GREEDY SPARSE ARRAY DESIGN FOR OPTIMAL LOCALIZATION UNDER SPATIALLY PRIORITIZED SOURCE DISTRIBUTION

Yotam Gershon *Yaakov Buchris* *Israel Cohen*

Andrew and Erna Viterby Faculty of Electrical Engineering
Technion – Israel Institute of Technology, Technion City, Haifa 3200003, Israel

ABSTRACT

A common approach for acoustic source localization is based on finding the maximum of a spatial cost function, such as the steered response power (SRP) function. The shape of the SRP highly depends on the constellation of sensors within the array layout, and have a direct impact on the performance. Thus, an array may be specifically designed to produce high localization performance and small error regions, especially when a spatially prioritized source location distribution function is taken into account. We introduce a new measure called power spread, which quantifies the localization error region. Then, we propose a greedy algorithm for a sparse array design, aiming to minimize the power spread for optimal localization error region in a given area of interest. Simulations demonstrate that the proposed design, compared to standard linear array design and random array design, obtains superior performance in terms of power spread and localization error, with a reasonable computational effort.

Index Terms— Source localization, greedy sparse design

1. INTRODUCTION

A common approach in acoustic source localization (ASL) [1, 2, 3, 4, 5] is based on defining a spatial cost function that is calculated from the signals received at different synchronized microphones, and searching for a spatial point in which this function obtains its maximal value. A well-known such function, called steered response power (SRP) [6, 7, 8, 9, 10], calculates the output power of a delay-and-sum (DAS) beamformer matched to each point in space. The properties of this power function, and especially the sharpness of its peak in the source's vicinity, highly depends on the positioning of microphones within the array, and have a significant effect on the size of localization error regions in noisy environments. Furthermore, these properties varies significantly for different source locations. Thus, an array design for optimized localization performance in a certain area of interest, must take into account the SRP properties produced by the array for sources in that area. A previous work dealing with the error based optimization of sensor positioning was proposed in [11], in which an error predictor is estimated by the delay variance of each sensors pair, and the array design is done by a Monte Carlo approach for minimizing the average localization standard deviation in particular positions. We propose a new measure for quantifying the error region which called power spread. This measure enables an evaluation of the array expected localization performance in terms of the likely error region size. Based on this measure, a greedy algorithm for a sparse array design is developed, aiming to minimize the power spread around the true location peak, thus optimizing the expected localization performance, for a set of randomly chosen source locations drawn from a spatially prioritization distribution function.

This work was supported by the Israel Science Foundation (grant no. 576/16) and the ISF-NSFC joint research program (grant No. 2514/17).

2. ASL USING SRP

2.1. Signal Model

For simplicity, we may focus on the 2-D localization case, although the models and algorithms in this paper may easily be generalized to the 3-D case. We consider $s(t)$ to be an acoustic signal generated from a point source at location $\mathbf{r}_\mathrm{s} \in \mathbb{R}^2$. The signal is received in an array of M synchronized microphones, with locations $\{\mathbf{r}_m\}_{m=1}^M \in \mathbb{R}^2$. The mathematical model for the mth microphone signal after propagation is

$$x_m(t) = s(t - \tau_m), \tag{1}$$

where $\tau_m = ||\mathbf{r}_\mathrm{s} - \mathbf{r}_m||/c$ is the propagation delay from the source to the mth microphone, c is the speed of sound and $||\cdot||$ is the euclidean norm. By arbitrarily setting the first microphone to be a reference, the model can be rewritten as [12]

$$y_m(t) = x_1(t - \tau_{m1}) + n_m(t), \tag{2}$$

where $\tau_{m1} = \tau_m - \tau_1 = [||\mathbf{r}_\mathrm{s} - \mathbf{r}_m|| - ||\mathbf{r}_\mathrm{s} - \mathbf{r}_1||]/c$ is the difference between propagation delays from the source to the mth microphone and to the first microphone, and $n_m(t)$ is a noise term assumed to be a statistically independent white gaussian noise with uniform variance, i.e,

$$\mathbf{n}(t) = [n_1(t), \cdots, n_M(t)]^T \sim \mathcal{N}(\underline{0}, \sigma^2 I_M). \tag{3}$$

2.2. Steered Response Power

Given a hypothesized source location $\mathbf{r} \in \mathbb{R}^2$ it is immediate to calculate the corresponding propagation delays

$\{\tau_m(\mathbf{r})\}_{m=1}^{M}$. The SRP of location $\mathbf{r}$ is defined as the output power of a DAS beamformer that matches this location, i.e.

$$P_{\mathrm{SRP}}(\mathbf{r}) = \int_{-\infty}^{\infty} \left| \sum_{m=1}^{M} y_m(t - \tau_m(\mathbf{r})) \right|^2 dt. \quad (4)$$

By transforming to the frequency domain, replacing the order of summation and integration, and taking a desired frequency range, Eq. (4) can be written as [7],[12]

$$P_{\mathrm{SRP}}(\mathbf{r}) = \int_{f_{\min}}^{f_{\max}} \sum_{k=1}^{M} \sum_{l=1}^{M} Y_k(f) Y_l^*(f) e^{j2\pi f \tau_{kl}(\mathbf{r})} df, \quad (5)$$

where $(\cdot)^*$ represents complex conjugate and $Y_m(f) = e^{-j2\pi f \tau_{m1}} X_1(f) + N_m(f)$. The SRP function is maximized when the hypothesized location coincides with the actual source location (i.e. $\mathbf{r} = \mathbf{r}_{\mathrm{s}}$). Thus, the localization problem can be formulated as

$$\hat{\mathbf{r}}_{\mathrm{s}} = \arg\max_{\mathbf{r}} P_{\mathrm{SRP}}(\mathbf{r}). \quad (6)$$

An efficient and elegant method for calculating the SRP function by means of geometric projection of $M-$dimensional complex vectors is introduced in [12].

3. POWER SPREAD AND ERROR REGION

3.1. Array Effect on Error Region

In a noisy environment, the SRP function blurs and the identification of the maximum position becomes vulnerable to errors, which is highly effected by the shape of the SRP function in the source's vicinity. An example to this effect is encountered in distant source localization using a uniform linear array (ULA), for which the acoustic wavefront is assumed to be planar upon reaching the array, resulting in the steering vector

$$\mathbf{d}_{\mathrm{ULA}}(\mathbf{r}_{\mathrm{s}}, f) \approx \left[1, e^{-j2\pi \frac{\delta}{c} \sin\theta}, \cdots, e^{-j2\pi (M-1) \frac{\delta}{c} \sin\theta}\right], \quad (7)$$

where δ and θ are the array spacing and angle to the source, respectively. The distance independent steering vector produces a large error region along the θ axis. Using numerous smaller ULAs with different deployment angles, the error region can be significantly reduced, as illustrated in Fig. 1.

3.2. Power Spread and Distribution Ellipse

A quantification of the error region size is needed to analyze the effect of the array constellation. Inspired by the confidence ellipse concept [13, 14, 15], herein we introduce the power spread quantity, which is a measurement for the error region. The idea is to estimate an ellipse of 2 standard deviations in each orthogonal axis of high power values, describing the significant probability mass of source location estimation, and then to quantify its size. Unlike a confidence ellipse of localization error, calculated after multiple localizations of a single source, our proposed ellipses are based on the spatial power function that the array induces. Given $P_{\mathrm{SRP}}(\mathbf{r})$, the maximal power value $\mathcal{M}$ is found, and a threshold value is set as a predefined percentage of the maximum (e.g. $\mathcal{T} = 0.9\mathcal{M}$). All grid points $\mathbf{r} = (x, y)$ in which the power is greater than the threshold are stacked into vectors,

$$(\underline{x}, \underline{y}) = \{(x_i, y_i), i = 1, 2, ... : P_{\mathrm{SRP}}(x_i, y_i) > \mathcal{T}\}. \quad (8)$$

The covariance matrix of the joint set $(\underline{x}, \underline{y})$ is

$$\mathcal{C}_{\underline{x},\underline{y}} = \frac{1}{N-1} \sum_{i=1}^{N} \begin{bmatrix} (x_i - \bar{x})^2 & (x_i - \bar{x})(y_i - \bar{y}) \\ (x_i - \bar{x})(y_i - \bar{y}) & (y_i - \bar{y})^2 \end{bmatrix}, \quad (9)$$

where N is the number of elements and $\bar{x}, \bar{y}$ are the arithmetic averages of $\underline{x}, \underline{y}$ respectively. The eigenvalue decomposition of the covariance matrix is $\mathcal{C}_{\underline{x},\underline{y}} = U \Sigma U^{-1}$, where $U = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$, and $\sigma_1 \geq \sigma_2$ so that $\underline{u}_1 = [u_{11}, u_{21}]^T$ is defined as the principle axis. The distribution ellipse of the set is defined as

$$\begin{bmatrix} x_{\mathrm{e}}(t) \\ y_{\mathrm{e}}(t) \end{bmatrix} = \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} + \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} x_{\mathrm{c}}(t) \\ y_{\mathrm{c}}(t) \end{bmatrix}, \quad (10)$$

where $\alpha = \tan^{-1}\left(\frac{u_{11}}{u_{21}}\right)$, and

$$[x_{\mathrm{c}}(t), y_{\mathrm{c}}(t)] = [2\sigma_1 \cos(t), 2\sigma_2 \sin(t)], t \in [0, 2\pi]. \quad (11)$$

The power spread is defined as the area of the distribution ellipse, i.e.

$$S = 4\pi\sigma_1\sigma_2. \quad (12)$$

The power function $P_{\mathrm{SRP}}(\mathbf{r})$, and thus the power spread, depend on both the source and the microphone locations. Thus, the power spread can be written as $S(\mathbf{r}_{\mathrm{s}}, \{\mathbf{r}_m\}_{m=1}^{M})$. Figure 1 illustrates these quantities.

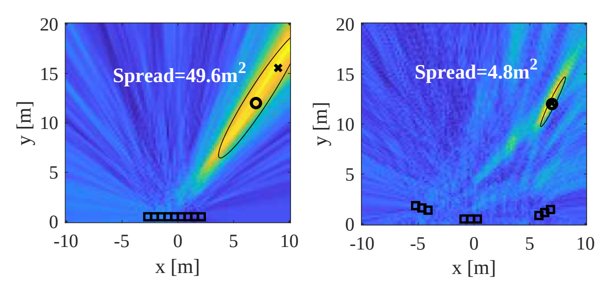

Fig. 1. *SRP function (colormap) produced for a given true source location (black circle), and a nine microphones array (black squares), as a single ULA (left) and as three distributed subarrays (right). The estimated distribution ellipse (black line) and the estimated source location (black cross) are also presented.*

4. SPARSE ARRAY DESIGN ALGORITHM

4.1. Motivation

When designing a microphone array for localization, the main performance measure is the error region. The array is often constrained to a certain size, number of microphones and region of deployment. Moreover, in many cases there is some kind of area of interest for which the localization performance is more important. An example of such a case may occur when the array placement is constrained to a corner of some room, and the main area of interest is the center of this room (which corresponds to a certain range of angles). This area of interest will be described by a spatial priority function. We hereby propose an algorithm for a greedy sparse array design, subject to a given spatial priority function, that iteratively selects microphones which statistically maximize the localization performance.

4.2. Algorithm

We begin with a spatial priority function $0 \leq W(\mathbf{r}) \leq 1$, where lower values indicate low priority for sources located in the corresponding locations, and vice versa. Based on $W(\mathbf{r})$, we then randomly choose N_{r_s} source locations, each located in $\mathbf{r}_\mathrm{s}(i), i \in \{1, 2, ..., N_{r_\mathrm{s}}\}$. We define a grid of M potential microphone locations $\{\mathbf{r}_\mathrm{g}(j)\}_{j=1}^{M}$, determined according to given physical deployment constraints. The algorithm's goal is to choose $K \ll M$ microphone locations that maximize the average localization performance for all the chosen N_{r_s} source locations, simultaneously. This is done iteratively, where in each step $k \in \{1, ..., K\}$ we begin with a fixed array $\{\mathbf{r}_\mathrm{a}\}^{(k-1)}$ of $k-1$ microphones from the predefined grid, and add a new microphone from the remaining grid, such that the average power spread over all sources locations $\mathbf{r}_\mathrm{s}(i), i \in \{1, 2, ..., N_{r_\mathrm{s}}\}$, is minimized for the next step. The array is initialized with the first microphone $\{\mathbf{r}_\mathrm{a}\}^{(1)} = \mathbf{r}_\mathrm{g}(j_1)$, where $j_1 \in \{1, 2, ..., M\}$ is chosen deterministically or randomly. We also initialize an array containing the remaining (unused) grid indices, i.e $J_\mathrm{rem} = \{1, 2, ..., M\} \backslash \{j_1\}$. In each following step $k \in \{2, 3, ..., K\}$, we initialize a histogram $\mathbf{h}^{(k)}$ as an all-zeros vector of size M. We then iterate over all source locations $\mathbf{r}_\mathrm{s}(i), i \in \{1, ..., N_{r_\mathrm{s}}\}$. For each location $\mathbf{r}_\mathrm{s}(i)$ we define an area A_i as a square of size $D_x \times D_y[\mathrm{m}]$ with $\mathbf{r}_\mathrm{s}(i)$ as its center. We temporarily add each of the remaining microphones to the previous array, i.e. $\{\mathbf{r}_\mathrm{a}\}^{(k-1,j)} = \{\mathbf{r}_\mathrm{a}\}^{(k-1)} \cup \mathbf{r}_\mathrm{g}(j), j \in J_\mathrm{rem}$, and calculate the SRP in the area A_i and its corresponding power spread, $S_{i,j} \triangleq S(\mathbf{r}_\mathrm{s}(i), \{\mathbf{r}_\mathrm{a}\}^{(k-1,j)})$. We then may search for the index $\hat{j} \in J_\mathrm{rem}$ for which the power spread $S_{i,j}$ is minimized, and increase the histogram $\mathbf{h}^{(k)}$ value in the index $\hat{j}$, thus counting the corresponding new microphone choosing occurrences over different source locations. After going through all source locations, we search for the index $\hat{l}$ of the histogram's maximal value (or the lowest indexed maximal value if there is a tie), and we update the designed array by adding the microphone of corresponding index, i.e. $\{\mathbf{r}_\mathrm{a}\}^{(k)} = \{\mathbf{r}_\mathrm{a}\}^{(k-1)} \cup \mathbf{r}_\mathrm{g}(\hat{l})$.

We also remove this index from the remaining grid, i.e. $J_\mathrm{rem} = J_\mathrm{rem} \backslash \{\hat{l}\}$. The new array locally minimizes the average power spread over all source locations, with respect to all possible new microphones added to the previous array. At the end of K iterations, we get the array $\{\mathbf{r}_\mathrm{a}\}^{(K)}$, with K microphones. The algorithm is described in Alg. 1.

Algorithm 1 Minimal Spread Algorithm (MSA)

Input: $W(\mathbf{r})$, K, $\{\mathbf{r}_g\}$, N_{r_s}
Output: array constellation $\{\mathbf{r}_a\}^{(K)}$

Randomly choose N_{r_s} source locations $\{\mathbf{r}_s\}$
Initialize array: $\{\mathbf{r}_a\}^{(1)} = \mathbf{r}_g(j_1), j_1 \in 1, 2, ..., M$
Initialize remaining grid: $J_\mathrm{rem} = \{1, 2, ..., M\} \backslash \{j_1\}$
for $k = 2$ to K **do**
  Initialize histogram $\mathbf{h}^{(k)}$ as an all-zeros vector of size M
  for $i = 1$ to N_{r_s} **do**
    Retrieve: $\mathbf{r}_s(i) = (x_s^i, y_s^i) \in \{\mathbf{r}_s\}$
    Define source area:
$$A_i = \{\mathbf{r} = (x, y) : |x - x_s^i| \leq D_x/2, |y - y_s^i| \leq D_y/2\}$$
    for $j \in J_\mathrm{rem}$ **do**
      Retrieve: $\mathbf{r}_g(j)$
      Append: $\{\mathbf{r}_a\}^{(k-1,j)} = \{\mathbf{r}_a\}^{(k-1)} \cup \mathbf{r}_g(j)$
      Calculate the SRP within area A_i
      Estimate the power spread: $S_{i,j} \triangleq S(\mathbf{r}_\mathrm{s}(i), \{\mathbf{r}_\mathrm{a}\}^{(k-1,j)})$
    end for
    Find minimal: $\hat{j} = \arg\min_j \{S_{i,j}\}$
    Increase histogram: $\mathbf{h}^{(k)}(\hat{j}) = \mathbf{h}^{(k)}(\hat{j}) + 1$
  end for
  Find highest historgram value: $\hat{l} = \arg\max_l \{\mathbf{h}^{(k)}(l)\}$
  Update current array: $\{\mathbf{r}_a\}^{(k)} = \{\mathbf{r}_a\}^{(k-1)} \cup \mathbf{r}_\mathrm{g}(\hat{l})$
  Update remaining grid: $J_\mathrm{rem} = J_\mathrm{rem} \backslash \{\hat{l}\}$
end for

5. SIMULATIONS AND RESULTS

5.1. Simulations

The performance analysis of our algorithm was done with dedicated simulations in MATLAB environment. A clean speech signal recorded at 8kHz was taken as the source's signal, and the received signals were simulated by a suitable propagation delay from the source location, $\mathbf{r}_\mathrm{s}$, to each microphone with addition of a white gaussian noise. The signals were processed in the STFT domain, calculated with non-overlapping 64ms (512 samples) hamming windows, as done in [12]. The simulation parameters were chosen as follows: microphone array grid of 9×3 points, with $1.5\mathrm{m} \times 1\mathrm{m}$ spacing between adjacent microphones, for a total number of $M = 27$ points in a range of $12\mathrm{m} \times 2\mathrm{m}$, source area of $D_x = D_y = 4\mathrm{m}$, power spread threshold of $\mathcal{T} = 0.85 \max\{P\}$, resolution of SRP calculation of $0.24\mathrm{m} \times 0.2\mathrm{m}$, 0[dB] SNR, and $N_{r_\mathrm{s}} = 10$ random source locations. The bottom middle microphone was set as the spatial point $(0, 0)$m, thus defining the coordinates system. For spatial priority functions, two cases were examined: (1) Uniform priority over $[-12, 12]\mathrm{m} \times [5, 20]\mathrm{m}$, and (2) angle priority, with $\theta \sim N(62, 20)[°]$, $R \sim U[5, 20]\mathrm{m}$. This geometry is illustrated in Fig. 2. The uniform priority generalize the problem to a non-prioritized array design, while the angle priority exemplifies a non-trivial, reasonable area of interest. The algorithm was compared with two opponents: (1) Randomly chosen array, (2) uniform linear array, through two relevant criteria: (1) Average power spread $[\mathrm{m}^2]$, (2) average estimation error $(||\hat{\mathbf{r}}_\mathrm{s} - \mathbf{r}_\mathrm{s}||)\,[\mathrm{m}]$. The random array was averaged over $N_{\text{rand}} = 10$ different array randomizations, in which it remained unchanged for all source locations. For the ULA, we began with the center microphone and added an adjacent microphone from left and right alternatively. This entire process was repeated for $N_{\text{rep}} = 10$ array design repetitions, and the average performance was examined.

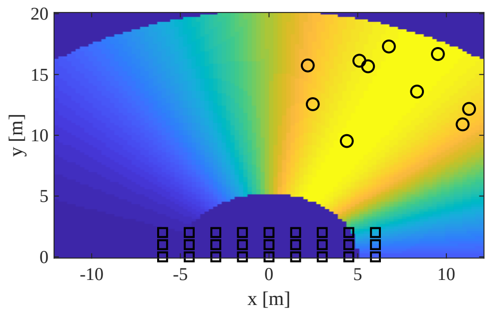

Fig. 2. *Angle spatial priority function (colormap), with the $M = 27$ microphone locations grid (black squares) and $N_{r_s} = 10$ randomly chosen locations (black circles).*

5.2. Results

The algorithm running time for the described parameters was measured to 35.1 minutes on an Intel i5 dual core PC, with 8GB RAM, and the results are hereby presented. An example of the source locations randomization for angle priority, with respect to the M microphone locations grid, is illustrated in Fig. 2. Histograms of microphones index choosing for the first two steps are illustrated in Fig. 3. It is seen that in each step there is a preferred microphone which reduces the average power spread the most.

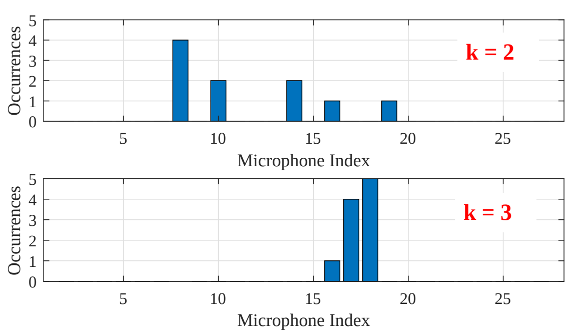

Fig. 3. *Histograms of microphone choosing based on the minimal power spread, for the first two steps. The microphone that is added to the array at each step is that with the highest histogram value.*

The distribution ellipses that are calculated in the first three steps are presented in Fig. 4, for the three compared design approaches. The steepest reduction in the power spread is achieved using the MSA. The algorithm performance with comparison to linear and random designs is shown for the angle and uniform priority distributions in Fig. 5. For any number $k \leq K = 9$ of microphones, the MSA outperforms linear and random designs in both power spread and average estimation error (when only the power spread is directly optimized by the algorithm), for both priority functions. The MSA advantage in the uniform case shows its effectiveness when there is no area prioritization as well.

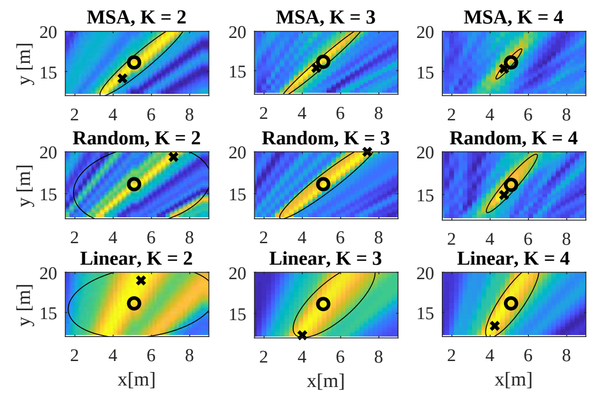

Fig. 4. *SRP in the source's area A_1, for the first 3 steps (left to right) of the MSA, random and linear designs (top to bottom). The distribution ellipse (black line) is shown along with the true (black circle) and estimated (black cross) locations.*

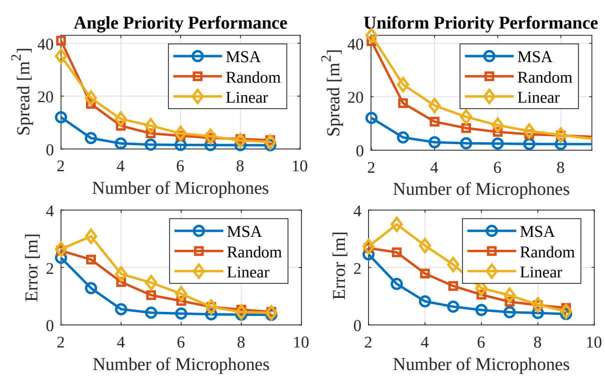

Fig. 5. *Average power spread and estimation error as a function of microphones number, for the MSA, random and linear designs - for angle (left) and uniform (right) spatial priority functions.*

6. CONCLUSIONS

In this paper, we introduced a method for quantifying an error region size in acoustic source localization, called power spread. We then used this quantity in a new greedy algorithm for sparse array design, which minimizes the average error region in a given area of interest described by a spatial priority function. The proposed algorithm was compared to linear and random array designs, and demonstrated a significant advantage in terms of error region and average estimation error.

7. REFERENCES

[1] M. Brandstein, J. Adcock, and H. Silverman, "A closed-form method for finding source locations from microphone-array time-delay estimates," *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1995.

[2] X. Sheng and Y. Hu, "Energy based acoustic source localization," *IEEE Conference on Information Processing in Sensor Networks*, 2003.

[3] A. Tellakula, "Acoustic source localization using time delay estimation," *M.Sc. Thesis, Indian Institute of Science*, 2007.

[4] R. Levorato and E. Pagello, "Doa acoustic source localization in mobile robot sensor networks," *IEEE International Conference on Autonomous Robot Systems and Competitions*, 2015.

[5] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," *J. Acoust. Soc. Am., Vol 107*, 2000.

[6] M. Omologo and P. Svaizer, "Acoustic event localization using a crosspower-spectrum phase based technique," *IEEE International Conference on Acoustics, Speech and Signal Processing*, p. 273–276, 1994.

[7] H. DiBiase, "A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays," *Ph.D. Thesis, Brown University*, 2000.

[8] M. Lima, M. Wallace, L. Nunes, L. Biscainho, T. Ferreira, C. Mauricio, and B. Lee, "Efficient steered-response power methods for sound source localization using microphone arrays," *arXiv, 1407.2351v3*, 2015.

[9] J. Dmochowski, J. Benesti, and S. Affes, "A generalized steered response power method for computationally viable souce localization," *IEEE Transactions on Audio, Speech and Language Processing*, 2007.

[10] H. Do, H. Silverman, and Y. Yu, "A real-time srp-phat source location implementation using stochastic region contraction (src) on a large-aperture microphone array," *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2007.

[11] M. Brandstein, J. Adcock, and H. Silverman, "A localizations-error-based method for microphone-array design," *IEEE In Proceedings of ICASSP96*, 1996.

[12] T. Long, J. Chen, G. Huang, J. Benesty, and I. Cohen, "Acoustic source localization based on geometric projection in reverberant and noisy environments," *IEEE J. on Selected Topics in Signal Process*, 2018.

[13] E. Beh, "Elliptical confidence regions for simple correspondence analysis," *Journal of Statistical Planning and Inference 140*, 2010.

[14] M. Cadoret and F. Husson, "Construction and evaluation of confidence ellipses applied at sensory data," *Food Quality and Preference, Vol. 28*, 2013.

[15] W. Hoover, "Algorithms for confidence circles and ellipses," *NOAA Technical Report, NOS 107, C&GS 3*, 1984.