Sensitivity-based region selection in the steered response power algorithm

Published
DOI:10.1016/j.sigpro.2018.07.002

Sensitivity-Based Region Selection in the Steered Response Power Algorithm

Daniele Salvati, Carlo Drioli, Gian Luca Foresti

Department of Mathematics, Computer Science and Physics, University of Udine

Abstract

The steered response power (SRP) algorithm is a well-studied method for sound source localization using a microphone array. Recently, different improvements based on the accumulation of all time difference of arrival (TDOA) information have been proposed in order to achieve spatial resolution scalability of the grid search map and reduce the computational cost. However, the TDOA information distribution is not uniform with respect to the search grid, as it depends on the geometry of the array, the sampling frequency, and the spatial resolution. In this paper, we propose a sensitivity-based region selection SRP (R-SRP) algorithm that exploits the nonuniform TDOA information accumulation on the search grid. First, high and low sensitivity regions of the search space are identified using an array sensitivity estimation procedure; then, through the formulation of a peak-to-peak ratio (PPR) measuring the peak energy distribution in the two regions, the source is classified to belong to a high or to a low sensitivity region, and this information is used to design an ad hoc weighting function of the acoustic power map on which

Email addresses: `daniele.salvati@uniud.it` (Daniele Salvati), `carlo.drioli@uniud.it` (Carlo Drioli), `gianluca.foresti@uniud.it` (Gian Luca Foresti)

the grid search is performed. Simulated and real experiments show that the proposed method improves the localization performance in comparison to the state-of-the-art.
Keywords:
acoustic source localization, microphone array, SRP-PHAT, sensitivity map, region selection

---

1. Introduction

Sound source localization using microphone arrays received significant attention by the scientific community due to its importance in acoustic scene analysis, signal enhancement, and speaker recognition and tracking [1, 2, 3, 4, 5, 6].

In general, the localization can be computed with indirect and direct methods. The former are based on the computation of a set of time difference of arrivals (TDOAs), obtained by measurements across various combinations of microphones [7, 8], and on the estimation of the source position using geometric reasoning [9, 10, 11]. Direct methods are based on the steered response power (SRP) beamformers [12, 13, 14], on subspace algorithms [15, 16, 17], or on maximum-likelihood estimators [18, 19, 20]. They are very attractive for acoustic applications due to their robustness in noisy and reverberant conditions.

The conventional SRP algorithm is based on the delay-and-sum beamforming technique [21]. Broadband SRP is typically implemented with the phase transform (PHAT) pre-whitening [7], which provides a normalization of narrowband SRPs and increases the spatial resolution [22]. This allows a better identification of direct path and early reflections in a reverberant environment. SRP-PHAT has the advantage that it can be computed by considering the generalized cross-correlation (GCC) [7] between each microphone pair, and by summing TDOA values related to the search space [13]. This implementation is computationally more efficient if compared to methods that require a computation of narrowband SRP maps and their fusion [22]. However, the search procedure can be very expensive. Thus, iterative volume-search-based procedures have been recently proposed [23, 24, 25], which aim at reducing the computational complexity of this step. These methods take into account the accumulation of TDOA information [26, 24, 25] to achieve the reduction of the spatial grid resolution without loss of information, and uses sequentially volumetric refinement steps for increasing the localization accuracy.

It has been demonstrated, using the geometrically sampled grid (GSG) algorithm [27], that the accumulation of all TDOA values from GCC functions is not uniform within the search space, and as a consequence the acoustic map is characterized by high and low sensitivity regions. The advantage of using all TDOA information is to obtain a robust localization in the high sensitivity region with adverse noisy and reverberant conditions. If the sound source is located in a low sensitivity region, however, its localization is more prone to be unstable and affected by errors. This is due to the fact that the acoustic map energy peak corresponding to the actual source position might be lower than the peaks corresponding to noise and reverberation in the high sensitivity region, emphasized by the prominent TDOA accumulation. SRP-based methods that use all TDOA information were proposed in

[26, 24, 27]. In [23], it was also proposed a SRP method that uses all TDOA information, providing however a power normalization in each volume with respect to the number of TDOA values. This approach mitigates the problem due to the nonuniform TDOA accumulation, but also reduces the robustness in the high sensitivity region. In [25], a SRP method based on the use of two grids (a coarser one, and a finer one) was proposed. This method uses an uniform TDOA accumulation in each volume, mitigating the problem of nonuniform distribution, but it discards part of the information available, reducing the TDOA accumulation that can be positively used in the high sensitivity region.

In this paper, we consider the localization of a single source in noisy and reverberant conditions. This scenario can be of interest in different practical applications such as videoconferencing systems or in human-computer interaction systems. With the aim of using all the TDOA information from the GCC functions and of exploiting the robustness in the high sensitivity region, we propose a sensitivity-based region selection SRP algorithm, named R-SRP, which is organized in two steps: first, it establishes if the source is positioned in a high or low sensitivity region, through the formulation of a peak-to-peak ratio (PPR) measuring the peak energy distribution in the high and low sensitivity regions of the array, determined through the GSG algorithm. Then, it proceeds with the search of the acoustic source in the selected region using, when opportune, the sensitivity function to weight the power acoustic map and reduce the impact of noise. It will be shown that this array sensitivity-informed method effectively reduces the localization errors due to the nonuniform distribution of the TDOA accumulation in the power acoustic map.

2. Steered Response Power

Let us consider a reverberant room G, M microphones positioned at coordinates $\mathbf{r}_m = [x_m, y_m, z_m]^T (m = 1, 2, \ldots, M)$, where $(\cdot)^T$ denotes the transpose operator, and a single source $\mathbf{r}_s(k) = [x_s(k), y_s(k), z_s(k)]^T$ active at time k. The SRP-PHAT based on all the TDOA information can then be expressed in terms of GCC functions as [13, 26, 23, 24, 27]

$$\phi(\mathbf{r}, k) = \sum_{m_1=1}^{M-1} \sum_{m_2=m_1+1}^{M} \sum_{\tau=\tau_{m_1 m_2}^{\min}(\mathbf{r})}^{\tau_{m_1 m_2}^{\max}(\mathbf{r})} R_{m_1 m_2}(\tau, k), \tag{1}$$

where $\mathbf{r} = [x, y, z]^T \in G$ is a generic grid position with spatial resolution Δ, $\tau_{m_1 m_2}^{\min}(\mathbf{r})$ and $\tau_{m_1 m_2}^{\max}(\mathbf{r})$ denote the bounds of the accumulated TDOAs between the microphone m_1 and m_2 for the position $\mathbf{r}$, and the GGC-PHAT [7] function is

$$R_{m_1 m_2}(\tau, k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{X_{m_1}(w, k) X_{m_2}^*(w, k)}{|X_{m_1}(w.k) X_{m_2}^*(w, k)|} e^{jw\tau} dw, \tag{2}$$

where τ is the time lag, w is the angular frequency, $X_m(w, k)$ is the transform of the signal observed at microphone m, $(\cdot)^*$ denotes the complex conjugate, j denotes the imaginary unit, and $|\cdot|$ denotes absolute value. The GCC-PHAT is computed in the frequency domain using the discrete Fourier transform, and hence the SRP is computed on a block-by-block basis. If $\tau_{m_1 m_2}^{\min}(\mathbf{r}) = \tau_{m_1 m_2}^{\max}(\mathbf{r})$, equation (1) represents the conventional SRP-PHAT algorithm [13]. The accumulation limits can be determined with different strategies which can rely on the gradient of the inter-microphone time delay function corresponding to each microphone pair in the M-SRP [26], on the gradient of the inter-microphone time delay function exploiting the mean of the accumulated GGC-PHAT values for each volume in the I-SRP [23], on the surrounding cube taking into account vertices of the volume in the H-SRP [24], or on discrete representations of the the hyperboloids related to all possible TDOA values in the GSG-based method (G-SRP) [27].

Once the array steered response power funtion $\phi(\mathbf{r}, k)$ is available, the source position can be estimated by searching its maximum in the search region

$$\widehat{\mathbf{r}}_s(k) = \underset{\mathbf{r}}{\operatorname{argmax}}[\phi(\mathbf{r}, k)]. \tag{3}$$

3. Geometrically Sampled Grid

The proposed R-SRP algorithm extends the G-SRP [27] algorithm by including a region selection procedure. The G-SRP is based on the GSG method, in which the search space is obtained by discretizing, with a given spatial resolution, the hyperboloids representing the surface on which the TDOAs are constant, and by finally computing a grid related to the intersections between these discrete curves. It thus allows the accumulation of the whole TDOA information provided by the GCC functions into the search space, the design of an acoustically-coherent space grid, and the design of a sensitivity map.

Let now consider the dicretization of the search space G with a spatial resolution Δ. A discrete hyperboloid related to a microphone pair (m_1, m_2) and a TDOA $\tau_{m_1 m_2}$ can be represented as a finite set $\Lambda_{\tau_{m_1 m_2}}$ of points in $\mathbb{R}^3$, describing the hyperboloid when the x, y, and z-axis are discretized with spatial resolution Δ (for a detailed discussion on the hyperboloid discretization procedure, see [27]).

In the implementation of the G-SRP, the discrete hyperboloids and the TDOA information are stored in four look-up tables. The tables are computed off-line, and then used on-line to estimate the acoustic energy and computing the accumulation of the GCC-PHAT function information due to all the sensor pairs involved. To each discrete hyperboloid point, we assign an index q, so that we have a table $\gamma_r(q)$ for the position, a table $\gamma_p(q)$ for the pair index, and a table $\gamma_\tau(q)$ for the TDOA. The last look-up table, $\delta(\mathbf{r})$, is the GSG sensitivity map, which contains the number of all the discrete surfaces intersecting in the position $\mathbf{r}$. The sensitivity map provides information on the distribution of TDOAs into the search space, and thus it defines a measure of the localization accuracy of the array and a mean to identify those areas for which it is more accurate.

If we call $T_{m_1m_2} = \text{fix}\left(\frac{||\mathbf{r}_{m_1}-\mathbf{r}_{m_2}||f_s}{c}\right)$ the maximum TDOA in samples for the sensor pair (m_1, m_2), where $\text{fix}(\cdot)$ denotes the round toward zero operation, f_s is the sampling frequency, c is the speed of sound, and $||\cdot||$ denotes Euclidean norm, we have $(2T_{m_1m_2}+1)M(M-1)/2$ discrete hyperboloids. The procedure to build the GSG grid and the sensitivity map $\delta(\mathbf{r})$ is given by the following steps:

1. Initialize $\delta(\mathbf{r}) = 0$ for all $\mathbf{r} \in G$ and of index q=0;
2. For each sensor pair (m_1, m_2) and for all TDOA values $\tau_{m_1m_2}$ in the range [-$T_{m_1m_2}$,$T_{m_1m_2}$], calculate the discrete hyperboloid $\Lambda_{\tau_{m_1m_2}}$, and for each grid position $\mathbf{r} \in \Lambda_{\tau_{m_1m_2}}$, fill the look-up tables $\gamma_r(q)$, $\gamma_p(q)$,

Algorithm 1 GSG Algorithm

M: number of microphones
Δ: spatial resolution
Initialization: for each grid position $\mathbf{r} \in G$, $\delta(\mathbf{r}) = 0$, $q = 0$
for $m_1 = 1$ to $M - 1$ **do**
  for $m_2 = m_1 + 1$ to M **do**
    for $\tau_{m_1 m_2} = -T_{m_1 m_2}$ to $T_{m_1 m_2}$ **do**
      Calculate the discrete hyperboloid $\Lambda_{\tau_{m_1 m_2}}$
      for all $\mathbf{r} \in \Lambda_{\tau_{m_1 m_2}}$ **do**
        $\gamma_r(q) = \mathbf{r}$, $\gamma_p(q) = [m_1, m_2]^T$, $\gamma_\tau(q) = \tau_{m_1 m_2}$
        $\delta(\mathbf{r}) = \delta(\mathbf{r}) + 1$
        q=q+1
      end for
    end for
  end for
end for
Apply the constraint $\delta(\mathbf{r}) < \mu \Rightarrow \delta(\mathbf{r}) = 0$, $\forall \mathbf{r} \in G$
Update $\gamma_r(q)$, $\gamma_p(q)$, and $\gamma_\tau(q)$,
Calculate the GSG grid $\Gamma_r = \{\mathbf{r} : \delta(\mathbf{r}) \neq 0\}$.

and $\gamma_\tau(q)$, incremet by one the value of the look-up table $\delta(\mathbf{r})$, and increment q by one;

3. After the geometric discrete analysis of the hyperboloids has terminated, apply the constraint $\delta(\mathbf{r}) < \mu \Rightarrow \delta(\mathbf{r}) = 0$, $\forall \mathbf{r} \in G$, where $\mu = 3$ and $\mu = 2$ in case of 3D and 2D localization, respectively. The constraint has the goal of discarding those space grid points that are useless for the localization. Finally, update the look-up tables $\gamma_r(q)$, $\gamma_p(q)$, and $\gamma_\tau(q)$, and calculate the acoustically-coherent GSG grid $\Gamma_r = \{\mathbf{r} : \delta(\mathbf{r}) \neq 0\}$.

The GSC algorithm is summarized in Algorithm 1.

Finally, we can write the G-SRP as

$$\phi(\mathbf{r},k) = \sum_{m_1=1}^{M-1} \sum_{m_2=m_1+1}^{M} \sum_{z\in Z_{r,m_1m_2}} R_{m_1m_2}(\gamma_\tau(z),k), \tag{4}$$

where

$$Z_{r,m_1m_2} = \{q : [\gamma_r(q) = \mathbf{r}] \wedge [\gamma_p(q) = [m_1, m_2]^T]\}, \tag{5}$$

are the look-up table indices corresponding to the TDOAs for the position $\mathbf{r} \in \Gamma_r$ of the sensor pair (m_1, m_2).

4. Sensitivity-Based Region Selection

We model the power function $\phi(\mathbf{r},k)$ given by (1) as the sum of the contribution of the source $\phi_s(\mathbf{r},k)$ and the contribution of noise $\phi_v(\mathbf{r},k)$. For simplicity, we drop the time index k from now on. If we assume that the noise component $\phi_v(\mathbf{r})$ has normal distribution $N(0,\sigma^2)$, we can write the acoustic map as

$$\phi(\mathbf{r}) = \phi_s(\mathbf{r}) + \phi_v(\mathbf{r}) = \phi_s(\mathbf{r}) + \sigma^2\delta(\mathbf{r}). \tag{6}$$

Note that the noise component σ^2 is related to the noise actually present in the GCC-PHAT functions. If we consider that the source is not active, we can write that $R_{m_1m_2}(\tau) = \sigma^2$, and we can see from (1) that the accumulation of TDOA values in each grid position is given by the number of sample values from all sensor pairs, i.e the information contained in the sensitivity map $\delta(\mathbf{r})$, resulting in $\phi_v(\mathbf{r}) = \sum_{m_1=1}^{M-1} \sum_{m_2=m_1+1}^{M} \sum_{\tau=\tau_{m_1m_2}^{\min}(\mathbf{r})}^{\tau_{m_1m_2}^{\max}(\mathbf{r})} \sigma^2 = \sigma^2\delta(\mathbf{r})$. According to [27], we can divide the search space sensed by the array into two regions with different sensitivity:

$$
\begin{aligned}
H &= \{\mathbf{r} \in G : \delta(\mathbf{r}) \geq \eta\}, \\
L &= \{\mathbf{r} \in G : \delta(\mathbf{r}) < \eta\},
\end{aligned}
\tag{7}
$$

where H and L denote the high and low sensitivity region respectively, and η is a threshold computed as

$$
\eta = \frac{\max[\delta(\mathbf{r})] + \min[\delta(\mathbf{r})]}{2}, \tag{8}
$$

with $\max[\cdot]$ and $\min[\cdot]$ denoting the maximum value and the minimum value, respectively. Based on the available data, i.e. the power function $\phi(\mathbf{r})$ and the function $\delta(\mathbf{r})$, with $\mathbf{r} \in G$, a rough region classification criterion would check if the maximum of $\phi(\mathbf{r})$ was found in L or H, and assign the source to that region. Figure 1 and 2 represent two qualitative examples of SRP functions for the source in H and L, respectively. Due to the additive noise component, this criterion would misclassify the region in those cases in which, even though the source is located in L (i.e, $\phi_s(\mathbf{r})$'s maximum is in L), the maximum of $\phi(\mathbf{r})$ is found in H due to the additive noise component, amplified in H by the function $\delta(\mathbf{r})$ (see Figure 2). The opposite situation, i.e. occurring when the source is in H but the maximum of $\phi(\mathbf{r})$ is found in L, is very unlikely since the function $\delta(\mathbf{r})$ is low-valued in this region and would hardly be responsible for a high-energy noise peak able to affect the global maximum. We thus aim at improving the baseline criterion by finding a more effective, data-dependent threshold for the region selection. We define the following

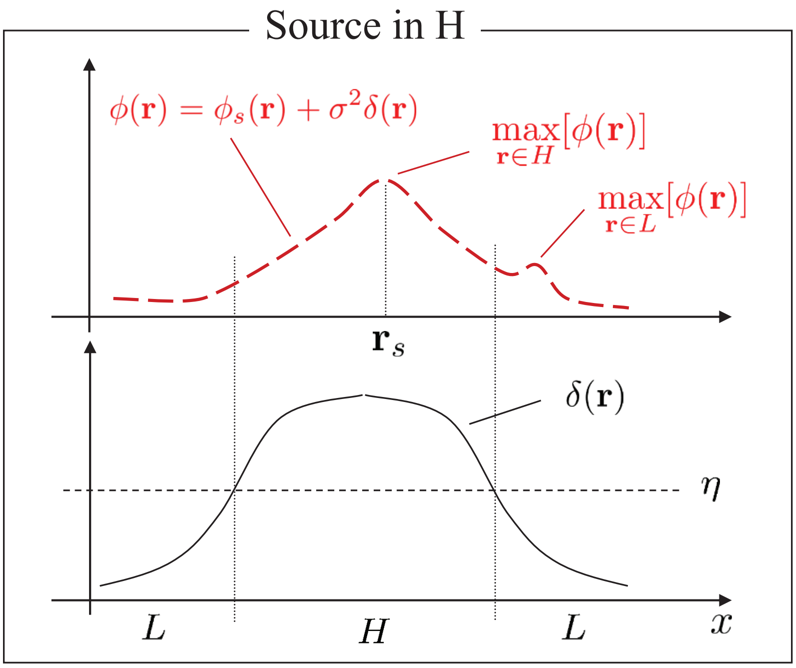

Figure 1: A schematic representation of the SRP profile along x axis when the source is positioned in the high sensitivity region.

peak-to-peak ratio

$$\mathrm{PPR} = \frac{\max_{\mathbf{r}\in H}[\phi(\mathbf{r})]}{\max_{\mathbf{r}\in L}[\phi(\mathbf{r})]}, \tag{9}$$

which is a measure of the difference between the maximum energy peak in the high sensitivity region and the one in the low sensitivity region. The baseline criterion would classify the source as belonging to L if $\mathrm{PPR} < 1$, and to H otherwise. Since this criterion can be assumed robust for $\mathrm{PPR} < 1$ (and thus maximum of $\phi(\mathbf{r})$ in L), we will focus on the $\mathrm{PPR} \geq 1$ case in what follows.

Let us call $\bar{\mathbf{r}}$ the position of $\phi(\mathbf{r})$'s maximum, and let suppose now that the source is actually positioned in the high sensitivity region H. From what

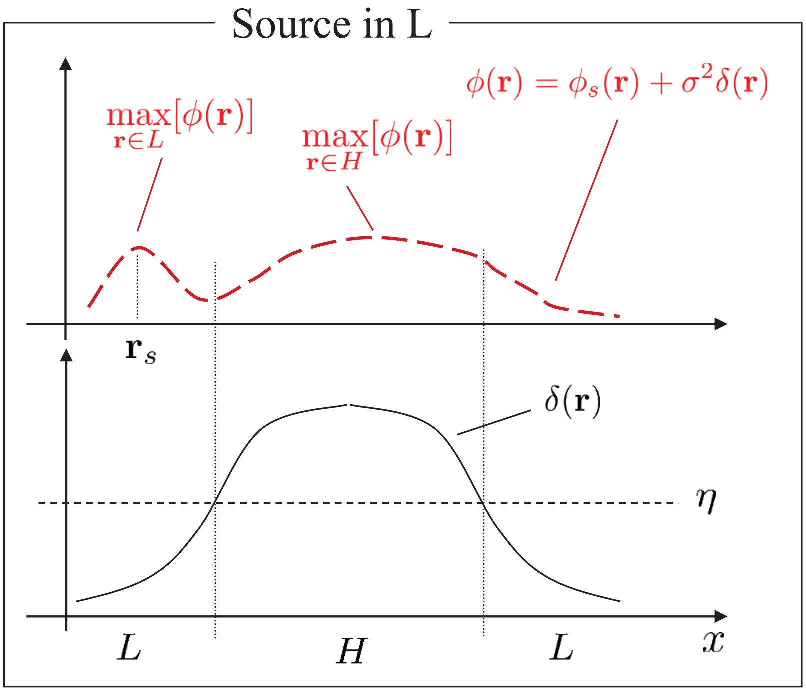

Figure 2: A schematic representation of the SRP profile along the x axis when the source is positioned in the low sensitivity region and the maximum of the overall region is positioned in the H region.

said so far, we can assume that $\bar{\mathbf{r}}$ will fall in H and thus restrict the maximum search to the high sensitivity region, i.e. $\bar{\mathbf{r}} = \underset{\mathbf{r}\in H}{\operatorname{argmax}}[\phi(\mathbf{r})]$. We can say in this case that

$$\max_{\mathbf{r}\in L}[\phi_s(\mathbf{r}) + \sigma^2\delta(\mathbf{r})] \approx \max_{\mathbf{r}\in L}[\sigma^2\delta(\mathbf{r})], \tag{10}$$

i.e. the contribution of the source will be negligible in the computation of the SRP maximum in L. We can thus write the PPR as

$$\mathrm{PPR} = \frac{\max_{\mathbf{r}\in H}[\phi_s(\mathbf{r}) + \sigma^2\delta(\mathbf{r})]}{\max_{\mathbf{r}\in L}[\sigma^2\delta(\mathbf{r})]} = \frac{\phi_s(\bar{\mathbf{r}}) + \sigma^2\delta(\bar{\mathbf{r}})}{\sigma^2\max_{\mathbf{r}\in L}[\delta(\mathbf{r})]}. \tag{11}$$

Since it is $\max\limits_{\mathbf{r}\in L}[\delta(\mathbf{r})] = \eta$, equation (11) leads to the condition PPR $\geq \frac{\delta(\bar{\mathbf{r}})}{\eta}$. We can now show that this threshold also correctly classifies the sensitivity region when the source is located in L but the maximum of $\phi(\mathbf{r})$ is found in H due to the effect of noise. In this case, we can write

$$\max_{\mathbf{r}\in H}[\phi_s(\mathbf{r}) + \sigma^2\delta(\mathbf{r})] \approx \max_{\mathbf{r}\in H}[\sigma^2\delta(\mathbf{r})] = \sigma^2\delta(\bar{\mathbf{r}}), \tag{12}$$

and the peak-to-peak ratio becomes

$$\mathrm{PPR} = \frac{\sigma^2\delta(\bar{\mathbf{r}})}{\max\limits_{\mathbf{r}\in L}[\Phi_s(\mathbf{r}) + \sigma^2\delta(\mathbf{r})]} < \frac{\delta(\bar{\mathbf{r}})}{\eta}. \tag{13}$$

We can thus adopt the following L-H classification criterion:

$$\widehat{\mathbf{r}}_s \in \begin{cases} L & \text{if} \quad \mathrm{PPR} < 1, \\ L & \text{if} \quad 1 \leq \mathrm{PPR} < \frac{\delta(\bar{\mathbf{r}})}{\eta}, \\ H & \text{if} \quad \mathrm{PPR} \geq \frac{\delta(\bar{\mathbf{r}})}{\eta}. \end{cases} \tag{14}$$

We can now note that

$$1 \leq \frac{\delta(\bar{\mathbf{r}})}{\eta} \leq \frac{\max[\delta(\mathbf{r})]}{\eta}. \tag{15}$$

The threshold for the PPR region selection will be equal to 1 when $\delta(\bar{\mathbf{r}}) = \eta$, i.e when the maximum of the power response is positioned on the boundary between the two regions. In this case, the amplification of the noise in the high sensitivity region is ininfluent. On the other hand, we have a larger noise amplification when $\delta(\bar{\mathbf{r}}) > \eta$, which is ininfluent on the classification if the source is in H, but might affect it if the source is in L. Therefore, a threshold value larger then 1 has the effect of compensating the amplification of noise due to the sensitivity of the array and to improve the decision on which is the region where the source should be searched.

When the PPR criterion selects the high sensitivity region as the searching region, the source position is estimated as

$$\widehat{\mathbf{r}}_s = \underset{\mathbf{r}\in H}{\operatorname{argmax}}[\phi(\mathbf{r})]. \tag{16}$$

On the other hand, when the PPR criterion indicates to search in the low sensitivity region, the source is localized by searching the maximum of the steered response power, uniformed through the array sensitivity map:

$$\widehat{\mathbf{r}}_s = \underset{\mathbf{r}\in L}{\operatorname{argmax}}\left[\frac{\phi(\mathbf{r})}{\delta(\mathbf{r})}\right]. \tag{17}$$

This equation provides a more robust sound localization in the region L, since it permits to reduce the nonuniform accumulation and the ambiguity that may arise when the maximum value for the L region is positioned close to the boundary of the two regions. Figure 3 illustrates the situation in which the L region maximum is positioned close to the boundary and it is larger than the source maximum (continous line). Equation (17) provides an uniform TDOA accumulation (dotted line) that allows the correct estimation of the source position in this case. The proposed R-SRP increases the localization accuracy in the low sensitivity region keeping an high accuracy in the high sensitivity region due to the accumulation of all TDOA information. Note that by using an uniform steered response power in the overall region, the localization performance in the high sensitivity region considerably degrades,

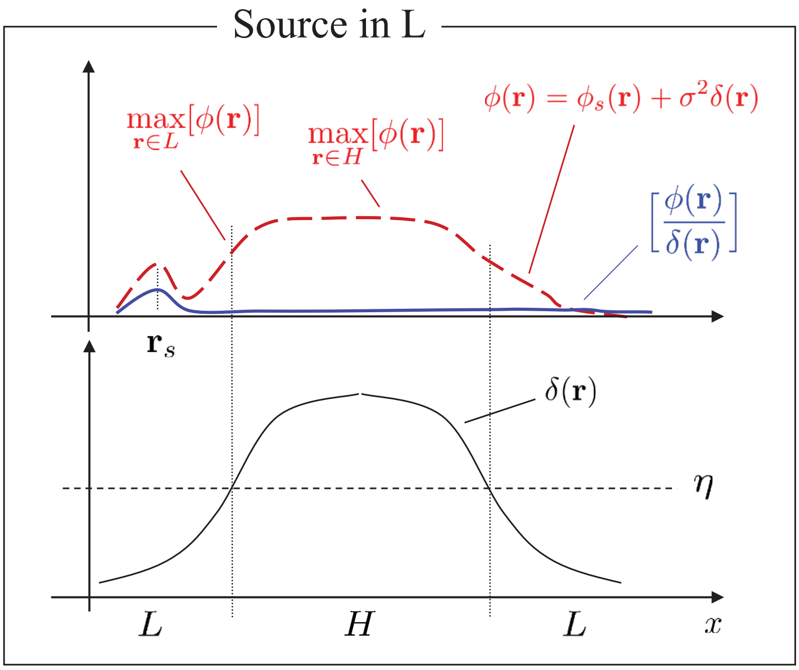

Figure 3: A schematic representation of the SRP profile along x axis when the source is positioned in the low sensitivity region and the L region maximum is not positioned on the source position.

since the mean operation attenuates the TDOA accumulation in the grid points corresponding to the highest number of hyperboloid intersections. An example of uniform steered response power in the overall region was proposed in [23] (I-SRP), in which the normalization allows the reduction of the problem due to nonuniform accumulation. However, it also discards part of the information in the high sensitivity region that can be positively used to improve the localization performance in that region [27].

5. Region Selection Steered Response Power

The implementation of the R-SRP method can be divided in two steps. In the off-line step, the sampled space grid is computed with the GSG method (Algorithm 1) providing the look-up tables ($\gamma_r(q)$, $\gamma_p(q)$, $\gamma_\tau(q)$), linking the all TDOA values of the microphone pairs with the grid positions in space, and the sensitivity function $\delta(\mathbf{r})$. From equation (7), the high- and low-sensitivity regions cn be identified, providing two sets of discrete grid positions, H and L, one for each region. In the on-line step, the G-SRP is computed on a frame-by-frame basis to estimate the source position. For each analysis frame, the R-SRP is computed through the following steps:

1. The values from the estimated GCC-PHAT functions are accumulated in the grid map (4);
2. The maximum values of the SRP for the low and high sensitivity regions are identified, and the PPR is estimated through equation (9);
3. By using the classification criterion in (14), the region selection is computed to estimate the area in which the source is positioned;
4. The source position is finally estimated using (16) or (17), depending on whether it was estimated to lie in the high or in the low sensitivity region.

6. Experimental Results

Experiments for the 2D sound source localization on simulated data and on real-world data are reported. We compare the performance of the proposed R-SRP algorithm, with the following ones: SRP [13], M-SRP [26] , I-SRP [23], H-SRP [24], and G-SRP [27]. Note that we not consider the volumetric refinement steps of I-SRP and H-SRP, since we focus on the evaluation of localization performance with coarser grid. Hence, the same grid resolution was used for all SRP methods. Specifically, the spatial resolution Δ was set to 0.25 m and 0.5 m in two different experiments. We have used a coarser grid since it allows the reduction of the computational cost, and it may be used to compute a further volumetric refinement step for increasing the localization accuracy [23, 24]. Performance is reported in terms of root mean square error (RMSE) and of accuracy rate (AR) for the estimated source that is inside the area surrounding the grid point given by the spatial resolution Δ:

$$\begin{aligned} |\widehat{x}_s(k) - x_s(k)| &\leq \frac{\Delta}{2}, \\ |\widehat{y}_s(k) - y_s(k)| &\leq \frac{\Delta}{2}. \end{aligned} \tag{18}$$

6.1. Simulation

The localization performance has been evaluated with several Monte Carlo simulations, using 100 run trials for each condition test. The image-source model was used to simulate reverberant audio data in room acoustics [28, 29]. A room of $(9 \times 6 \times 3)$ m was used. The tests were conducted with different signal-to-noise ratios (SNRs), which were obtained by adding mutually independent white Gaussian noise to each channel. A randomly distributed sensor array of 8 microphones was used. The room setup, the sensitivity map, and the high and low sensitivity regions with Δ=0.25 m are shown in Figure 4. Both microphones and the source were positioned at a distance from the floor of 1.3 m. A speech signal source was randomly

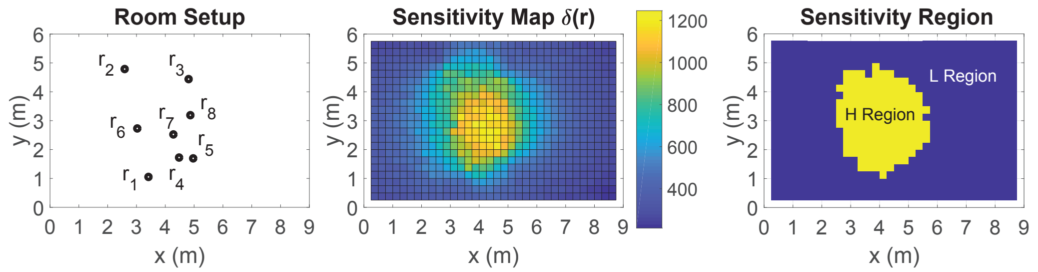

Figure 4: The simulated room with the position of 8 microphones, the sensitivity map and the high and low sensitivity regions with spatial resolution $\Delta = 0.25$ m.

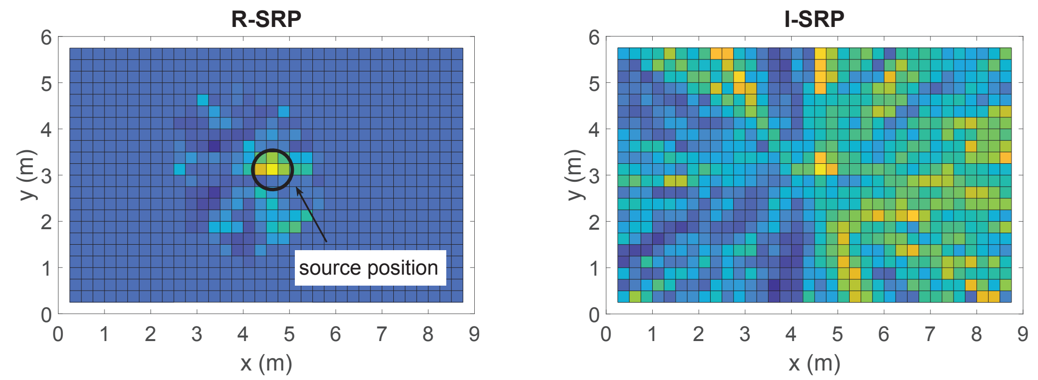

Figure 5: The power response maps for a source positioned in the high sensitivity region for the R-SRP and I-SRP in a frame with $\mathrm{RT}_{60} = 0.7$ s, SNR$=$ 20 dB and spatial resolution $\Delta = 0.25$ m. R-SRP localizes correctly the source position.

located in each trial so that the minimum distance between walls was 0.4 m and the minimum distance between source and microphones was 0.2 m. The sampling frequency was 44.1 kHz and the analysis frame was 8192 samples.

Table 1 and Table 2 report the AR and the RMSE localization performance for the whole search space G_s with spatial resolution $\Delta = 0.25$ m and $\Delta = 0.5$ m, respectively. The reverberant time (RT_{60}) was set to 0.3 s. As it can be observed, the R-SRP algorithm delivers a better performance than other SRP-based methods. We can especially see the improvement due to the region selection operation of the R-SRP in comparison with the G-SRP

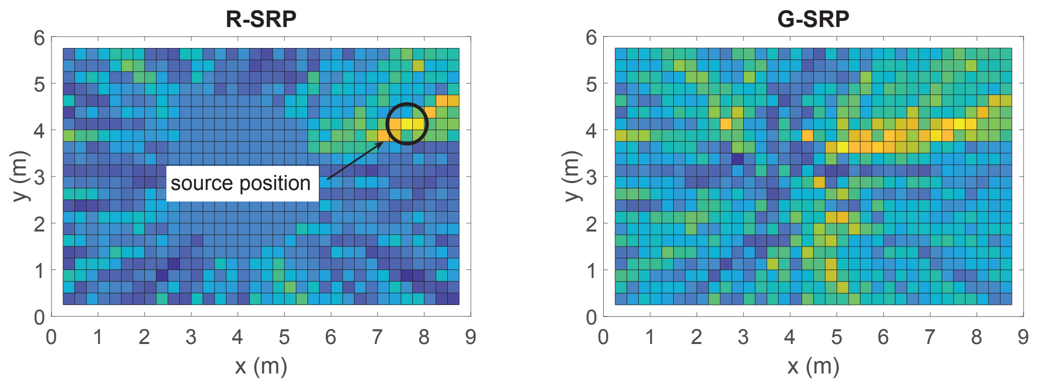

Figure 6: The power response maps for a source positioned in the low sensitivity region for the R-SRP and G-SRP in a frame with $\mathrm{RT}_{60} = 0.7$ s, SNR$= 20$ dB and spatial resolution $\Delta = 0.25$ m. R-SRP localizes correctly the source position.

Table 1: AR (%) and RMSE (m) localization performance in G_s with $\mathrm{RT}_{60} = 0.3$ s and spatial resolution $\Delta = 0.25$ m using simulated data.

SNR (dB)		R-SRP	G-SRP	SRP	M-SRP	I-SRP	H-SRP
20	AR	77.20	77.11	50.75	72.73	65.72	67.79
	RMSE	0.147	0.296	0.628	0.244	0.330	0.261
10	AR	74.00	71.15	48.82	68.16	64.85	64.02
	RMSE	0.356	0.607	0.816	0.526	0.351	0.534
0	AR	63.36	58.82	32.62	58.17	55.99	54.95
	RMSE	0.891	1.135	1.798	1.051	0.928	1.018

when the SNR decreases.

Next, Table 3 and 4 show the results of two simulations with $\Delta = 0.25$ m and $\Delta = 0.5$ m for a RT_{60} of 0.7 s and a SNR of 20 dB. The tables report also the AR an RMSE localization performance for the two regions H_s (high sensitivity) and L_s (low sensitivity). R-SRP outperforms other methods for both spatial resolutions in the overall region G_s. R-SRP has a similar AR and RMSE in the H_s region if compared to other accumulated TDOA methods (G-SRP, M-SRP, H-SRP), and a better AR in the L_s region

Table 2: AR (%) and RMSE (m) localization performance in G_s with $\mathrm{RT}_{60} = 0.3$ s and spatial resolution $\Delta = 0.5$ m using simulated data.

SNR (dB)		R-SRP	G-SRP	SRP	M-SRP	I-SRP	H-SRP
20	AR	77.59	74.24	35.64	74.76	48.19	67.42
	RMSE	0.342	0.562	1.169	0.541	0.825	0.528
10	AR	73.07	68.43	34.90	70.39	51.33	63.85
	RMSE	0.532	0.786	1.384	0.733	0.776	0.709
0	AR	64.28	58.97	23.55	61.73	47.99	56.61
	RMSE	0.946	1.115	2.219	1.022	1.120	0.996

if compared to the I-SRP, that, however, has a better RMSE in the L_s, but it provides a minor localization performance in the H_s region since it uses an uniform steered response power by computing the mean of the accumulated GGC values for each volume. Figures 5 and 6 show the comparison of power response maps, in which we can see the effective correct localization of the source with the R-SRP. In Figure 5, we can observe how the uniform steered response power in the I-SRP reduces the robustness in the high sensitivity region. In Figure 6, we can see the localization improvement due to the proposed region selection. In accordance to [26, 27], the conventional SRP degrades the localization accuracy when a coarser grid is used due to the loss of information of GCC functions, which are not linked with any grid position.

6.2. Real Data

Real-world tests have been computed in a room of dimensions $(6.4 \times 3 \times 3.6)$ m, and a RT_{60} of 0.6 s. A grid resolution Δ of 0.25 m was used for all SRP methods. A distributed array of 6 microphones was positioned with a distance from the floor of 0.88 m. Four source positions have been considered: $\mathbf{s}_1$ and $\mathbf{s}_4$ located in the low sensitivity region, and $\mathbf{s}_2$ and $\mathbf{s}_3$ located in

Table 3: AR (%) and RMSE (m) with $RT_{60} = 0.7$ s, SNR= 20 dB and spatial resolution $\Delta = 0.25$ m using simulated data.

Region		R-SRP	G-SRP	SRP	M-SRP	I-SRP	H-SRP
G_s	AR	67.46	60.53	35.11	58.55	54.59	52.82
	RMSE	0.948	1.445	1.746	1.297	1.017	1.302
L_s	AR	44.30	30.86	29.86	29.01	36.06	24.69
	RMSE	1.299	2.021	1.790	1.813	0.958	1.821
H_s	AR	90.621	90.197	40.363	88.090	73.121	80.95
	RMSE	0.329	0.309	1.702	0.280	1.073	0.279

Table 4: AR (%) and RMSE (m) with $RT_{60} = 0.7$ s, SNR= 20 dB and spatial resolution $\Delta = 0.5$ m using simulated data.

Region		R-SRP	G-SRP	SRP	M-SRP	I-SRP	H-SRP
G_s	AR	62.48	55.01	23.32	58.00	37.62	51.80
	RMSE	1.122	1.539	2.135	1.316	1.340	1.286
L_s	AR	42.51	27.09	25.37	29.92	38.43	26.77
	RMSE	1.433	2.112	2.242	1.815	1.060	1.766
H_s	AR	82.45	82.93	21.27	86.07	36.81	76.83
	RMSE	0.682	0.523	2.022	0.414	1.572	0.433

the high sensitivity region. A source speech signal was reproduced with a loudspeaker at each position. Figure 7 depicts the room setup, the sensitivity map, and the sensitivity regions calculated with the GSG algorithm with $\Delta = 0.25$ m. The result of localization performance are reported in Table 5 for the whole search space G_r. We can observe that the R-SRP algorithm outperforms the other SRP methods, providing an accuracy rate of about 36% whereas the others reach a 26% accuracy rate at best.

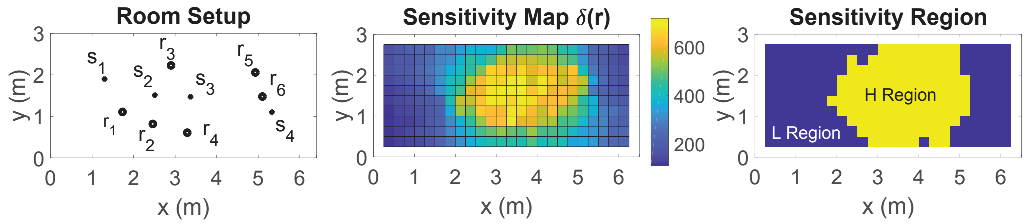

Figure 7: The real room with the position of 6 microphones and 4 sources, the sensitivity map, and the high and low sensitivity regions with spatial resolution $\Delta = 0.25$ m.

Table 5: AR (%) and RMSE (m) with spatial resolution $\Delta = 0.25$ m using real data with RT_{60}=0.6 s.

Region		R-SRP	G-SRP	SRP	M-SRP	I-SRP	H-SRP
G_r	AR	36.09	23.13	10.681	26.318	23.272	25.181
	RMSE	1.334	1.671	1.788	1.576	1.818	1.824

7. Conclusions

A sensitivity-based region selection method for the SRP-PHAT using GSG accumulated TDOA functions was presented. The proposed R-SRP is based on a definition of a PPR between high and low sensitivity regions calculated by the GSG algorithm. A classification criterion taking into account the sensitivity map was formulated. Our experiments demonstrate that the error of localization can be reduced especially when the source is positioned in the low sensitivity region.

[1] B. Laufer-Goldshtein, R. Talmon, S. Gannot, Semi-supervised sound source localization based on manifold regularization, IEEE/ACM Transactions on Audio, Speech, and Language Processing 24 (8) (2016) 1393–1407.

[2] D. Salvati, C. Drioli, G. L. Foresti, A weighted MVDR beamformer

based on SVM learning for sound source localization, Pattern Recognition Letters 84 (2016) 15–21.

[3] L. Kumar, R. M. Hegde, Near-field acoustic source localization and beamforming in spherical harmonics domain, IEEE Transactions on Signal Processing 64 (13) (2016) 3351–3361.

[4] D. Yook, T. Lee, Y. Cho, Fast sound source localization using two-level search space clustering, IEEE Transactions on Cybernetics 46 (1) (2016) 20–26.

[5] D. Salvati, C. Drioli, G. L. Foresti, Sound source and microphone localization from acoustic impulse responses, IEEE Signal Processing Letters 23 (10) (2016) 1459–1463.

[6] L. Petrica, An evaluation of low-power microphone array sound source localization for deforestation detection, Applied Acoustics 113 (2016) 162–169.

[7] C. Knapp, G. Carter, The generalized correlation method for estimation of time delay, IEEE Transactions on Acoustics, Speech, and Signal Processing 24 (4) (1976) 320–327.

[8] J. Benesty, Adaptive eigenvalue decomposition algorithm for passive acoustic source localization, Journal of the Acoustical Society of America 107 (1) (2000) 384–391.

[9] J. O. Smith, J. S. Abel, Closed-form least-squares source location estimation from range-difference measurements, IEEE Transactions on Acoustics, Speech, and Signal Processing 35 (12) (1987) 1661–1669.

[10] Y. Huang, J. Benesty, G. W. Elko, R. M. Mersereau, Real-time passive source localization: a practical linear-correction least-squares approach, IEEE Transactions on Speech and Audio Processing 9 (8) (2001) 943–956.

[11] P. Stoica, J. Li, Source localization from range-difference measurements, IEEE Signal Processing Magazine 23 (3) (2006) 63–66.

[12] M. Omologo, P. Svaizer, R. De Mori, Spoken Dialogue with Computers, Academic Press, 1998, Ch. Acoustic Transduction.

[13] J. H. DiBiase, H. F. Silverman, M. S. Brandstein, Microphone Arrays: Signal Processing Techniques and Applications, Springer, 2001, Ch. Robust localization in reverberant rooms.

[14] P. Pertilä, T. Korhonen, A. Visa, Measurement combination for acoustic source localization in a room environment, EURASIP Journal on Audio, Speech, and Music Processing 2008 (2008) 1–14.

[15] R. O. Schmidt, Multiple emitter location and signal parameter estimation, IEEE Transactions on Antennas and Propagation 34 (3) (1986) 276–280.

[16] R. Roy, T. Kailath, ESPRIT - estimation of signal parameters via rotational invariance techniques, IEEE Transactions on Acoustics, Speech, and Signal Processing 37 (7) (1989) 984–995.

[17] B. D. Rao, K. V. S. Hari, Performance analysis of root-music, IEEE Transactions on Acoustics, Speech, and Signal Processing 37 (12) (1989) 1939–1949.

[18] K. Harmanci, J. Tabrikian, J. L. Krolik, Relationships between adaptive minimum variance beamforming and optimal source localization, IEEE Transactions on Signal Processing 48 (1) (2000) 1–12.

[19] C. Zhang, D. Florencio, D. E. Ba, Z. Zhang, Maximum likelihood sound source localization and beamforming for directional microphone arrays in distributed meetings, IEEE Transactions on Multimedia 10 (3) (2008) 538–548.

[20] J. Traa, D. Wingate, N. D. Stein, P. Smaragdis, Robust source localization and enhancement with a probabilistic steered response power model, IEEE/ACM Transactions on Audio, Speech, and Language Processing 24 (3) (2016) 493–503.

[21] M. S. Bartlett, Smoothing periodograms from time-series with continuous spectra, Nature 161 (1948) 686–687.

[22] D. Salvati, C. Drioli, G. L. Foresti, Incoherent frequency fusion for broadband steered response power algorithms in noisy environments, IEEE Signal Processing Letters 21 (5) (2014) 581–585.

[23] A. Marti, M. Cobos, J. J. Lopez, J. Escolano, A steered response power iterative method for high-accuracy acoustic source localization, Journal of the Acoustical Society of America 134 (4) (2013) 2627–2630.

[24] L. O. Nunes, W. A. Martins, M. V. S. Lima, L. W. P. Biscainho, M. V. M. Costa, F. M. Gonalves, A. Said, B. Lee, A steered-response power algorithm employing hierarchical search for acoustic source local-

ization using microphone arrays, IEEE Transactions on Signal Processing 62 (19) (2014) 5171–5183.

[25] M. V. S. Lima, W. A. Martins, L. O. Nunes, L. W. P. Biscainho, T. N. Ferreira, M. V. M. Costa, B. Lee, A volumetric SRP with refinement step for sound source localization, IEEE Signal Processing Letters 22 (8) (2015) 1098–1102.

[26] M. Cobos, A. Marti, J. J. Lopez, A modified SRP-PHAT functional for robust real-time sound source localization with scalable spatial sampling, IEEE Signal Processing Letters 18 (1) (2011) 71–74.

[27] D. Salvati, C. Drioli, G. L. Foresti, Exploiting a geometrically sampled grid in the steered response power algorithm for localization improvement, Journal of the Acoustical Society of America 141 (1) (2017) 586–601.

[28] J. B. Allen, D. A. Berkley, Image method for efficiently simulating small-room acoustics, Journal of the Acoustical Society of America 65 (4) (1979) 943–950.

[29] E. Lehmann, A. Johansson, Prediction of energy decay in room impulse responses simulated with an image-source model, Journal of the Acoustical Society of America 124 (1) (2008) 269–277.