INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2013, VOL. 59, NO. 4, PP. 383–390
Manuscript received November 4, 2013; revised December, 2013. DOI: 10.2478/eletel-2013-0047

Optimized Acoustic Localization with SRP-PHAT for Monitoring in Distributed Sensor Networks

Sergei Astapov, Julia Berdnikova, and Jürgo-Sören Preden

***Abstract*—Acoustic localization by means of sensor arrays has a variety of applications, from conference telephony to environment monitoring. Many of these tasks are appealing for implementation on embedded systems, however large dataflows and computational complexity of multi-channel signal processing impede the development of such systems. This paper proposes a method of acoustic localization targeted for distributed systems, such as Wireless Sensor Networks (WSN). The method builds on an optimized localization algorithm of Steered Response Power with Phase Transform (SRP-PHAT) and simplifies it further by reducing the initial search region, in which the sound source is contained. The sensor array is partitioned into sub-blocks, which may be implemented as independent nodes of WSN. For the region reduction two approaches are handled. One is based on Direction of Arrival estimation and the other – on multilateration. Both approaches are tested on real signals for speaker localization and industrial machinery monitoring applications. Experiment results indicate the method's potency in both these tasks.**

***Keywords*—Acoustic localization, wireless sensor networks, direction of arrival, multilateration, SRP-PHAT**

I. Introduction

In recent years acoustic signal analysis has grown in popularity in environment monitoring applications. Acoustic signal analysis has a wide area of application because one-dimensional audio signals are relatively easy to process, they are highly comprehensive without additional manipulations, and because acoustic signal acquisition does not require either full direct sight of view of the monitored object, or sufficient highlighting. On the other hand, acoustic signals are prone to noise pollution, especially in unconfined environments, where ambient noise variance and the nature of different background noises are undetermined. For single-sensor solutions, noise poses a great problem because these solutions are unable to efficiently filter unknown noise types [1]. However, the situation changes radically if the acoustic sensors are used in array configurations. Multi-sensor solutions enable concentrating on a specific region of the monitored area and consequently filtering the sound incoming from that region alone. Another application of sensor array solutions lies in sound source localization and tracking.

There exists a variety of methods for acoustic localization, e.g. [2]–[5]. These methods are all based on simple principles of acoustic wave propagation. Having several sensors set in a specific configuration, the direction and distance to the sound source can be estimated by the time delays of wave arrival to these sensors, also called Time Difference of Arrival (TDOA). The Direction of Arrival (DOA) can be estimated from the TDOA or by other methods, as for example in the MUSIC algorithm [4]. For our application we employ the localization method of Steered Response Power with Phase Transform (SRP-PHAT). The method is established to be robust and tolerant to both noise and acoustic reverberation.

One of the main problems related to acoustic localization methods is high computational complexity. Multi-channel signal processing requires large amounts of computational resources for real-time operation. The significantly reduced resources of embedded hardware of Wireless Sensor Networks (WSN) aggravate the situation. Furthermore, for WSN the amounts of data exchanged between nodes must also be maximally reduced. For these reasons the main focus of research in the area lies in the simplification of localization algorithms. Yet, WSN applications with small embedded hardware solutions allow to widen the ordinary localization techniques with more complex multi-node sound source detection and recognition solutions, e.g. [6]–[9].

In this paper we propose a method of Initial Search Region Reduction (ISRR) for the SRP-PHAT, that significantly reduces computational load. For the implementation we use several linear microphone arrays, that together constitute a large-aperture array with a wide area of observation. The ISRR is performed by estimating the DOA for every sub-array and finding the region of common direction. Alternatively we also use multilateration for the region estimation. For final localization we apply the optimized version of SRP-PHAT, which uses Stochastic Region Contraction (SRC) for global energy maximum search. The proposed method is tested on real signals for moving speaker localization and industrial machinery monitoring [10] applications. Based on the results, we consider the advantages and shortcomings of the DOA and multilateration approaches to ISRR.

II. Acoustic Localization with SRP-PHAT

Acoustic localization may be performed either in a three or two-dimensional space. For our grounded applications we focus on the horizontal plane, thus acoustic source coordinates

This research was supported by the Innovative Manufacturing Engineering Systems Competence Centre IMECC, by the European Union Regional Development Fund (project EU30006), co-financed by the Estonian ICT doctoral school and the Nations Support Program for ICT in Higher Education "Tiger University".

S. Astapov and J.-S. Preden are with the Laboratory for Proactive Technologies, Department of Computer Control, Tallinn University of Technology, Ehitajate tee 5, 19086, Tallinn, Estonia (e-mails: sergei.astapov@dcc.ttu.ee; jurgo.preden@dcc.ttu.ee).

J. Berdnikova is with the Department of Radio and Communication Engineering, Tallinn University of Technology, Ehitajate tee 5, 19086, Tallinn, Estonia (e-mail: juliad@lr.ttu.ee).

(x, y) are estimated. In the two-dimensional space the use of linear arrays is sufficient and computationally less complicated. Linear arrays consist of several microphones with equal distances between each other. The TDOA from one microphone to another then specifies the DOA of the source. The calculation of DOA relies firstly on the speed of sound (in air in our case), the dependence of which on the ambient temperature is expressed by the following equation:

$$c = 331.45\sqrt{1 + \theta/273}, \tag{1}$$

where c is the speed of sound and θ is the temperature in Celsius. Secondly, DOA depends on the assumption of near or far field signal source location. For our implementation we assume the far field disposition of the sound source. The near and far field assumptions specify the trigonometry to be used for DOA computation. Sound waves propagate spherically, and while in the near field this curvature of the wave front is accounted for in DOA calculation, in the far field the fronts are well spread and considered linear. We combine several linear array blocks to achieve a large-aperture array with a Field of View (FOV) of up to 25 m^2. A FOV is an area where the sound source is localizable, it directly depends on the array's configuration. Large FOV require much time and resources for the source to be localized.

A. *Conventional SRP-PHAT*

SRP-PHAT is a technique of estimating the DOA of sound signals in a reverberant environment. The SRP $P(\vec{a})$ is a real-valued functional of a spatial vector $\vec{a}$, which is defined by the FOV of a specific array. The high maxima in $P(\vec{a})$ indicate the estimates of the sound source location. $P(\vec{a})$ is computed for each direction as the cumulative Generalized Cross-Correlation with Phase Transform (GCC-PHAT) value across all pairs of microphones at the theoretical time-delays associated with the chosen direction. Consider a pair of signals $x_k(t)$, $x_l(t)$ of an array consisting of M microphones. The time instances of sound arrival from a point $a \in \vec{a}$ for the two microphones are $\tau(a, k)$ and $\tau(a, l)$ respectively. Hence the time delay between the signals is $\tau_{kl}(a) = \tau(a, k) - \tau(a, l)$. The SRP-PHAT for all pairs of signals is then defined as

$$P(a) = \sum_{k=1}^{M} \sum_{l=k+1}^{M} \int_{-\infty}^{\infty} \Psi_{kl} X_k(\omega) X_l^*(\omega) e^{j\omega(\tau(a,k) - \tau(a,l))} d\omega, \tag{2}$$

where $X(\omega)$ is the spectrum (the Fourier transform) of signal x and $X^*(\omega)$ is the conjugate of the spectrum [11]. Ψ_{kl} is the PHAT weight of the inverse of the spectral magnitudes:

$$\Psi_{kl} = \frac{1}{|X_k(\omega) X_l^*(\omega)|}. \tag{3}$$

The PHAT is an effective weighting of a GCC for finding TDOA from signals in a highly-reverberant environment.

Computing the SRP for every point in the area $\vec{a}$ results in a SRP image of the whole observable FOV. These images are highly suitable for manual analysis as they portray signal energy distributions and reverberation effects very clearly. For example, consider a result of speaker localization in a room performed by conventional SRP-PHAT, presented in Fig. 1. However, for automated processing (i.e. global maximum estimation) these images contain an overwhelming amount of information. Consequently, the processing, as the image generation itself, is highly time and resource consuming. Several propositions have been made for SRP optimization [11]–[13]. For our work we choose the method of locating high maxima of SRP energy by applying Stochastic Region Contraction (SRC).

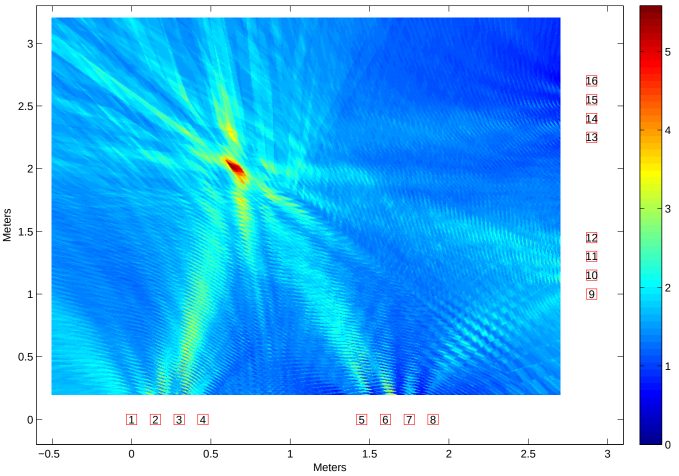

Fig. 1. SRP image of speaker localization using the conventional SRP-PHAT.

B. *SRP-PHAT with SRC*

The conventional SRP-PHAT performs as many functional evaluations (2), or FE, as there are points in $\vec{a}$, the number of which is defined by the dimensionality of the FOV and the accuracy measure, that partitions the area into small discrete regions. This analysis is highly resource demanding, particularly when applied for large areas of observation. The number of computations is significantly reduced by applying Stochastic Region Contraction, which iteratively reduces the search volume for the global maximum. SRC starts with the initial search volume (i.e. the whole FOV), stochastically explores the functional of that volume by randomly picking a specific number of points, and then contracts the volume into the sub-volume containing the desired global optimum and proceeds iteratively until the global maximum can be located with a finite precision [11]. The procedure may be described in pseudo code as:

1) Initialize iteration $i = 0$.
2) Set initial parameters: $V_0 = V_{\text{room}}$ – initial volume; J_0 – the number of random points that need to be evaluated to ensure, that one or more is likely to reside in the sub-volume of higher values, surrounding the global maximum; N_0 – number of points used to define the new sub-volume V_{i+1}.
3) Calculate $P(\vec{a})$ for J_i points.
4) Sort out the best (highest) $N_i \ll J_i$ points.
5) Contract the search volume to the smaller volume V_{i+1}, defined by a rectangular boundary vector $B_{i+1} = [x_{\min}(i+1), x_{\max}(i+1), y_{\min}(i+1), y_{\max}(i+1), z_{\min}(i+1), z_{\max}(i+1)]$, that contains these N_i points.

6) **IF** $V_{i+1} < V_u$ (a sufficiently small sub-volume, in which the global optimum is contained) **AND** $FE_i < \Phi$ (the total number of FE-s for iteration i is less than the maximum number of allowed FE-s), **THEN** determine the global maximum, **STOP**.
7) **ELSE IF** $FE_i \geq \Phi$, **STOP**, discard results.
8) **ELSE** Among the N_i points keep a subset G_i of points, which have values greater than the mean μ_i of the N_i points.
9) Evaluate J_{i+1} new random points in V_{i+1}.
10) Form the set of N_{i+1} as the union of G_i and the best $N_{i+1} - G_i$ points from the J_{i+1} just evaluated. This gives N_{i+1} high points for iteration $i+1$.
11) $i = i+1$, **GO TO** Step 5.

There are several proposed ways of selecting N_i and J_i depending on the specific FOV and on the condition of monotonic or non-monotonic increase of the mean μ_i. The one, emphasized in [11], consists of fixing N_i and adjusting J_i incrementally in the following fashion: N_i is chosen as $N_i \equiv N = 100$; J_i is the number of FE-s to find $N - G_i$ points greater than μ_i. For our system we propose a different method, which is presented in Section III-C.

III. Initial Search Region Reduction

However, optimized localization algorithms still require a significant amount of resources while starting the evaluation on the initial search area. Furthermore, the convergence on a sharp maximum may be guaranteed only if it exists in the FOV. For many applications and monitored objects this is not always true. Large objects, like vehicles or other machinery, do not have a single point of sound emission, rather they appear as distributed regions of heightened acoustic energy with several maxima. On the other hand, if no sound source is present, the localization algorithms will search for maxima in ambient noise, which produces useless results while consuming resources. The reduction of the initial search area, firstly, allows to estimate the presence of the sound source in the FOV, and secondly, greatly reduces the computational load of localization.

We focus on an array setup targeted for use in WSN. The sub-array blocks are places in different positions in the environment, their orientation may be at all random. The position of the sub-array is specified by the coordinates of its center (which may be found using [14] or [15]) and the angle α, by which the array is steered from the global zero angle, as it is shown in Fig. 2. Knowing the coordinates of a block center (x_0, y_0), ith sensor before rotation (x_i, y_i) and the angle α, the steering is performed as

$$\begin{bmatrix} x_i^{(rot)} \\ y_i^{(rot)} \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \begin{bmatrix} x_i - x_0 \\ y_i - y_0 \end{bmatrix}. \quad (4)$$

Such a configuration is convenient for WSN, where each sub-block may be implemented on a separate network node. Sub-arrays with common FOV form large-aperture arrays and cooperate on localization. Such a configuration enables ad-hoc array composition and increases robustness due to high decentralization. Also a large number of sub-blocks simplifies multiple source localization, as the monitored area is divided into smaller local regions.

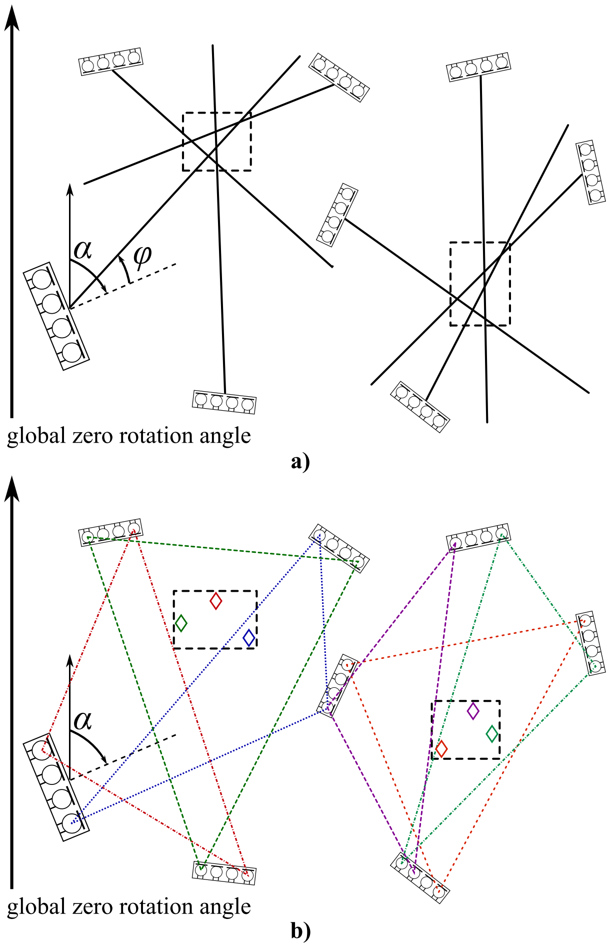

Fig. 2. Initial search region estimation by (a) DOA calculation and (b) multilateration in random configuration of sensor array blocks.

The ISRR is performed by estimating the DOA for every sub-array and finding the region of common DOA (i.e. the intersection of vectors pointed by the DOA) as is shown in Fig. 2a. We also consider an alternative approach of choosing sensor triplets and performing multilateration to retrieve the source coordinate estimates. The aggregate of these estimates then denotes the sought-for region (Fig. 2b).

A. The DOA Approach to ISRR

Having K microphone arrays, each consisting of M microphones, observing a common FOV, the ISRR is performed in the following steps:

1) Estimate the DOA for each of K arrays.
2) Generate vectors spanning from the arrays' centers to the bounds of the FOV in the directions of DOA.
3) Find points of intersections of these vectors.
4) Find groups of points no farther than $D_{\max}$ distance units (meters) from their centroid and enclose the areas, in which these groups coincide, in rectangles.
5) Perform control of false detection, discard areas not meeting specific criteria (optional).

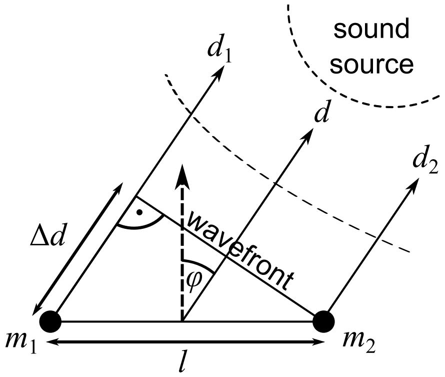

Fig. 3. DOA estimation for a pair of microphones.

The DOA are estimated for the array front, i.e. from $-90°$ to $90°$. As we operate in the horizontal plane, it is sufficient to acquire the azimuth (angle of arrival) of the incoming signal to estimate the DOA [16]. The estimation is performed for all $\binom{M}{2}$ combinations of M microphone pairs. Considering Fig. 3, the sound wave originating from a source in the far field is acquired by the microphones m_1 and m_2 with a time delay $\tau \in [-\tau_{\max}, \tau_{\max}]$, where $\tau_{\max}$ is the delay of sound traveling directly from one microphone to the other (i.e. at $\pm 90°$). To estimate τ we apply cross-correlation to the two signals:

$$R(\tau) = \sum_{k=0}^{n} m_1(k) \cdot m_2(k-\tau), \tag{5}$$

where n is the length of the signals in samples. The maximum of the cross-correlation defines the time delay, and the azimuth is obtained by

$$\varphi = \arcsin \frac{\tau \cdot c}{l} = \arcsin \frac{\Delta k / f_s \cdot c}{l}, \tag{6}$$

where c is the speed of sound, l is the distance between two microphones and τ is represented in terms of delay in samples Δk and the sampling frequency f_s. Depending on the chosen pair of microphones in the array, l will vary from l to $l(M-1)$. At this point data validation may be performed. If the maximum of correlation is less than some threshold, the azimuth φ may be discarded. This way, in absence of a sound source or in case of high noise, invalid estimates are avoided early on. We use the deviation from the mean for this metric:

$$\max\left(R(\tau)\right) > (1 + TH) \cdot \overline{R(\tau)}, \tag{7}$$

where TH is the threshold of deviation, which depends on the Signal to Noise Ratio (SNR). We use 0.2-0.3 in our experiments.

Having $C_i \leq \binom{M}{2}$ azimuth estimates for every array (varying slightly due to varying inter-microphone distance and accounting for the far field error), the final DOA for each ith array, $i \in (1, \ldots, K)$, is derived according to the following special cases:

1) DOA spread uniformly (leftmost pairs point left, rightmost – right, and center – straight): no common DOA, $\phi_i = \emptyset$.
2) DOA are consensual with slight variance: common DOA is the mean of pair-wise ones

$$\phi_i = \frac{1}{C_i} \sum_{j=1}^{C_i} \varphi_{i,j}. \tag{8}$$

3) Same as *Case* 2), but with some DOA outside variance of consensual group: exclude these DOA from mean.
4) Several distinct groups of consensual DOA: choose one with more members and less variance (several may be considered for heavily multi-source applications), calculate mean.

Having estimated $K_1 \leq K$ of the existing array DOA ϕ_{i^*}, $i^* \in (1, \ldots, K_1)$ and added the nodes' rotation angles α_i to them, vectors $\overrightarrow{AB}_{i^*}$ are computed with the starting point $A_{i^*} = (x_{1,i^*}, y_{1,i^*})$ being the coordinate of i^*th array's center and the ending point $B_{i^*} = (x_{2,i^*}, y_{2,i^*})$ being the point at a bound of the FOV steered by ϕ_{i^*} from the array's center. Intersection points of all pairs $\overrightarrow{AB}_h$, $\overrightarrow{AB}_k$ are calculated by

$$\begin{aligned}&\mathrm{I}_{hk} = (I_x, I_y) = \\ &\left(\frac{(x_{1,h}y_{2,h} - y_{1,h}x_{2,h})(x_{1,k} - x_{2,k}) - (x_{1,h} - x_{2,h})(x_{1,k}y_{2,k} - y_{1,k}x_{2,k})}{(x_{1,h} - x_{2,h})(y_{1,k} - y_{2,k}) - (y_{1,h} - y_{2,h})(x_{1,k} - x_{2,k})}, \right. \\ &\left. \frac{(x_{1,h}y_{2,h} - y_{1,h}x_{2,h})(y_{1,k} - y_{2,k}) - (y_{1,h} - y_{2,h})(x_{1,k}y_{2,k} - y_{1,k}x_{2,k})}{(x_{1,h} - x_{2,h})(y_{1,k} - y_{2,k}) - (y_{1,h} - y_{2,h})(x_{1,k} - x_{2,k})} \right).\end{aligned} \tag{9}$$

As a result we have a set of $\mathrm{I}_{i^{**}}$ intersections, $i^{**} \in (1, \ldots, K_2)$, $K_2 \leq \binom{M}{2}$. To get the initial search areas, these intersection points are partitioned by their relative distance. For the maximum distance $D_{\max}$ the partitioning is performed in the following manner:

1) **IF** no points $\mathrm{I} = \emptyset$ **THEN** no partitions $\mathrm{P} = \emptyset$, **STOP**
2) **ELSE IF** only 1 point I_1 **THEN** $\mathrm{P}_1 = \mathrm{I}_1$ **STOP**
3) **ELSE** number of partitions $j = 0$
4) **WHILE** $|\mathrm{I}| > 0$, where $|\mathrm{I}|$ is the cardinality of the set I, calculate centroid of free points $\mathrm{Cent} = 1/|\mathrm{I}| \cdot \sum \mathrm{I}$.
5) Calculate Euclidean distance of all free points to centroid $D_k = \sqrt{\sum_{s=1,2} (\mathrm{I}_{k,s} - \mathrm{Cent}_s)^2}$, choose point with minimal distance, $j = j + 1$, insert point to P_j, remove point from set of free points I.
6) Calculate partition centroid $\mathrm{Cent}(P_j) = {}^{1}/_{|P_j|} \cdot \sum \mathrm{P}_j$, get Euclidean distance for all free points $D_k = \sqrt{\sum_{s=1,2} (\mathrm{I}_{k,s} - \mathrm{Cent}(\mathrm{P}_j)_s)^2}$.
7) **IF** $\min(D) \leq D_{\max}$ **THEN** insert point corresponding to $\min(D)$ into P_j, delete point from set of free points I, **GO TO** Step 6.
8) **ELSE IF** $|\mathrm{I}| > 1$ **THEN GO TO** Step 4.
9) **ELSE** $j = j + 1$, put last remaining point to P_j.

After obtaining the partitions, their areas are enclosed by rectangles with the edges denoted by the partitions' minimal and maximal values of x and y, added a constant in order to ensure minimal area (in the experimental part we choose 0.1 m). As a result several initial regions may occur in the same FOV. Also while a vector of DOA from one array may cross with several other vectors, redundant "echoing" regions may arise. These may be removed by additional control metrics or by analyzing previously estimated positions (i.e. tracking).

B. Multilateration Approach to ISRR

Multilateration is a technique of estimating the signal source coordinates based on TDOA from the source to the receiving sensors. The distance between the sensor with coordinates x_i, y_i, z_i and the acoustic object could be defined as the length of vector $\vec{d}$:

$$\left\|\vec{d}\right\| = \sqrt{\left(x_i - x\right)^2 + \left(y_i - y\right)^2 + \left(z_i - z\right)^2}, \qquad (10)$$

where x, y, z are the acoustic source coordinates. For the multi-sensor WSN ground applications we simplify the solution with constant z dimension. Thus having a TDOA τ_{ij} between two nodes i and j, the acoustic source location coordinates are calculated directly by

$$\begin{gathered} d_{ij} = c \cdot \tau_{ij} = c\left(\tau_i - \tau_j\right) = \\ \sqrt{\left(x_i - x\right)^2 + \left(y_i - y\right)^2} - \sqrt{\left(x_j - x\right)^2 + \left(y_j - y\right)^2}, \end{gathered} \qquad (11)$$

where d_{ij} is the distance difference estimate between sensors i and j, and $(x_i, y_i)$and (x_j, y_j) are the sensors' respective coordinates [16]. If τ_{ij} is represented in terms of delay in samples Δk_{ij} with sampling frequency f_s, then the difference, similar to (6), is computed as $d_{ij} = \Delta k_{ij}/f_s \cdot c$. The delay τ_{ij} is calculated using cross-correlation, as in (5), also applying the control metric (7). For any three separate sensors $(1, 2, 3)$ acoustic source is localizable by the following system of equations:

$$\begin{cases} d_{12} = \sqrt{\left(x_1 - x\right)^2 + \left(y_1 - y\right)^2} - \sqrt{\left(x_2 - x\right)^2 + \left(y_2 - y\right)^2} \\ d_{13} = \sqrt{\left(x_1 - x\right)^2 + \left(y_1 - y\right)^2} - \sqrt{\left(x_3 - x\right)^2 + \left(y_3 - y\right)^2} \end{cases} \qquad (12)$$

To estimate the solution to this system of nonlinear equations we apply a numerical method called Trust-Region Dogleg [17].

We use multiple sensor triplets in order to establish several triangles for multilateration. Every triplet gives a separate position estimate and then all the estimates are partitioned by minimal distance, as in Section III-A, in order to get the reduced regions. The general direct multilateration solution in real-time WSN applications is solved with larger number of nodes [18], where the incorrectly placed regions or multiple sound sources are eliminated by feedback from the object tracking stage. We, however, do not expand beyond three sensor batches in order to simplify and accelerate the solution estimation procedure.

C. Application of SRP-PHAT with SRC to ISRR Estimates

Our approach initializes the SRP-PHAT on already contracted areas and often more than once for a single signal frame. The typical approach to SRC suggests choosing fixed values for $N_i \equiv N = 100$ and $J_0 = 3000$ for a FOV of approximately 20 m^2, however this is not suitable for constantly varying initial search areas. In our approach the parameters are rather estimated by linear functions. Building on the test results in [11], considering peak estimation quality, and performing our own testing, we derive the two functions for the task:

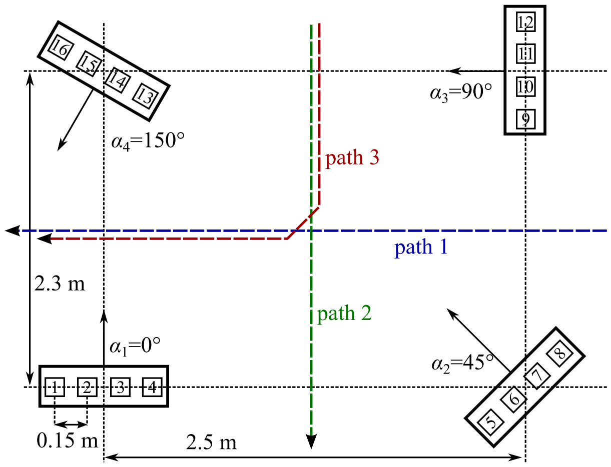

Fig. 4. Layout of the experiment with one speaker and four array blocks.

$$\begin{aligned} J_0(s) &= \begin{cases} [297.6 \cdot s + 24], & S < 10 \\ 3000, & S \geq 10 \end{cases}, \\ N(s) &= \begin{cases} [9.9 \cdot s + 1], & S < 10 \\ 100, & S \geq 10 \end{cases}, \end{aligned} \qquad (13)$$

where s is the area of the FOV in m^2 and $[\cdot]$ denotes the operation of rounding to the nearest integer. The application of these functions optimizes the SRC process by greatly reducing the number of SRP evaluations for reduced regions of acoustic source search [19].

IV. Experimental Results

For the experimental installation we use Vansonic PVM-6052 condenser microphones. The microphones are mounted with a spacing of 15 cm between each other. We use 4 sub-arrays with 4 microphones in each sub-array (width of a single sub-array is thus equal to 45 cm), which results in a large aperture 16-microphone array. For signal acquisition an Agilent U2354A data acquisition device (DAQ) is used with the sampling rate set to 8 kS/s per channel. The data is acquired to and processed in the Matlab environment using the Data Acquisition Toolbox. Processing is performed in frames with a step of 0.2 seconds by conventional SRP-PHAT and by ISRR followed by SRP-PHAT SRC on estimated regions.

A. Human Speaker Localization

For the human speaker localization experiment the microphones are placed in a room as it is portrayed in Fig. 4. The FOV is set to be 1 meter wider in every direction than the corner points of the array (approximately 18 m^2). Sub-arrays SA1, SA2 and SA3 form a triangular array, while sub-array SA4 is diverted from the common direction of view, simulating the belonging to a different group. The speaker takes 3 paths while walking with an average pace (approximately 1-1.5 m/s) and reciting the rainbow passage (designed to contain all the English phonemes and used in speech evaluations). For the DOA approach to ISRR all 16 microphones are used, for

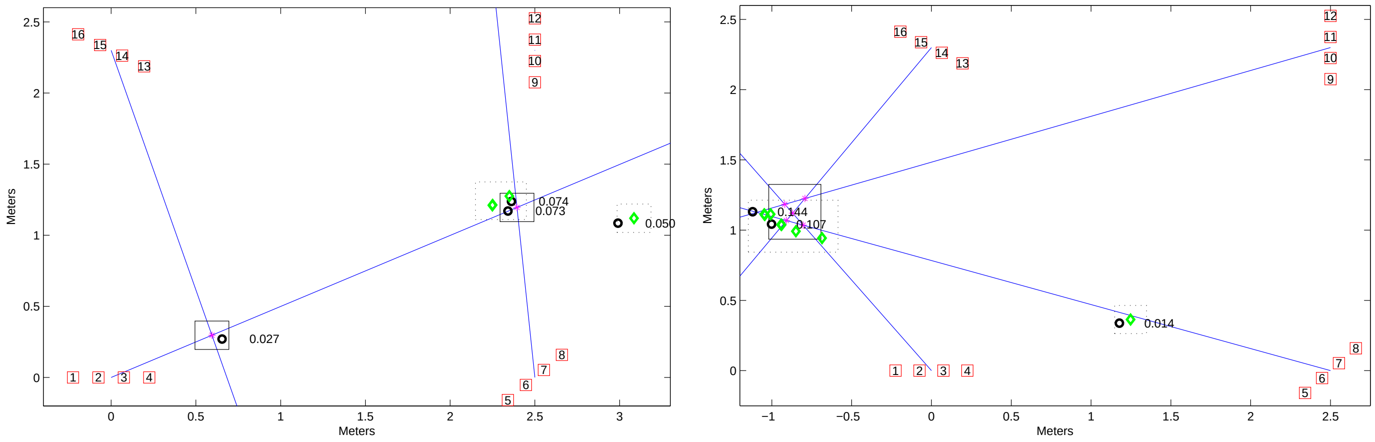

Fig. 5. Two instances of ISRR and localization for the speaker experiment. Blue lines denote sub-array DOA, pink stars – the intersections of DOA rays, black boxes – the estimated regions and black circles - the energy maximum of a region followed to the right by its value. For triangulation the coordinate estimates are denoted by green diamonds, the regions – by dotted boxes and energy maxima – by black circles.

TABLE I
RESULTS OF REGION REDUCTION FOR SPEAKER LOCALIZATION

FOV≃18 m^2	DOA estimation	Multilateration
Mean area (m^2)	0.1374	0.0621
RMSE x (m)	0.1143	0.1227
RMSE y (m)	0.1107	0.1230

multilateration the triplets are chosen in the following order: 1 4 12; 1 4 16; 5 8 12; 5 8 16; 1 8 12; 1 8 16; 4 5 9; 4 5 13. Several resulting triangles with small areas may perform better on closer distances to the source and those with larger areas – on greater distances.

In this experiment the ISRR with DOA estimation and multilateration operate with approximately equal accuracy. Problems arise for both approaches in the region behind and between SA2 and SA3 (path 1), where neither SA2 or SA3 have a sufficient view of the source and SA1 and SA4 are overly steered away (Fig. 5 left). For SA4 the DOA totally exceeds its limits. A slight advantage of multilateration is, however, evident due to its non-directional approach. The latter part of path 1 and both paths 2 and 3 are well traceable by both approaches. In the leftmost region of the FOV, where SA4 is also active, the ISRR achieves the best results (Fig. 5 right).

The impact of the ISRR is substantial, the mean area is reduced from 18 m^2 of the whole FOV range to a fraction of a square meter (see Tab. I). To estimate the divergence from the global maximum estimated over the FOV, we find the difference between the result of conventional SRP-PHAT and the result of our method. Error variation over time is presented in Fig. 6, and the Root Mean Square Errors (RMSE) are presented in Tab. I. As it can be seen from the x-axis values in Fig. 6, the DOA approach discards less frames than multilateration (due to non-detection operation) and is therefore more sensitive to the sound source. Also the RMSE is slightly lower for the DOA approach. The overall errors are sufficiently low for speaker localization. Rare bursts of error do occur, however they are instantaneous and appear only during moments of speaker acceleration.

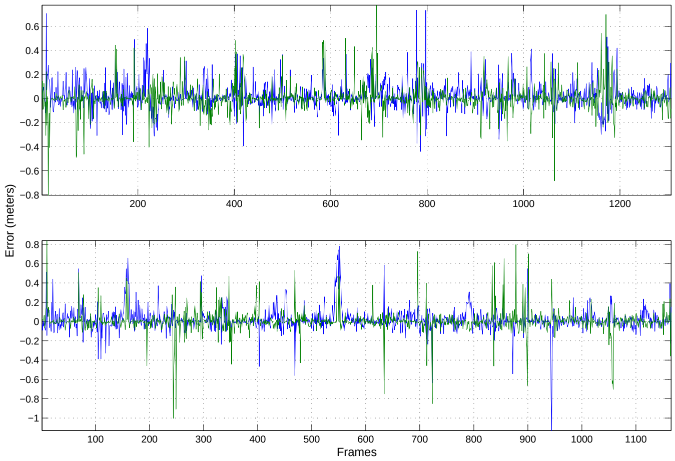

Fig. 6. Difference in localization between SRP-PHAT over the whole FOV and using ISRR: DOA approach (upper) and multilateration (lower). Blue line denotes the x and the green line – the y coordinates.

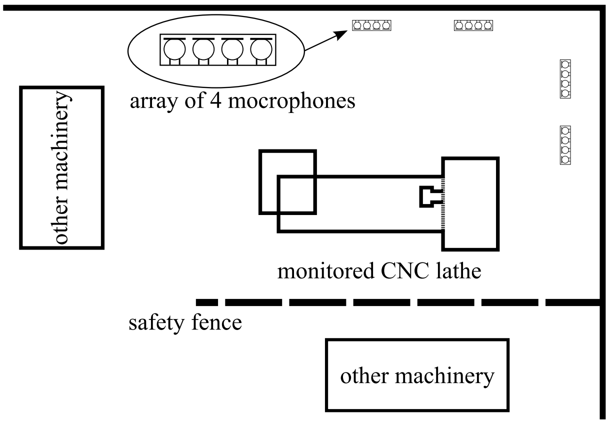

Fig. 7. Array placement at an industrial facility for CNC lathe monitoring.

B. Industrial Machinery Monitoring

For the industrial machinery monitoring experiment the same hardware implementation as for speaker localization is used. The microphones and their triplets are chosen in the

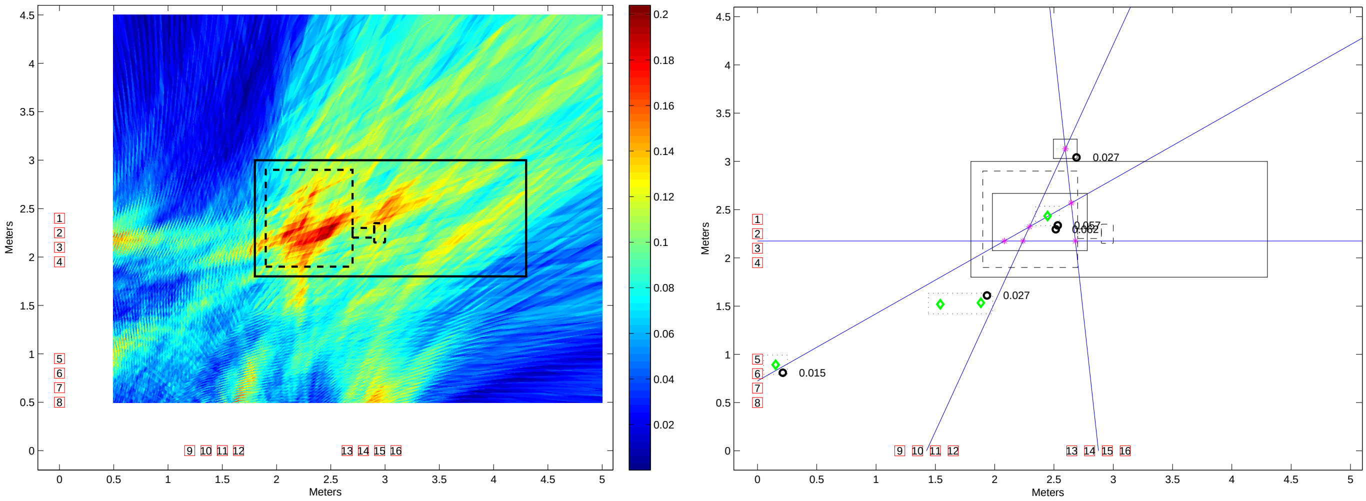

Fig. 8. CNC lathe noise localization with conventional SRP-PHAT (left) and using ISRR followed by SRP-PHAT with SRC (right).

same manner. The placement scheme and the room layout is presented in Fig. 7. The two sub-arrays are placed near a wall at a right angle to the other two sub-arrays, i.e. $\alpha = \{0, 0, 90, 90\}$. The monitored object in the experiment is a large Computer Numerical Control (CNC) lathe. The main noise sources of the lathe are the motor with the gear box and the spindle. The lathe is put through a short working cycle: the motor is activated, the spindle rotates and the carriage with the cutting tool moves beside the cutting area, after which the spindle and then the motor are shut down.

In this experiment the DOA approach to ISRR performs significantly better than multilateration. It seems that multilateration, considering this specific configuration of the array and the large sound emitting area of the lathe, cannot estimate the region confined by the specified maximal distance $D_{\max} = 0.5$ (see Section III-A). A frame corresponding to the active motor and spindle noise is presented in Fig. 8. The region of the motor noise is correctly located by the DOA approach and corresponds to the result of the conventional SRP-PHAT. Multilateration estimates only one small correct region. The triangular array configuration used is evidently not appropriate for localization of large sound sources by multilateration.

V. Discussion and Future Work

The ISRR method has proven to perform well for both experimental tasks. The DOA and multilateration approaches perform differently in various situations. The non-directional nature of multilateration enables it to locate the sources out of view for DOA. On the other hand, the directional approach eliminates the possibility of regions duplicating on the opposite side of the array due to reverberation. For a more complicated task of localizing a large noise region, the supremacy of the DOA approach is more evident. The DOA method performs better in a triangular configuration and worse in a square-like configuration. The situation with multilateration is absolutely opposite. Thus, ISRR type may be chosen based on the configuration of the array and the specific application. Both approaches may be used in conjunction for mutual reassurance.

For future work we intend to develop a fully embedded system with array blocks implemented on individual devices. The dataflows between the devices must be thoroughly researched in order to achieve smooth real-time operation. As SRP-PHAT demands information from different sub-arrays, a cooperation scheme with resource allocation must be developed. The operation may proceed in an ad-hoc manner, where the operations are equally distributed between nodes, or a separate node may be allocated for sophisticated computations.

A. *Aspects of WSN Organization*

The underlying computation and communication system realizing the localization method described in the paper must be able to cope with the real-time requirements set by the employed algorithms. The data delivered to the fusion algorithm from the individual microphone arrays must be temporally valid, i.e. the age of the data delivered to the fusion algorithm must not exceed a set maximum and the data from distinct sources processed by an algorithm must be coherent in time and space. In the current application the spatial aspect is of special importance as due to uncertainties inherent in a distributed architecture – the locations of the individual arrays are not known beforehand and the configuration of the system may change over time. This means that the spatial aspects must be explicitly considered in communication and computation. As the communication and processing delays are dynamic, the system must be also able to cope with these variances. In order to manage with these uncertainties we suggest the use of a proactive middleware [20] as an active mediator of data to and from the individual computing nodes. The middleware enables deterministic data exchange between autonomous (sensing) systems according to the constraints set by individual fusion algorithms and devices. This is achieved by performing constraint propagation from the computing nodes and online data validation based on the propagated constraints. In addition to the data propagation tasks, the middleware can be also used as a tool to synchronize the spatial properties of devices, such as location and orientation, making it for example possible to perform data alignment for the angles and coordinates among the WSN nodes at the level of the middleware. The proactive middleware can also distribute the individual tasks (e.g. to compute the SRP per FOV) among the WSN nodes using a prescribed scenario (partiture) as proposed in [21].

VI. Conclusion

The paper proposes a method of search region reduction for the purpose of optimizing acoustic localization. The method targets the sensor array configurations implementable on separate nodes of WSN. Two approaches to the method are proposed and tested on two real experimental signals. The results are positive with the method succeeding with substantially reducing the search region and localizing with small amounts of error. The differences in localization quality for the two approaches under different circumstances do not show definite supremacy of either approach. The results suggest the application of both approaches to region reduction in the final implementation.

References

[1] S. Astapov and A. Riid, "A multistage procedure of mobile vehicle acoustic identification for single-sensor embedded device," *International Journal of Electronics and Telecommunications*, vol. 59, no. 2, pp. 151–160, 2013.

[2] D. Blatt and A. O. Hero, "APOCS: a rapidly convergent source localization algorithm for sensor networks," in *Proc. IEEE/SP 13th Workshop Statistical Signal Processing*, 2005, pp. 1214–1219.

[3] F. Masson, D. Puschini, P. Julian, P. Crocce, L. Arlenghi, P. Mandolesi, and A. G. Andreou, "Hybrid sensor network and fusion algorithm for sound source localization," in *Proc. IEEE Int. Symp. Circuits and Systems ISCAS 2005*, 2005, pp. 2763–2766.

[4] C. T. Ishi, O. Chatot, H. Ishiguro, and N. Hagita, "Evaluation of a MUSIC-based real-time sound localization of multiple sound sources in real noisy environments," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems IROS 2009*, 2009, pp. 2027–2032.

[5] H. Hongsen, L. Jing, and G. Yang, "Acoustic direction of arrival estimation based on spatial circular prediction," in *Information Technology and Applications, 2009. IFITA '09. International Forum on*, vol. 3, 2009, pp. 177–180.

[6] A. Dhawan, R. Balasubramanian, and V. Vokkarane, "A framework for real-time monitoring of acoustic events using a wireless sensor network," in *Proc. IEEE Int Technologies for Homeland Security (HST) Conf*, 2011, pp. 254–261.

[7] T. Liu, Y. Liu, X. Cui, G. Xu, and D. Qian, "MOLTS: Mobile object localization and tracking system based on wireless sensor networks," in *Proc. IEEE 7th Int Networking, Architecture and Storage (NAS) Conf*, 2012, pp. 245–251.

[8] Z. Merhi, M. Elgamel, and M. Bayoumi, "A lightweight collaborative fault tolerant target localization system for wireless sensor networks," *IEEE Trans. Mobile Comput.*, vol. 8, no. 12, pp. 1690–1704, 2009.

[9] G. Vakulya and G. Simon, "Fast adaptive acoustic localization for sensor networks," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 5, pp. 1820–1829, 2011.

[10] S. Astapov, J. S. Preden, T. Aruvali, and B. Gordon, "Production machinery utilization monitoring based on acoustic and vibration signal analysis," in *Proc. 8th Int DAAAM Baltic Industrial Engineering Conf*, 2012, pp. 268–273.

[11] H. Do, H. F. Silverman, and Y. Yu, "A real-time SRP-PHAT source location implementation using stochastic region contraction (SRC) on a large-aperture microphone array," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing ICASSP 2007*, vol. 1, 2007.

[12] M. Cobos, A. Marti, and J. J. Lopez, "A modified SRP-PHAT functional for robust real-time sound source localization with scalable spatial sampling," *IEEE Signal Process. Lett.*, vol. 18, no. 1, pp. 71–74, 2011.

[13] X. Zhao, J. Tang, L. Zhou, and Z. Wu, "Accelerated steered response power method for sound source localization via clustering search," *Science China Physics, Mechanics and Astronomy*, vol. 56, no. 7, pp. 1329–1338, 2013.

[14] E. Mangas and A. Bilas, "FLASH: Fine-grained localization in wireless sensor networks using acoustic sound transmissions and high precision clock synchronization," in *Proc. 29th IEEE Int. Conf. Distributed Computing Systems ICDCS '09*, 2009, pp. 289–298.

[15] Q. Wang, R. Zheng, A. Tirumala, X. Liu, and L. Sha, "Lightning: A hard real-time, fast, and lightweight low-end wireless sensor election protocol for acoustic event localization," *IEEE Trans. Mobile Comput.*, vol. 7, no. 5, pp. 570–584, 2008.

[16] Y. Liu and Z. Yang, *Location, Localization, and Localizability: Location-awareness Technology for Wireless Networks*. Springer, 2010.

[17] J. Nocedal and S. J. Wright, *Numerical optimization*. Springer verlag, 1999.

[18] S. Zhang, G. Li, W. Wei, and B. Yang, "A novel iterative multilateral localization algorithm for wireless sensor networks," *Journal of Networks*, vol. 5, no. 1, pp. 112–119, 2010.

[19] S. Astapov, J.-S. Preden, and J. Berdnikova, "Simplified acoustic localization by linear arrays for wireless sensor networks," in *Digital Signal Processing (DSP), 2013 18th International Conference on*, 2013, pp. 1–6.

[20] L. Motus, M. Meriste, and J. Preden, "Towards middleware based situation awareness," in *Military Communications Conference, 2009. MILCOM 2009. IEEE*. IEEE, 2009, pp. 1–7.

[21] J. Preden and J. Helander, "Situation aware computing in distributed computing systems," in *Proceedings of the 10th Symposium on Programming Languages and Software Tools*, 2007, pp. 280–292.