation

5. PERFORMANCE EVALUATION

5.1 Source localization within the Soundbox

5.1.1 Initial Localization Performance

To assess the initial localization performance within the Soundbox, the SRCP value is calculated for all grid points of the scanning plane using all nine microphones. The resulting spectrum map is illustrated in Figure 3a. It can be observed that the calculated source position deviates from the actual position due to multiple false peaks with larger spectrum amplitudes, resulting in incorrect source detection.

A detailed analysis of the measured time signals highlighted significant amplitude variations, zero time lag redundancies, and microphone clustering effects. For exactly this reason, the microphone selection algorithm presented in Section 3 was developed, in order to refine the subset of microphones used for localization.

5.1.2 Localization Performance Using the Microphone Selection Algorithm

Applying the microphone selection algorithm, the optimal subset of microphones is determined based on amplitude differences and time lag characteristics. The selection process identified microphones 1, 3, 7, and 9 (see Figure 1b) as the most effective for source localization, ensuring a balanced distribution while eliminating sources of error identified in the initial analysis.

A detailed examination of the selection process revealed that microphones 4 and 5, positioned at the cor-

ners of the Soundbox, exhibited significantly higher amplitude signals compared to the others, violating the amplitude criterion for reliable selection. Furthermore, microphones 3 and 2 recorded nearly identical signals with zero time lag, providing no useful spatial separation. Additionally, microphones 6, 7, and 8, due to their close proximity and the low sampling rate, formed a cluster that produced near-zero time lags, leading to redundant and less informative data.

By recalculating the spectrum map using only this refined subset (Figure 3b), false peaks are eliminated, and the detected source position closely matches the actual location. This demonstrated a substantial improvement in localization accuracy, confirming the effectiveness of the microphone selection strategy.

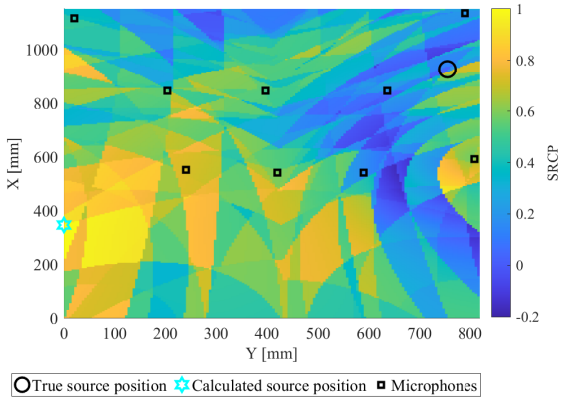
5.2 Source localization within the LPBF cavity

Building upon the results obtained in the Soundbox, the localization performance of SRP-PHAT-β is evaluated in the LPBF cavity, a more complex and acoustically challenging environment. Initially, the developed microphone selection algorithm is applied to analyze the captured time signals, aiming to identify potential issues such as amplitude imbalances or zero time lag that could affect localization accuracy. Unlike the Soundbox case, no microphones were excluded from the localization process. This outcome can be attributed to the significantly higher sampling rate, which ensured sufficient time resolution to distinguish signal arrivals effectively. Furthermore, the microphone positions were carefully selected to avoid placement near reflective surfaces or dominant noise sources, reducing the risk of excessive amplitude variations or redundant signals.

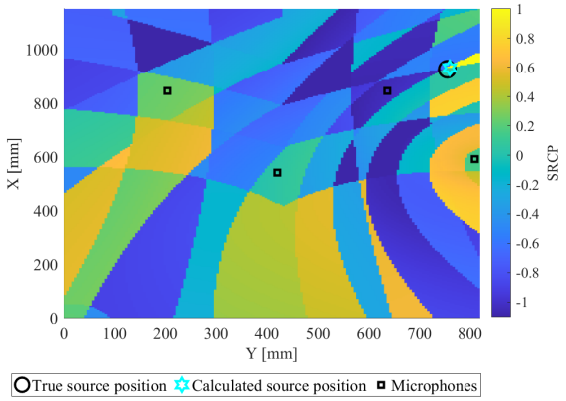
The calculated spectrum map, using all eight sensors, is illustrated in Figure 4. Compared to the Soundbox case, the SRP-PHAT-β response shows a stronger concentration of energy around the actual source position, benefiting from improved spatial resolution due to the higher sampling frequency. This suggests that the method remains effective even in the presence of increased reverberation and background noise, demonstrating its robustness in challenging industrial environments.

6. CONCLUSIONS

In small industrial reverberant cavities, reflections can be as strong as the direct signal, resulting in complex sound fields that make accurate source localization challenging. In this paper, the SRP-PHAT-β algorithm is evaluated in



(a)



(b)

Figure 3: 2D source localization within the Soundbox using nine microphones – (a) Without and (b) With microphone selection algorithm

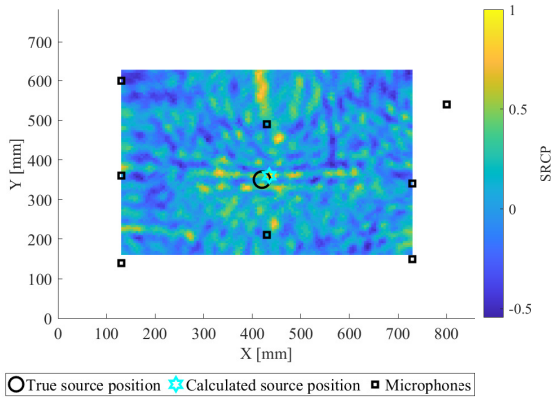


Figure 4: 2D source localization within the LPBF cavity using eight microphones

two different setups, each using broadband sources in distinct frequency ranges. The findings highlight its capability to achieve reliable source localization, demonstrating its potential for use in compact reverberant settings.

While the method consistently located the source, performance variations were observed due to variations in signal amplitudes and the impact of sampling rate. In the first setup, a microphone selection algorithm is developed based on amplitude differences and time lags to address issues related to microphone positioning and quantization errors caused by the low sampling rate. In the second setup, where the significantly higher sampling rate improved time resolution, all microphones contributed effectively to the localization. These findings emphasize the critical role of both microphone placement and sampling rate in localization accuracy. Optimizing microphone configurations, particularly in space-constrained reverberant environments, is crucial for reliable performance.

7. ACKNOWLEDGMENTS

This research was partially supported by Flanders Make, the strategic research centre for the manufacturing industry. Furthermore, Internal Funds KU Leuven are gratefully acknowledged for their support.

8. REFERENCES

[1] M. Meissner, "Acoustics of small rectangular rooms: Analytical and numerical determination of reverberation parameters," *Applied Acoustics*, vol. 120, pp. 111–119, 2017.

[2] J. H. Dibiase, H. F. Silverman, and M. Brandstein, "Robust localization in reverberant rooms," *Springer eBooks*, pp. 157–180, 01 2001.

[3] X. Wan and Z. Wu, "Improved steered response power method for sound source localization based on principal eigenvector," *Applied Acoustics*, vol. 71, pp. 1126–1131, 12 2010.

[4] K. D. Donohue, J. Hannemann, and H. G. Dietz, "Performance of phase transform for detecting sound sources with microphone arrays in reverberant and noisy environments," *Signal Processing*, vol. 87, pp. 1677–1691, 07 2007.

[5] A. Ramamurthy, H. Unnikrishnan, and K. D. Donohue, "Experimental performance analysis of sound source detection with srp phat-β," *IEEE Southeastcon 2009*, 03 2009.

[6] K. D. Donohue and P. M. Griffioen, "Computational strategy for accelerating robust sound source detection in dynamic scenes," *IEEE Southeastcon 2014*, 03 2014.

[7] D.-B. Zhuo and H. Cao, "Fast sound source localization based on srp-phat using density peaks clustering," *Applied Sciences*, vol. 11, p. 445, 01 2021.

[8] E. Grinstein, E. Tengan, B. Çakmak, T. Dietzen, L. Nunes, T. van Waterschoot, M. Brookes, and P. A. Naylor, "Steered response power for sound source localization: a tutorial review," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2024, 11 2024.