Performing Acoustic Localization in a Network of Embedded Smart Sensors

Sergei Astapov, Johannes Ehala, and Jürgo-Sören Preden

***Abstract*—Situation awareness is an important aspect of ubiquitous computer systems, as these systems of systems are highly integrated with the physical world and for successful operation they must maintain high awareness of the environment. Acoustic information is one of the most popular modalities, by which the environment states are estimated. Multi-sensor approaches also provide the possibility for acoustic source localization. This paper considers an acoustic localization system of dual channel smart sensors interconnected through a Wireless Sensor Network (WSN). The low computational power of smart sensor devices requires distribution of localization tasks among WSN nodes. The Initial Search Region Reduction (ISRR) method is used in the WSN to meet this requirement. ISRR, as opposed to conventional localization methods, performs significantly less complex computations and does not require exchange of raw signal between nodes. The system is implemented on smart dust motes utilizing Atmel ATmega128RFA1 processors with integrated 2.4GHz IEEE 802.15.4 compliant radio transceivers. The paper discusses complications introduced by low power hardware and ad-hoc networking, and also reviews conditions of real-time operation.**

***Index Terms*—Acoustic localization, Wireless Sensor Networks, Direction of Arrival, smart dust, distributed computing.**

I. Introduction

THE continuous process of computer systems integration into all aspects of everyday life paves the way for cyber-physical systems with diverse abilities for interfacing with human operators and the environment, in which these systems exist. Future Internet of Things applications are also envisioned to be ubiquitous systems, which must maintain good situation awareness in order to be able to provide the expected services proactively. Situation awareness is achieved by constant analysis of environment states by sensing different modalities (e.g. acoustic, video, vibration, magnetic, etc.) and sophisticated decision-making through data fusion and system component cooperation. One of the most popular modalities for the majority of environments and human-machine interaction is acoustic signals. Acoustic information is widespread and may be acquired during various physical processes accompanied by sound emission and during human speech analysis.

Acoustic signal analysis has been applied for a great variety of tasks concerning both environment monitoring and human-machine interfaces (HMI). Applications for open environments span from traffic monitoring [1] and military reconnaissance [2] to monitoring woodland and aquatic wildlife [3]. For confined environments the main applications are person and process monitoring, raging from home automation [4] and security systems [5] to industrial process control [6]. The majority of monitoring tasks assume pattern matching and classification and use single-sensor solutions. Single channel information is sufficient for low-noise environments with a well defined list of expected events and signal patterns (e.g. HMI with a finite list of voice commands). For the majority of open environments, however, multi-sensor systems are beneficial from several standpoints. Firstly, the monitored area is observed from multiple points of view, which provides more information than a single-sensor system. Secondly, if joint analysis is applied, position of the acoustic source may be localized in observed space. Thirdly, if localization is possible, the sound emitted by a specific source may be filtered from sounds incoming from other directions via beamforming.

The advantages of multi-sensor acoustic systems are even more evident if implemented in Wireless Sensor Networks (WSN). Compact sensor network solutions allow to widen the ordinary localization techniques with more complex multi-node source detection and recognition solutions, e.g. [7]–[10]. A WSN consists of several smart sensor nodes distributed in the observed environment and communicating with each other. The aggregate of local measurements from single nodes can be used to generate a global assessment of the situation. The downside of WSN application is low computational power of sensor nodes. In order to ensure the small size of smart sensors and the longevity of their power supplies, the hardware used in even modern smart sensors is quite limited in terms of computational power. However, these limitations may be overcome to a certain extent through node cooperation and distributed computations.

In this paper we consider a localization approach of Initial Search Region Reduction (ISRR), previously developed by our research group [11], [12], and its implementation in WSN. In our work we use smart sensors of particularly small size and low computational power, known as smart dust motes. Low power imposes many restrictions on signal acquisition and processing, e.g. low sampling rates, limited memory. The paper addresses these restrictions and proposes a system architecture for workload distribution, as well as discusses inter-node communication problems and system real-time operation capabilities. The decrease in localization quality resulting from the sampling rate decrease is demonstrated on a practical example of single speaker localization performed on an eight sensor system.

This research was partly supported by the Estonian Ministry of Defence, the European Defense Agency project IN4STARS, cofinanced by the Estonian ICT doctoral school and the Estonian IT Foundation for Education: IT Academy program.

All authors are with the Research Laboratory for Proactive Technologies, Department of Computer Control, Tallinn University of Technology, Ehitajate tee 5, 19086, Tallinn, Estonia
e-mail: sergei.astapov@ttu.ee, johannes.ehala@ttu.ee, jurgo.preden@ttu.ee

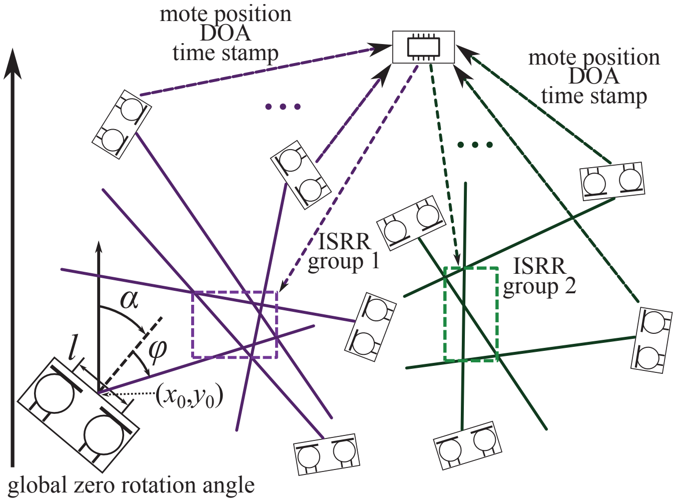


Fig. 1. Schematic diagram of proposed WSN architecture.

II. Distributed Localization in WSN

The WSN system is designed for localizing grounded acoustic sources. The motes are placed in the monitored environment in the horizontal plane and localization is performed by estimating the coordinates (x, y) of sound emitting objects. Each mote is equipped with two acoustic sensors spaced by a specific distance l from one another. Localization is based on estimating the time delays of acoustic wave arrival to the sensors, also called Time Difference of Arrival (TDOA). The Direction of Arrival (DOA) of sound from a specific acoustic source is calculated using TDOA. The whole localization system consists of a large number of motes, each one of which plays its specific role in the localization process. This section presents the proposed WSN architecture along with the distributed approach to localization.

A. WSN Architecture

The proposed WSN architecture is designed for applications in both open (urban, woodland, etc.) and confined (home, office, industrial facility, etc.) environments. The network consists of two types of motes: smart sensors and fusion nodes. Dual channel smart sensors acquire acoustic information and perform DOA estimation. Fusion nodes gather DOA estimates and perform further steps of localization, which are discussed in Section II-C. The schematic diagram of the architecture is presented in Fig. 1.

The sensor motes are dispersed in the monitored environment either in an orderly or random fashion. In confined environments an orderly placement is more likely, because sensors are usually mounted on room walls or ceilings. In open environments, however, its is rarely the case — the sensors may be attached to buildings, light posts, trash bins, etc. in urban and to trees, rocks, etc. in natural environments. Thus a general case is assumed, where the sensor's location is defined by the coordinates of its point of reference (x_r, y_r) and the angle α, by which the sensor is steered from the global angle reference. For example, sensor location may be estimated via the Global Positioning System (GPS), in which case the point of reference is the GPS unit. For environments, where GPS signals are unavailable, other location algorithms based on Radio Frequency (RF) [13] or sound [14] may be adopted. The global angle reference may be defined by Earth's magnetic field and the angle α estimated using a magnetometer. The central point of the microphone pair (x_0, y_0), for which the DOA is actually estimated, is defined by the reference point (x_r, y_r) and may coincide with it. The coordinates of the i-th microphone (x_i, y_i) are shifted by $\pm l/2$ from (x_0, y_0) and then the steering by α is performed as

$$\begin{pmatrix} x_i^{(rot)} \\ y_i^{(rot)} \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix} \begin{pmatrix} x_i - x_0 \\ y_i - y_0 \end{pmatrix}. \quad (1)$$

Sensor motes are partitioned into groups, where a single mote can belong to any number of groups. Each group must have a common Field of View (FOV), i.e. all motes observe the same area. The whole network may consist of several groups or each group can constitute a separate sub-network. Group partitioning is in essence a clustering task, for which two aspects are taken into consideration. Firstly, motes must have a common field of view as the considered localization procedure uses a directional approach. In this regard, the observed area is not necessarily enclosed by motes, as shown in Fig. 1, but may be observed from one or several sides. Secondly, a group must have certain homogeneity. Motes located too far from the group's centroid may be useless to the localization effort in low Signal to Noise Ratio (SNR) environments or when the sound emitted by the source of interest is too weak. Furthermore, non-homogeneous groups present additional challenges for wireless communication.

Fusion nodes of the WSN perform mote grouping during network initialization and later participate in localization. For an orderly configuration of motes a single fusion node may be assigned to coordinate the activity of the whole WSN. In a random configuration each sensor mote may be a part of several groups and each group may be governed by several fusion nodes. In order to ensure coverage of all groups, fusion nodes reach an agreement concerning which node will govern which mote group. In this process communication signal strength is taken into account, meaning that a fusion node will adopt a group, to which it has the strongest connection. However, if there exists an ungoverned group, a redundant (i.e. covering an already covered group) fusion node closest to it will switch to that group. Mote communication is further discussed in Section III-C.

B. Acoustic Source Localization

Acoustic localization consists of estimating the DOA and distance to the sound source. DOA, in turn, consists of estimating the Angle of Arrival (AOA) and elevation of the acoustic wave front. As we operate only in the horizontal plane, we assume zero elevation, and thus for DOA estimation only the AOA is needed to be computed. The AOA, as it was mentioned earlier, is calculated based on TDOA. In choosing the trigonometric approach to AOA calculation an assumption of near or far field source location must be made. As sound waves propagate spherically, wave front curvature must be accounted for in the calculations. The near field disposition assumes spherical fronts, whereas waves originating in the far field are spread enough by the time they reach the sensor to be considered linear. The far field assumption is met for a linear microphone array if the inequality

$$|r| > \frac{2\left(Md\right)^2}{\lambda_{\min}} \tag{2}$$

holds, where M is the number of microphones, d is the inter-microphone distance, $\lambda_{\min}$ is the minimal wave length of the wide-band acoustic signal and r is the radial distance from the array center to the source. For our implementation we assume the far field disposition.

There exists a variety of methods for acoustic localization, most of which also employ TDOA as a basic principal. The methods utilize sensor array structures, in which a large number of microphones is arranged in some specific manner (e.g. linear, tetrahedron, spherical, etc.). The TDOA and consequently DOA is generally estimated using some measure of correlation between different sensor signals. For example a popular method of Steered Response Power with Phase Transform (SRP-PHAT) computes cross-correlation across all pairs of microphones at the theoretical time delays associated with all possible DOA to estimate the cumulative signal energy for each discrete point of the FOV [15]. MUltiple SIgnal Classification (MUSIC) applies eigenspace analysis to the signal correlation matrix in order to get the largest eigenvalues corresponding to the most probable DOA [16]. Multilateration methods estimate distances from every sensor to the source and calculate the position of that source using geometry of triangles and circles (spheres for 3D cases). Distance estimation is usually based on TDOA [17].

Notice that typical acoustic localization methods utilize information from every sensor. This fact does not pose a problem for wired systems with a single powerful computational hub. In WSN, however, collecting raw signals from nodes is a real challenge, especially if the number of nodes is large and signal frames are long. Recent developments in distributed localization combine individual sensor estimates for source positioning by applying, for example, maximum likelihood iterative search [18], or fuzzy clustering [19]. We try to overcome problems associated with communicating signal frames by applying a simplified localization approach of Initial Search Region Reduction (ISRR), recently developed by our research group.

C. Initial Search Region Reduction in WSN

The main idea behind ISRR lies in maximally confining the region of acoustic source disposition as a preliminary procedure to SRP-PHAT or other localization method [12]. Having already established that SRP-PHAT requires raw information from all sensors in the network, we do not apply it for this specific implementation. For object or person monitoring applications, where localization to a single point is not obligatory, ISRR confined regions serve as a sufficient estimate of object location. This section presents ISRR for the specific implementation of dual sensor mote WSN.

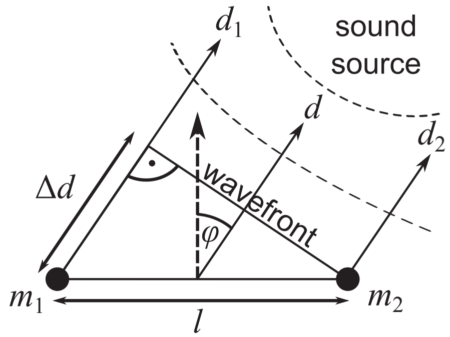


Fig. 2. DOA estimation for a pair of microphones.

Having a group of K dual sensor motes, the ISRR is performed in the following steps:

1) Estimate the DOA for each of K motes.
2) Generate vectors spanning from the mote sensor pair 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 reside, in rectangles.
5) Perform control of false detection, discard areas not meeting specific criteria (optional).

Step 1 is performed on each sensor mote, steps 2–5 are performed on the group's fusion node.

The DOA are estimated for the front view of the sensor pair, i.e. from -90° to 90°. Considering Fig. 2, the sound wave emitted by a source in the far field is acquired by the microphones m_1 and m_2 with a time delay $\tau = \Delta d / c$, where c is the speed of sound in m/s. The delay takes the values $\tau \in [-\tau_{\max}, \tau_{\max}]$, where $\tau_{\max}$ is the delay of sound traveling directly from one microphone to the other (i.e. at $\pm 90^\circ$). To estimate τ we apply cross-correlation to the two signals:

$$R(\tau) = \sum_{k=0}^{n-1} x_{m_1}(k) \cdot x_{m_2}(k - \tau), \tag{3}$$

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

$$\varphi = \sin^{-1} \frac{\tau \cdot c}{l} = \sin^{-1} \frac{\Delta k / f_s \cdot c}{l}, \tag{4}$$

where l is the distance between the microphones and τ is represented in terms of delay in samples Δk and the sampling frequency f_s. The speed of sound in air is dependent on the ambient temperature and is equal to

$$c = 331.45 \sqrt{1 + \theta / 273}, \tag{5}$$

where θ is the air temperature in Celsius.

At this point AOA validation is performed. If the correlation peak is not sharp and outstanding enough, the AOA φ is discarded. This way, in absence of a sound source or in case of high ambient noise, invalid estimates are avoided early on. We use the deviation from the mean for this metric:

$$\max\left(R(\tau)\right) > (1 + TH) \cdot \overline{R(\tau)}, \tag{6}$$

where TH is the threshold of deviation, which depends on the SNR in the environment. We use $TH = 0.2$ in our experiments. The angles φ from every mote are sent as DOA estimates to the fusion node.

The fusion node receives $K_1 \leq K$ DOA estimates ϕ_{i^*}, $i^* \in (1, \ldots, K_1)$ and adds the mote's 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 sensor pair'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 pair's center. Intersection points of all pairs $\overrightarrow{AB}_h$, $\overrightarrow{AB}_k$ are calculated by

$$\mathbf{I}_{hk} = (I_x, I_y) = \\ \left(\frac{(x_{1,h}y_{2,h}-y_{1,h}x_{2,h})(x_{1,k}-x_{2,k})-(x_{1,h}-x_{2,h})(x_{1,k}y_{2,k}-y_{1,k}x_{2,k})}{(x_{1,h}-x_{2,h})(y_{1,k}-y_{2,k})-(y_{1,h}-y_{2,h})(x_{1,k}-x_{2,k})}, \right. \\ \left. \frac{(x_{1,h}y_{2,h}-y_{1,h}x_{2,h})(y_{1,k}-y_{2,k})-(y_{1,h}-y_{2,h})(x_{1,k}y_{2,k}-y_{1,k}x_{2,k})}{(x_{1,h}-x_{2,h})(y_{1,k}-y_{2,k})-(y_{1,h}-y_{2,h})(x_{1,k}-x_{2,k})} \right). \tag{7}$$

As a result we have a set of $\mathbf{I}_{i^{**}}$ intersections, $i^{**} \in (1, \ldots, K_2)$, $K_2 \leq \binom{K_1}{2}$. 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 $\mathbf{I} = \emptyset$ **THEN** no partitions, $\mathbf{P} = \emptyset$ **STOP**
2) **ELSE IF** only 1 point $\mathbf{I}_1$ **THEN** $\mathbf{P}_1 = \mathbf{I}_1$ **STOP**
3) **ELSE** number of partitions $j = 0$
4) **WHILE** $|\mathbf{I}| > 0$, where $|\mathbf{I}|$ is the cardinality of the set $\mathbf{I}$, calculate centroid of free points $C_{\mathbf{I}} = 1/|\mathbf{I}| \cdot \sum \mathbf{I}$.
5) Calculate Euclidean distance of all free points to centroid $D_k = \sqrt{\sum_{s=1,2} \left(I_{k,s} - C_{\mathbf{I},s}\right)^2}$, choose point with minimal distance, $j = j + 1$, insert point to $\mathbf{P}_j$, remove point from set of free points $\mathbf{I}$.
6) Calculate partition centroid $C_{\mathbf{P}_j} = {}^{1}/_{|\mathbf{P}_j|} \cdot \sum \mathbf{P}_j$, get Euclidean distance for all free points $D_k = \sqrt{\sum_{s=1,2} \left(I_{k,s} - C_{\mathbf{P}_j,s}\right)^2}$.
7) **IF** $\min(D) \leq D_{\max}$ **THEN** insert point corresponding to $\min(D)$ into P_j, delete point from set of free points $\mathbf{I}$, **GO TO** Step 6.
8) **ELSE IF** $|\mathbf{I}| > 1$ **THEN GO TO** Step 4.
9) **ELSE** $j = j + 1$, put last remaining point to $\mathbf{P}_j$.

After obtaining the partitions $\mathbf{P}$, their areas are enclosed by rectangles with the edges denoted by the partitions lowest leftmost and highest rightmost points. As a result several regions may occur in the same FOV. Also while a vector from one array may cross with several other vectors, redundant 'echoing' regions may arise. These can be removed by applying SRP-PHAT to every region and comparing SRP values, or by using tracking filters. This work, however, does not focus on redundant region removal.

The procedure is applicable to multiple target localization. If more than two sensors are used in the array, several AOA may be estimated [12]. Each dual channel mote, however, points to a single direction of the strongest acoustic source. As sound pressure decreases exponentially with propagation, each mote group identifies a source closest to it. If a group is well spread, several targets may be identified within the FOV based on the same principle.

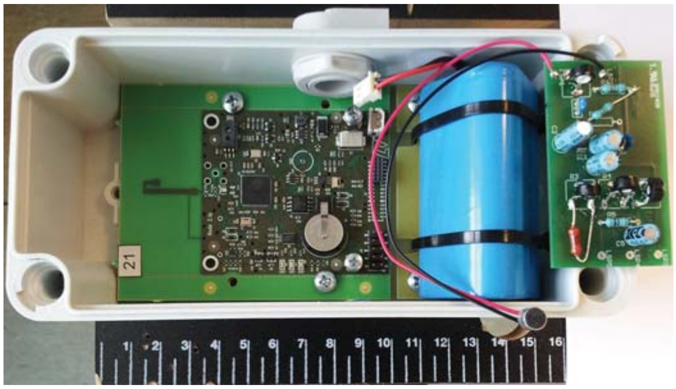


Fig. 3. Packaged WSN mote with a sensor amplification circuit (scale in cm).

III. WSN Implementation

The proposed distributed localization method with ISRR is implemented on a small WSN comprising several smart dust motes. The motes are equipped with Atmel ATmega128RFA1 microprocessors, which conveniently provide an on-chip AD converter for signal acquisition and a radio transceiver for WSN communication. The microprocessor has a clock speed of 16 MHz and provides 16kB of SRAM memory for operation with an additional 128kB of flash memory for program code. The on-chip Analog to Digital Converter (ADC) has a resolution of 10 bits and is able to sample with rates up to 330 kHz. However, actual experiments were carried out with a sampling rate of 2 kHz for each microphone channel, since higher sampling rates provided inconsistent and erroneous results during data acquisition. We were able to determine that erroneous results were caused by signal leakage from the previous ADC channel to the succeeding channel when switching between channels, but the cause of the leakage could not be determined. Mote-to-mote communication was realized with the IEEE standard 802.15.4 compliant radio transceiver with an effective indoor communication range of approximately 30 meters. The IEEE 802.15.4 standard supports transfer rates up to 250 kbit/s and packet sizes not larger than 127 bytes. Vansonic PVM-6052 electret condenser microphones were used for acoustic signal acquisition with additional circuitry performing signal amplification and the normalization needed for the microprocessor ADC input. For every mote a pair of microphones was mounted facing the same direction on a plastic board, which was then attached to the mote's plastic chassis.

The smart sensor mote chosen for the experiments is presented in Fig. 3. Microphone amplification circuitry is situated on the right and the microphone itself — in the bottom right corner. The mote is powered by a 3.7 V, 6600 mAh battery block (left from the sensor circuit). The motes are packaged in protective frames 16 cm in length. The poor computational characteristics listed above are typical for smart sensor motes. The reason for this is that these motes must work ubiquitously and autonomously with the battery they are provided for as long as possible. For example, the battery used in our configuration can sustain the motes for 1–1.5 years in a low duty cycle mode and approximately a month in constant

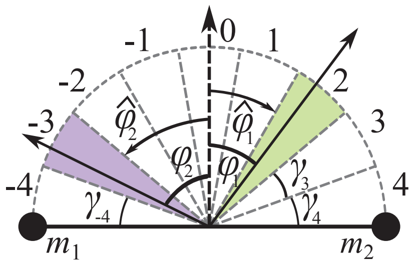


Fig. 4. Discretization of the AOA scope, defined by $\Delta k_{\max}$.

TABLE I
INTER-SENSOR DISTANCE FOR DIFFERENT SAMPLING RATES

$\Delta k_{\max}$	n_{AOA}	l (cm)		
		$f_s = 44.1$ kHz	$f_s = 2$ kHz	$f_s = 500$ Hz
1	3	0.8	17.2	68.7
2	5	1.6	34.4	137.4
3	7	2.4	51.6	206.1
10	21	7.8	171.7	686.8

operation mode. The goal here is to show that if localization and ISRR can be carried out on a smart sensor mote network, it is reasonable to assume it can also be implemented on larger networks with computationally more powerful motes.

A. *Implications of Using Low Sampling Rates*

The essential operation for AOA estimation is the signal cross-correlation (3). As our time delay τ is bounded by $\tau_{\max}$ and τ is expressed in delay in samples Δk, then Δk is also bounded by a maximal sample shift $\Delta k \in [-\Delta k_{\max}, \Delta k_{\max}]$, where $\Delta k_{\max}$ is calculated as

$$\Delta k_{\max} = \left\lfloor l \cdot \frac{f_s}{c} \right\rfloor, \tag{8}$$

where $\lfloor \cdot \rfloor$ denotes rounding to the largest previous integer (floor function). Consequently the view scope of the sensor pair is reduced to the number of possible AOA values $n_{AOA} = 2 \cdot \Delta k_{\max} + 1$. Fig. 4 depicts a view scope divided into 9 sectors. For any actual AOA φ, only its discrete margin $\hat{\varphi} \in [\gamma_{-\Delta k_{\max}}, \gamma_{\Delta k_{\max}}]$, corresponding to the correlation maximum, can be estimated. For devices capable of only low sampling rates this poses a problem in terms of compromise between the values of l and $\Delta k_{\max}$. Consider Table I. The standard CD sampling rate of 44.1 kHz is used for reference and l is calculated using (8): $l = \Delta k_{\max} \cdot c/f_s$. The table shows that to provide even the smallest n_{AOA} the inter-sensor distances must be quite considerable at low f_s. It is clear that if mote dimensions do not exceed 15–20 cm, it would not be reasonable to make $l = 1.7$ m to provide the sensor scope with 21 possible AOA.

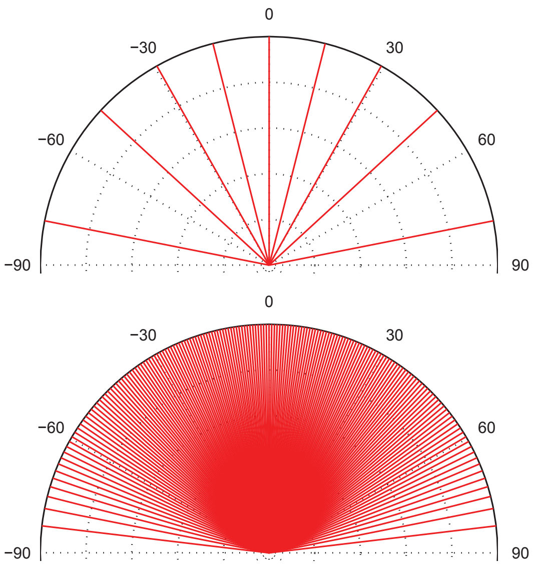
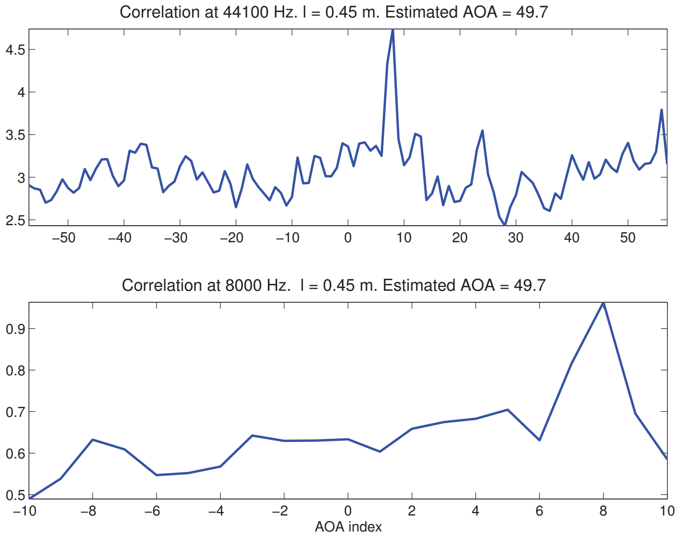


Fig. 5. Possible values of AOA for a sensor pair with $l = 0.7$ m and sampling rate $f_s = 2$ kHz (top); $f_s = 44.1$ kHz (bottom).


Correlation at 44100 Hz. l = 0.45 m. Estimated AOA = 49.7

Correlation at 8000 Hz. l = 0.45 m. Estimated AOA = 49.7

Fig. 6. Results of signal cross-correlation at different sampling rates.

We chose $l = 0.7$ m for our motes, which gives $n_{AOA} = 9$ possible AOA values with an average step of 19.7° at the used rate of $f_s = 2$ kHz. These are calculated using (4) and presented in Fig. 5. The substantial difference with high sampling rates is also evident from the figure — for the same sensor distance at a rate of $f_s = 44.1$ kHz, the AOA number is equal to 179 with an average step of 0.92°. A small n_{AOA} introduces additional error into the localization process as the ISRR estimated regions may become larger and get shifted from the true area occupied by the sound source. For example an angle step of 19.7° can give an error of 1.8 m if the sound source is situated only 5 m away from the sensors. To manage the situation a large number of motes must be used, preferably steered by different angles α (i.e. not facing in exactly the same direction). Random mote placements allow AOA uncertainty regions to superimpose on one another thus reducing the discrete gaps.

Low sampling rates also influence cross-correlation in two ways. Firstly, if the signal contains many strong components in the higher frequencies, they are not acquired at low sampling rates. As a result aliasing may occur, which in turn reduces the correlation reliability. A precise peak corresponding to a single Δk loses its steepness and spreads to several values. This makes AOA estimation and the control metric (6) less

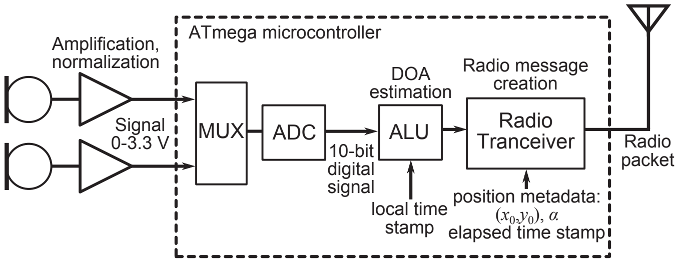


Fig. 7. Sensor mote architecture schematic.

reliable. Secondly, the cross-correlation yields exactly n_{AOA} coefficients, and if this number is low, the correlation peak cannot stand out from the average correlation level as much as in the case of high sampling rates. At low SNR the peak becomes almost uniform with the average level and control metric (6) declares the result invalid for the majority of signal frames. Both effects are evident from Fig. 6. The upper subplot shows the result of cross-correlation of two signals sampled at 44.1 kHz and the lower — at 8 kHz. For both cases the inter-microphone distance was equal to $l = 0.45$ m and the AOA from the acoustic source, as well as signal power, were the same. As n_{AOA} is more than five times larger in the case of $f_s = 44.1$ kHz and more signal energy information is contained in a single frame, the correlation peak is much more steep and evident than in the $f_s = 8$ kHz case. Generally at a fixed f_s correlation results are improved by increasing the signal frame length, thus providing more signal energy information. Here a compromise between correlation result reliability and the device refresh rate, as well as the amounts of required memory must be reached.

B. Resource Management and Scheduling

The smart sensor mote must divide its computational resource between two main tasks: sampling the ADC and performing cross-correlation (3) on the sensor signals. With the current hardware setup and computational power of the motes, sampling the ADC and doing correlation calculations at the same time is not possible, therefore currently these tasks are performed separately. A simplified schematic of sensor mote architecture and computational steps is presented in Fig. 7. First the ADC samples both channels and obtains a 0.2 second frame (totaling 400 samples at 2 kHz sampling rate) from each channel. Since we have a single ADC, both channels cannot be sampled simultaneously. Therefore there is a channel switching delay of about 150 microseconds. The phase shifts between channels due to these switching delays can be accounted for and they do not affect cross-correlation calculations significantly.

After the frames have been acquired, the ADC is stopped and the cross-correlation of frames is calculated. If a sound source is detected in the FOV of the sensor mote and the DOA of sound waves is found, then this information along with the spatial and temporal metadata is broadcast over the sensor network. Calculating the cross-correlation of frames and composing and transmitting the DOA message does not require much time (50 – 100 ms). Nevertheless, the smart sensor mote keeps track of this elapsed time and right before broadcasting the DOA message, appends the elapsed time to the message. Elapsed time is the time difference between the moment when the acoustic data reaches the processing unit and the moment of message composition. It indicates the 'age' (in milliseconds) of the calculated DOA value.

The fusion node has also two key tasks: listening to messages broadcast by sensor motes and performing sound source region estimation. As mentioned in Section II-A, sensor motes are partitioned into groups. Group membership is established by position metadata (coordinates and steering angle) found in the messages. Every fusion node maintains a lookup table consisting of coordinates of motes, which the node includes in its group. The table is updated every time a mote with new coordinates appears. A fusion node is only interested in messages arriving from motes in its own group and discards others. Sound source region estimation is performed when enough DOA messages with fresh data have been received. When the last data received becomes too old to be useful and no new data is received then the fusion node switches to idle mode until new messages arrive. Currently the expiration time for DOA information is 3 seconds, after which the fusion node discards the data.

C. Communication Strategy and Real-Time Operation

The benefit from using smart sensors is that preliminary sensor signal analysis is done on spot. With large networks it is not conceivable that raw sensor data is forwarded to some fusion unit. With the sampling parameters proposed for the WSN experiment (2 kHz sampling rate on both channels and frame lengths of 400 samples) it would take one sensor node in ideal conditions at least 0.15 seconds to communicate its entire measurement buffer to a fusion node. It is clear that with only a small number of motes the communication channel will be congested and system operation will be paralyzed. Hence, it is necessary to perform signal processing on the sensor motes and only communicate forward the results. This is what is proposed in our approach.

In the WSN experiment the communication scheme is built upon indirect messaging, i.e. motes broadcast the messages, which are to be received by fusion nodes. The indirect approach offers a quick and robust method for validating the data processing algorithm. The approach makes the network scalable to a certain extent and easily integrable into a larger system of systems [20].

In the algorithm validation experiment the sensor nodes broadcast messages about their latest DOA estimates and include their location and orientation metadata as well as temporal metadata with the broadcast message. A fusion node receiving these messages collects the DOA and metadata information and regularly initiates the ISRR algorithm to locate possible sound sources. Note that communication is performed in one direction — from sensor motes to fusion nodes. Therefore sensor motes do not possess any information concerning group partitioning.

The proposed WSN communication strategy is asynchronous, i.e. sensor motes do not have a global clock, which

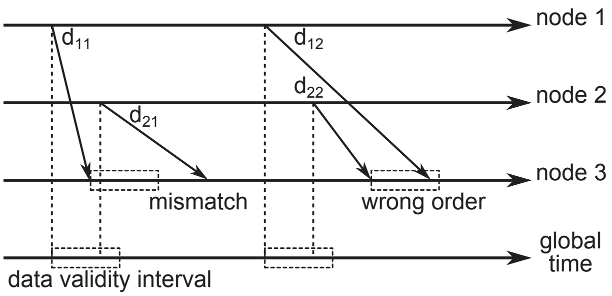


Fig. 8. Main problems situated with asynchronous data interchange.

would enable coordination of signal acquisition and message broadcasting. Rather every mote transmits a DOA estimate after every signal acquisition and processing loop (in our case it lasts 250–300 ms). The receiving fusion node can then estimate the time of DOA calculation in its own local time using the elapsed timestamp value and the common understanding of the millisecond time unit. For real-time operation two parameters must be strictly defined: the maximum duration of the DOA estimation loop and the maximum communication and processing delay for incoming messages. This must be done to enable estimating the validity of DOA estimates for position estimation. The asynchronous decentralized approach described above is robust and simple, suitable for algorithm validation, however in operational systems better control over data paths is desirable, which can be achieved by applying proactive middleware, as described in the next section.

D. Proactive Middleware and Data Validation

Performing computations in dynamic ad-hoc wireless sensor networks presents many challenges in terms of guaranteeing data correctness. The data used in the fusion process must satisfy certain temporal and spatial constraints (i.e. its age must not be greater than a pre-specified value or come from a certain location). It is easy to achieve such guarantees in a system with a fixed configuration, however in a dynamic setting the systems must evaluate these data properties at runtime.

For effective acoustic localization the DOA calculations ideally must be performed simultaneously. In real conditions a time interval must be specified in which the estimates are considered temporally coherent. Due to undefined transmission delays the data may arrive with considerable delays and therefore not satisfy the coherence requirement. Coherent data is vital for all signal processing tasks, like tracking and trajectory estimation. Fig. 8 graphically explains the validity interval mismatch and wrong order of message arrival. The validity interval specifies the time period, during which the data is considered coherent. Due to undetermined transmission delays the coherent messages may not fit in the interval. On the other hand, they may arrive in time, but in a wrong order, which may later cause errors in tracking procedures.

In [20], [21] we have presented the concept of proactive middleware, called ProWare, which is a lightweight distributed middleware component running on every element of the WSN system (see Fig. 9). ProWare implements a subscription based information exchange scheme, where the data consumer (fusion node) can subscribe for data from the providers (sensor nodes computing DOA estimates). ProWare also handles data validity checking ensuring that the data received at the fusion node satisfies the constraints for a given fusion operation (i.e., that the data is temporally coherent). ProWare manages the process of finding appropriate data providers (in our case sensor motes with overlapping fields of view) and setting up the data exchange paths with the consumers (fusion nodes). Both the validity checking and provider-consumer agreements are performed on-line. Among other tasks the middleware component keeps track of the different clock offsets of the motes and regularly checks and updates the change (caused by clock drift, jitter etc.) in these offsets. This temporal information is then used to estimate the time of measurement of the data in local time of the data consumer.

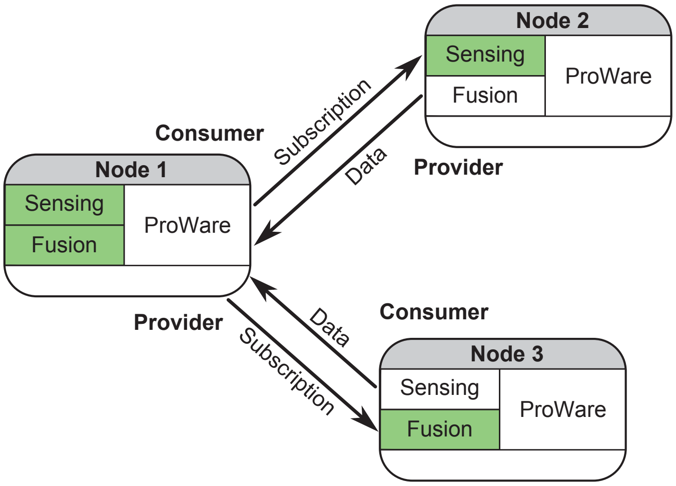


Fig. 9. Principle diagram of proactive middleware mediation.

IV. Experimental Results

We demonstrate the applicability of our proposed method of acoustic localization and the implications situated with using low sampling rates by performing an experiment of single speaker localization. For the initial experimental validation we performed data acquisition using an Agilent U2354A data acquisition device (DAQ) and performed localization offline in the MATLAB environment. Data acquisition is performed at two sampling rates: $f_s = 8$ kHz and $f_s = 2$ kHz per sensor (as the motes are able to sample the signal at 2 kHz). For the experiment we use four microphone pairs arranged in an angular configuration — two microphone pairs are placed perpendicular to the other two. For comparison with our proposed approach we apply the SRP-PHAT method, which was mentioned in Section II-B. The fact that SRP-PHAT is a highly resource demanding procedure is another reason for choosing MATLAB for computations. The ISRR procedure implemented in MATLAB is identical to the one running on the motes, therefore there is no difference in localization. Mote communication and asynchronous data validation is not considered in this experiment. Additionally we apply simplified SRP-PHAT with Stochastic Region Contraction (SRC) to the

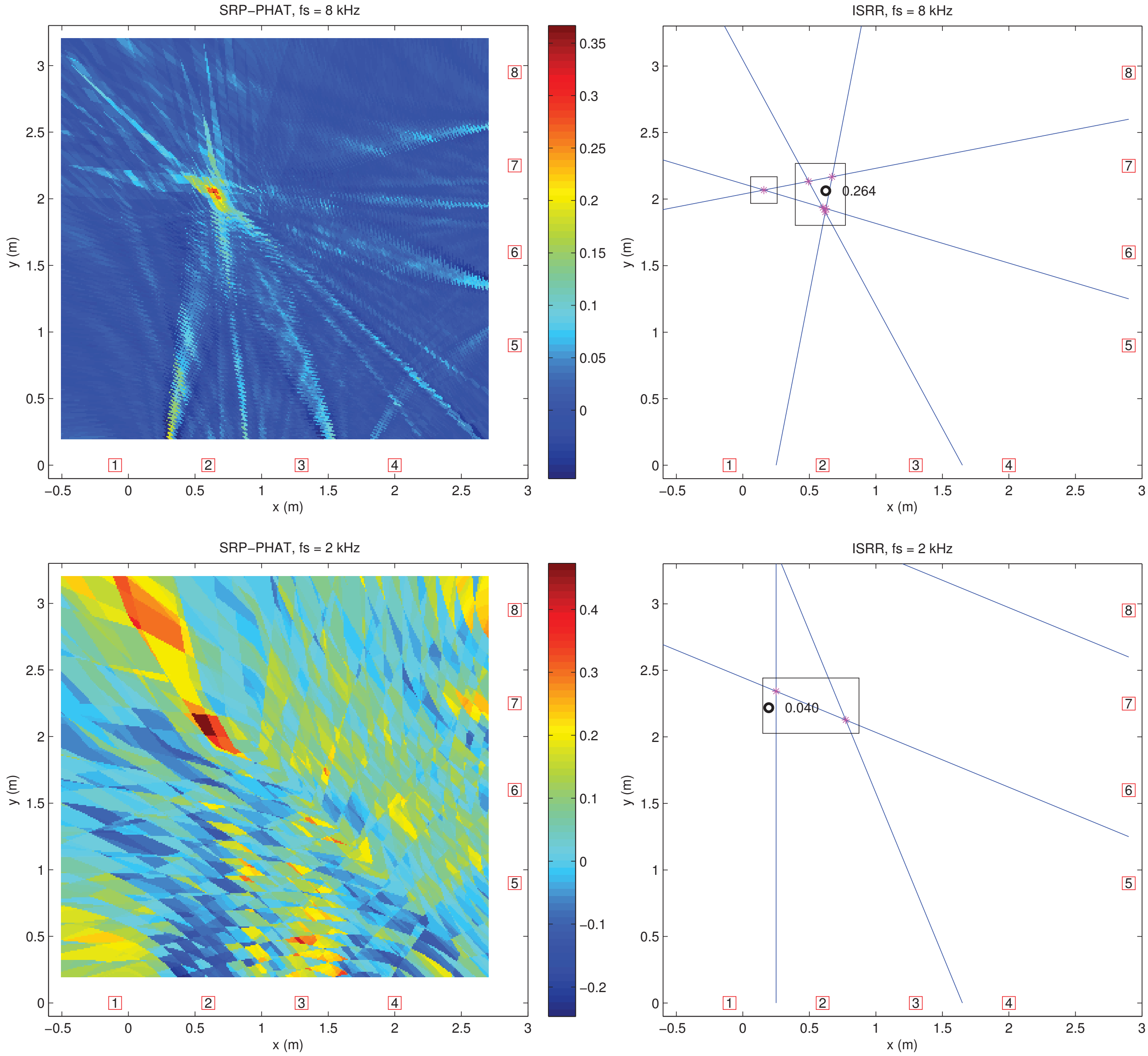


Fig. 10. Results of acoustic source localization using four pairs of microphones and two approaches: SRP-PHAT and the proposed ISRR.

source areas estimated by ISRR in order to determine the cumulative acoustic energy levels in these areas. The principles of SRP-PHAT and SRC are briefly introduced in the Appendix. For better comprehension of the result analysis it is advised to get familiarized with these localization methods.

The results of speaker localization are presented in Fig. 10. The speaker is placed at position $(0.7, 2)$ meters and a short speech recording is made. We analyze the signal frame by frame, as it is done in our WSN implementation, using a frame length of 200 ms. For SRP-PHAT the area discretization value is set to 1 cm^2. Fig. 10 portrays localization results for a single signal frame related to the same time instance. In the figure microphones are represented by red squares with numbers inside them. The four microphone pairs are then 1, 2; 3, 4; 5, 6; 7, 8. SRP-PHAT results of cumulative energy values are plotted using a red-green-blue color scale. DOA vectors of ISRR are denoted by blue lines, their intersections — by magenta stars, the estimated regions — by black rectangles and the maximum of SRP-PHAT with SRC — by black circles.

It can be seen from the two upper plots of Fig. 10, that at $f_s = 8$ kHz both SRP-PHAT and ISRR localize the sound source efficiently. The SRP-PHAT region of particularly high cumulative energy (i.e. orange to red on the scale) is reduced to approximately 0.01 m^2. The region estimated by ISRR is significantly larger, but proportionate to the SRP-PHAT region of medium cumulative energy (i.e. green on the scale). However, the sound source is fully confined by the region, as confirmed by the SRP-PHAT with SRC estimate on the region.

On the other hand, both methods suffer from the problems situated with the low sampling rate of $f_s = 2$ kHz, as it can be clearly seen in the lower plots of Fig. 10. SRP-PHAT high cumulative energy region enlarges to approximately 0.25×0.3 meters with an 'echoing' region situated in the top left corner of the FOV. The decrease of n_{AOA} and the number of signal samples in a frame, described in Section III-A, affects SRP-PHAT, producing a more rough and less comprehensive image. Nevertheless, SRP-PHAT localizes the source properly. ISRR performs worse, missing the source slightly up the y-axis. The reason lies in the fourth microphone pair failing to estimate the DOA correctly. Although the confined region is close to the source position, it does not confine it fully. This example clearly shows the need for a larger number of microphone pairs (motes) to be used for successful localization.

Generally the performance of ISRR is comparable to the performance of SRP-PHAT in terms of localization accuracy, with the deviation of ISRR estimates from the localization results over the whole FOV being less than 0.13 m [12]. Considering the dimensions of usual acoustic sources under localization (no less than 15–20 cm), this deviation is quite permissible. In the case of low sampling rates, however, both SRP-PHAT and ISRR become less reliable. Therefore it cannot be explicitly stated, that our proposed method suffers from the limitations of embedded hardware more than the other. On the contrary, ISRR significantly reduces the number of computations required for localization up to an area of a fraction of a square meter [11].

V. Discussion and Future Work

The paper mainly considers the limitations of data processing hardware and touches upon the implications of asynchronous data interchange slightly. The proactive middleware component, which has been tested on abstract data, was not fully implemented for the purposes of acoustic localization on a large number of smart dust motes. Thus further testing on a large network of motes is required in order to estimate the feasibility of the component for our specific purpose.

Considering our recent developments, we have achieved a reliable 4 kHz sampling rate per channel on the Atmel ATmega128RFA1. As a consequence, the localization quality has noticeably improved. The problem of signal leakage when changing channels is still not solved, but the higher sampling rate was reached by changing the ADC clock speed, which consequently affected the settling time (i.e., the time automatically inserted by hardware to clear and prepare the ADC registers for channel change) in-between changing ADC channels and alleviated the signal leakage problem. We do not plan on improving this hardware platform any further, instead we are considering more powerful embedded devices, such as the Gumstix Overo series, for our localization approach. The implementation on Atmel ATmega displays promising results, however, it requires a significant number of motes and their significant dispersion in the FOV to sustain the localization quality on low sampling rates, and not every application and environment will allow these things. For the applications, where only a few motes are permitted, the motes must estimate the DOA more accurately, and that requires more resources. With the increase of computational power it will be also possible to increase the number of sensors on each mote, which will increase the reliability of DOA estimates.

Increasing the number of sensors per mote will also allow for 3D acoustic localization. As the elevation AOA cannot be accurately estimated by a pair of horizontally placed microphones, 3D localization will require additional microphones to be utilized to estimate the AOA in the vertical plane. For this direction of future research the proposed ISRR method is to be expanded in order to be able to confine volumetric regions, as opposed to planar areas, discussed in this paper.

VI. Conclusion

The paper considers an acoustic source localization system and its implementation in a WSN consisting of dual channel low power smart sensors. A decentralized ad-hoc WSN structure for distributed computation is proposed, which reduces the number of computations per network node and introduces redundancy to the system, making it more reliable. The applied localization approach is presented and different problems situated with system implementation on specific hardware are handled. Computationally weak smart sensor hardware imposes limitations on the signal sampling rate, processing time and communication bandwidth. A compromise between a reliable sampling rate, suitable sensor pair geometry and localization accuracy is established. The applied asynchronous communication strategy reduces message interchange and does not overwhelm the network's fusion nodes. A practical experiment is held to test the proposed localization method and compare it to a popular and effective, but resource demanding approach. Experimental results show, that both methods suffer from the limitations induced by low power embedded hardware. However, the proposed method is capable of localization with permissible accuracy.

Appendix
Overview of SRP-PHAT and SRC

Steered Response Power with Phase Transform (SRP-PHAT) is a technique of estimating the DOA of sound signals. The SRP $P(\vec{a})$ is a real-valued functional of a spatial vector $\vec{a}$, defined by the FOV of a specific microphone array. The high maxima in $P(\vec{a})$ indicate the estimates of sound source location. $P(\vec{a})$ is computed for each direction as the cumulative Generalized Cross-Correlation with Phase Transform (GCC-PHAT) 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 times of sound arrival from point a to the two microphones are $\tau(a,k)$ and $\tau(a,l)$ respectively. Hence the time delay between the two signals is $\tau_{kl}(a) = \tau(a,k) - \tau(a,l)$. The SRP-PHAT for all pairs of signals is then defined as

$$P(a) = \sum_{k=1}^{M} \sum_{l=k+1}^{M} \int_{-\infty}^{\infty} \Psi_{kl} X_k(\omega) X_l^*(\omega) e^{j\omega\tau_{kl}(a)} \mathrm{d}\omega, \quad (9)$$

where $X_i(\omega)$ is the spectrum (the Fourier transform) of signal x_i and $X_i^*(\omega)$ is the conjugate of that spectrum. Ψ_{kl} is the PHAT weight of the inverse of spectral magnitudes:

$$\Psi_{kl} = \frac{1}{|X_k(\omega) X_l^*(\omega)|}. \quad (10)$$

Conventional SRP-PHAT performs as many evaluations (9), 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 (or volume) into small discrete regions. The method is highly resource demanding, particularly when applied to large areas of observation. The number of evaluations (9) is significantly reduced by applying Stochastic Region Contraction (SRC), which iteratively narrows down the search volume for the global maximum [15]. 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, then contracts the search volume into a sub-volume containing the desired global optimum and proceeds iteratively until the SRP maximum can be located with a finite precision.

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 (JET)*, vol. 59, no. 2, pp. 151–160, 2013.

[2] S. Astapov, J.-S. Preden, J. Ehala, and A. Riid, "Object detection for military surveillance using distributed multimodal smart sensors," in *Proc. 19th Int. Conf. on Digital Signal Processing (DSP 2014)*, 20–23 Aug. 2014, pp. 366–371.

[3] H. Lohrasbipeydeh, A. Zielinski, and T. Gulliver, "A new acoustic method for passive sperm whale depth tracking," in *Proc. IEEE Region 10 Conference TENCON 2012*, Nov 2012, pp. 1–5.

[4] J.-C. Wang, C.-H. Lin, E. Siahaan, B.-W. Chen, and H.-L. Chuang, "Mixed sound event verification on wireless sensor network for home automation," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 1, pp. 803–812, Feb 2014.

[5] Y. Lee, K. Kim, D. Han, and H. Ko, "Acoustic and visual signal based violence detection system for indoor security application," in *Proc. 2012 IEEE Int. Conf. on Consumer Electronics (ICCE)*, 2012, pp. 737–738.

[6] 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. Conf. DAAAM Baltic Industrial Engineering*, 2012, pp. 268–273.

[7] 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. Conf. Technologies for Homeland Security (HST)*, 2011, pp. 254–261.

[8] 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. Conf Networking, Architecture and Storage (NAS)*, 2012, pp. 245–251.

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

[10] G. Vakulya and G. Simon, "Fast adaptive acoustic localization for sensor networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 5, pp. 1820–1829, 2011.

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

[12] S. Astapov, J. Berdnikova, and J. S. Preden, "Optimized acoustic localization with SRP-PHAT for monitoring in distributed sensor networks," *International Journal of Electronics and Telecommunications*, vol. 59, no. 4, pp. 383–390, 2013.

[13] 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 Transactions on Mobile Computing*, vol. 7, no. 5, pp. 570–584, 2008.

[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*, 2009, pp. 289–298.

[15] 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*, vol. 1, 2007, pp. 121–124.

[16] 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.

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

[18] 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.

[19] Z. Merhi, M. Elgamel, and M. Bayoumi, "Acoustic target localization in sensor networks with FUZZYART," in *Proc. 50th Midwest Symp. Circuits and Systems MWSCAS 2007*, 2007, pp. 1536–1539.

[20] J. S. Preden, J. Llinas, G. Rogova, R. Pahtma, and L. Motus, "On-line data validation in distributed data fusion," in *SPIE Defense, Security and Sensing, Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR IV*, vol. 8742, 2013, pp. 1–12.

[21] J. S. Preden, L. Motus, R. Pahtma, and M. Meriste, "Data exchange for shared situation awareness," in *2012 IEEE Int. Multi-Disciplinary Conf. on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, March 2012, pp. 198–201.

Sergei Astapov received his M.Sc. degree in the field of Computer System Engineering at the Tallinn University of Technology in 2011. He continues his education as a PhD student at the Department of Computer Control at the Tallinn University of Technology and is a member of the Department's Research Laboratory for Proactive Technologies. His research interests include object tracking using wide-band signal analysis, classification tasks and distributed computing in embedded multi-agent systems. His recent research concerns object localization and identification in open environments and acoustic signal based diagnostics of industrial machinery.

Johannes Ehala received his M.Sc. degree in the field of Computer System Engineering at the Tallinn University of Technology in 2012. Currently he is a PhD student at the Department of Computer Control at the Tallinn University of Technology and is a member of the Research Laboratory for Proactive Technologies. His doctoral studies and research interests include self-organization and emergent behavior in cyber-physical systems and computational models of distributed computer systems. Currently he is involved in analyzing the temporal aspects of ProWare and developing a computational model to describe ProWare.

Jürgo-Sören Preden received his PhD degree in Computer Science from the Tallinn University of Technology, the topic being "Enhancing Situation-Awareness, Cognition and Reasoning of Ad-Hoc Network Agents". He is a senior researcher and the head of the Research Laboratory for Proactive Technologies at the Tallinn University of Technology. Jürgo's research interests are focused on distributed computing systems, more specifically on cognition and situation awareness of such systems. His spectrum of activities involves sensing technologies, data processing, computation and communication in ad-hoc sensing systems.