

Research Article

Indoor Self-Localization and Orientation Estimation of Smartphones Using Acoustic Signals

Héctor A. Sánchez-Hevia,¹ David Ayllón,^{1,2} Roberto Gil-Pita,¹ and Manuel Rosa-Zurera¹

¹Department of Signal Theory and Communications, University of Alcalá, Alcalá de Henares, Spain

²Research & Development Department, Fonetic, Madrid, Spain

Correspondence should be addressed to Héctor A. Sánchez-Hevia; hectoradrian.sanchez@uah.es

Received 19 May 2017; Accepted 22 August 2017; Published 25 September 2017

Academic Editor: Bowon Lee

Copyright © 2017 Héctor A. Sánchez-Hevia et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose a new acoustic self-localization and orientation estimation algorithm for smartphones networks composed of commercial off-the-shelf devices equipped with two microphones and a speaker. Each smartphone acts as an acoustic transceiver, which emits and receives acoustic signals. Node locations are found by combining estimates of the range and direction of arrival (DoA) between node pairs using a maximum likelihood (ML) estimator. A tailored optimization algorithm is proposed to simultaneously solve the DoA uncertainty problem that arises from the use of only 2 microphones per node and obtain the azimuthal orientation of each node without requiring an electronic compass.

1. Introduction

Locating the nodes in wireless networks is an essential step for many applications, where the location of the sensors gives meaning to the collected data. However, accurate knowledge about the nodes' locations and orientations is often not readily available. In indoor scenarios, where classic positioning systems such as GPS are not viable because of a lack of coverage or limited precision, it is common to resort to relative node distance and/or position measurements from acoustic, infrared, or radio frequency (RF) signals that are exchanged among devices. The most common measurements are time of arrival (ToA), direction of arrival (DoA), or angle of arrival and received signal strength (RSS) [1]. However, the use of these measurements is not straightforward because of the random component introduced by time-varying errors (e.g., additive noise and interferences) and environment-dependent errors (wall reflections, furniture obstructions, etc.). Traditional approaches for node localization rely on beacon nodes (sometimes called anchor nodes), whose position is known a priori to a certain degree. With the beacon nodes, the locations of the remaining sensors are estimated using multilateration or triangulation techniques [2, 3]. However, in ad hoc networks such as an opportunistic

network formed by smartphones, the probability of having beacon nodes is low because of their dynamic nature. Without the beacons, relative locations can be estimated using an arbitrary coordinate frame of reference, which is commonly called node self-localization. The relative location of the nodes provides us with sufficient spatial information to implement a wireless microphone array (WMA). WMAs have many potential applications in distributed audio processing, such as speech enhancement [4], blind source separation and echo cancellation [5], speaker localization and tracking [6, 7], and voice activity detection [8].

Current-generation smartphones pack sufficient hardware so that a group of devices with the correct software can be used for many applications such as indoor positioning, pedestrian tracking, smart cities, teleconferencing, and hearing impaired assistive technology [9]. However, despite their potential applications, smartphones have several hardware and software limitations that must be considered such as the limited number of available specialized sensors, their limited sampling rate, the lack of optimization of the operative system for real-time applications, and restricted hardware access.

Typically, the hardware in commercial smartphones is sufficient for different approaches to node self-localization. Some examples in the literature are the use of the RSS of RF

signals [10, 11], a combination of RF and ultrasonic signals [12, 13], and different data fusion schemes [14, 15]. In Höflinger et al. [16], the authors propose an acoustic-based system using high-pitch chirps and at least 3 known-location receivers to achieve a localization error of approximately 30 cm. Node orientations are commonly obtained using an electronic compass, which is composed of a magnetometer and an accelerometer; both sensors are readily available in most smartphones. Unfortunately, magnetometer measurements are sensitive to disturbances from electric equipment (and even large metallic objects) and must be frequently calibrated to avoid large errors [17]. Typically, RF-based solutions are intended for large areas (i.e., an entire building) because they can cover wider distances at the cost of localization errors in the range of meters, whereas acoustic-based methods are used for localization within a room and achieve errors in the tenths of centimeters.

When the localization procedure is based only on acoustic signals, we can discuss array geometry calibration [18]. This field encompasses different scenarios, of which *distributed array configuration calibration* is the most relevant for node localization because its objective is to infer the location and orientation of distributed microphone arrays with known local geometry (i.e., nodes with more than 1 microphone) using DoA measurements. A common approach is to assume a two-dimensional (2D) scenario as seen in Jacob et al. [19], where 4 arrays with 2 microphones are located to a precision up to 5 cm, which assumes that the nodes are located along the walls of a room of known dimensions. Similarly, in Plinge and Fink [20], 3 arrays with 5 microphones embedded on a table and synchronized to $22 \mu\text{s}$ are calibrated with a precision up to 1.2 cm and 1.3° using 300 s of white noise. In Anwar et al.'s work [21], nodes with 3 microphones and RSS measurement capabilities are located within an error of 11 cm and 1.7° . These proposals have in common the use of ad hoc hardware and all of them require 3 or more microphones per node to resolve 360° azimuthal orientation.

There are different types of self-localization methods such as those based on ToA measurements. Usually these approaches involve a number of acoustic sources and microphones at unknown locations from which the Time of Flight (ToF) between source-microphone pairs is obtained. The method described in Crocco et al.'s work [22] reports localization errors in the centimeter range. It represents an improvement upon classic methods such as Thrun [23] by introducing a closed-form solution as the initial state for the error function minimization.

In this work, we propose an algorithm for node self-localization and orientation estimation for smartphone networks using acoustic signals and assuming that each node is a state-of-the-art off-the-shelf smartphone with two microphones and a speaker. The algorithm is an extension of the ideas proposed in Ayllon et al.'s work [24], particularly a modification of the Maximum Likelihood-based Distributed Optimization for Node Localization (ML-DONL) algorithm in the said work. This modification does not require previous knowledge about node orientations. The main advantage of our proposal is that we avoid the error introduced by an uncalibrated compass, which is often in excess of 15° [17].

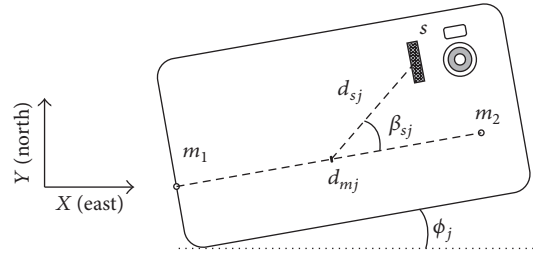


FIGURE 1: Configuration of the j th node.

Both the location and orientation estimates are based on closed-form expressions; however an optimization algorithm is used to resolve the DoA uncertainty needed to obtain 360° orientation estimates using only 2 microphones per node. RF signals are required for data exchange in the network, and the method assumes that the nodes are static during the localization procedure, which takes a few seconds. The proposed approach is intended for the localization of acoustic nodes in a room (there is line of sight between nodes) to create a WMA.

2. Problem Formulation

Let us consider a fully connected network composed of J nodes, where each node contains a microphone array of known geometry. If we also consider K acoustic sources that are emitted from unknown locations, we can obtain a series of DoA estimates from each node to each source, so that the network geometry can be found as a combination of all estimated angles by solving a minimization problem. However, DoA-based algorithms can only find the relative geometry, and additional information is required to scale it.

In our particular case, each node is a smartphone equipped with two microphones (m_{1j} and m_{2j}) and a speaker (s_j). Figure 1 represents a typical smartphone configuration that acts as the j th node, where d_{mj} is the distance between the microphone pair, d_{sj} and β_{sj} are the distance and angle between the center of the array and the speaker, respectively, and ϕ_j is the orientation of the node.

Our goal is to find the location and orientation of the J nodes that form the network. We focus on the 2D case, where all nodes lie on the XY plane, which is the typical scenario where various smartphones are resting on a table. A 3D generalization will require more than two microphones per node, which is an uncommon feature in current devices. Because we use an active approach by having the nodes as the sound sources, we define a node location as the location of its speaker. Then, the localization problem is reduced to the estimation of $2J$ speaker coordinates $\mathbf{p} = (x_1, \dots, x_J, y_1, \dots, y_J)^T$ and J orientations (azimuth) $\phi = (\phi_1, \dots, \phi_J)$ based on the combination of DoA and range estimates between node pairs.

2.1. DoA and Range Estimation. The proposed localization algorithm is based on the combination of $J \times J$ DoA (α_{jk}) and range (r_{jk}) estimates, where each α_{jk}, r_{jk} pair is an estimate

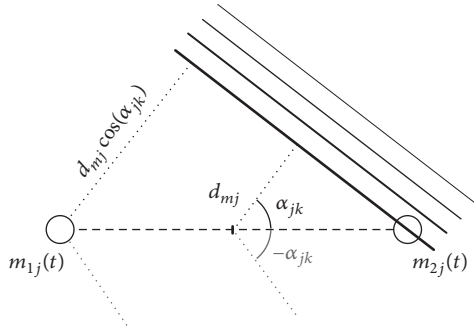


FIGURE 2: Illustration of the DoA estimation using a microphone pair with a source in their far field.

of the relative location of the k th speaker with respect to the j th node in polar coordinates.

Let us consider the microphones of node j as a linear array, so that if we assume that a source (k th speaker) is in the far field of the array, a plane wavefront impinges it with an angle α_{jk} . The DoA is obtained from the Time Difference of Arrival (TDoA) between the two sensors (see Figure 2), which is given by $\tau_{jk} = d_{mj} \cos(\alpha_{jk})/c$, where d_{mj} is the intermicrophone distance and c is the speed of sound. Unfortunately, a linear array (1D) in a 2D scenario can only discern DoAs between $-\pi/2$ and $\pi/2$ radians, which leads to a problem known as DoA uncertainty. Because $\cos(\alpha_{jk}) = \cos(-\alpha_{jk})$, for every τ_{jk} , there are two potential DoAs. Then, the measurement of the angle between node pairs is biased by the node orientation and affected by DoA uncertainty and measurement errors. Thus, we can define the estimated angle between node pairs (γ_{jk}) as

$$\gamma_{jk} = u_{jk} (\alpha_{jk} + \Delta\alpha_{jk}) + \phi_j \longrightarrow \gamma_{jk} \approx u_{jk} \alpha_{jk} + \phi_j, \quad (1)$$

where $u_{jk} = \{-1, 1\}$ is the DoA uncertainty correction variable and $\Delta\alpha_{jk}$ is the DoA measurement error. Please notice that, in Jacob et al.'s work [19], the DoA uncertainty is not considered as a problem because the 2-microphone arrays are always located along a wall, which eliminates the possibility of any sound impinging from the “back” of the array.

To obtain the DoA (α_{jk}) and range (r_{jk}) estimates between node pairs, each node emits a reference acoustic signal, which is received by every node in the network. Let these reference signals be known and denoted by $s_k(t)$, where k indicates the emitter node. In this work, we use the General Cross-Correlation PHASE Transform (GCC-PHAT) to obtain the DoA estimates because of its robustness to reverberation [25]. Let $Y_{1j}(\omega)$ and $Y_{2j}(\omega)$ be the Fourier transform of the signals received by the microphones of node j and let $S_k(\omega)$ be the Fourier transform of the reference signal emitted by node k . The GCC-PHAT of the microphone signals and reference signal is given by

$$R_{1jk}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{S_k(\omega) Y_{1j}(\omega)^*}{|S_j(\omega) Y_{1j}(\omega)^*|} e^{j\omega\tau} d\omega,$$

$$R_{2jk}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{S_k(\omega) Y_{2j}(\omega)^*}{|S_j(\omega) Y_{2j}(\omega)^*|} e^{j\omega\tau} d\omega, \quad (2)$$

where ω is the frequency and τ is the time lag. The time difference between the two signals corresponds to the point where the value of the GCC-PHAT function is at its maximum:

$$\begin{aligned} \tau_{1jk} &= \operatorname{argmax} (R_{1jk}(\tau)), \\ \tau_{2jk} &= \operatorname{argmax} (R_{2jk}(\tau)). \end{aligned} \quad (3)$$

Because we correlate with a known signal, τ_{1jk} and τ_{2jk} are the time of arrival (ToA) of that signal for each microphone. Then, the TDoA between microphones can be easily computed as the difference between ToAs: $\tau_{jk} = \tau_{2jk} - \tau_{1jk}$, from which the DoA is directly estimated as

$$\alpha_{jk} = \cos^{-1} \left(\tau_{jk} \frac{c}{d_{mj}} \right). \quad (4)$$

The range between node pairs is measured using ToF. Assuming that the nodes are synchronized, that is, every node in the network shares a common timebase and an identical sampling frequency f_s , the problem of range estimation is reduced to

$$r_{jk} = \frac{(tR_{1jk} + tR_{2jk})/2 - tS_k}{c} f_s, \quad (5)$$

where tS_k is the time when the k th node emits its signal and tR_{1jk} and tR_{2jk} are the time instants when the j th node receives that signal at both of its microphones (ToA). Notice that because the nodes are equipped with two microphones, we take the average of the ToAs to obtain the ToA at the center of the array. The specific methods to obtain internode synchronization fall outside the scope of this paper, although there are multiple solutions in the literature, for example, Sur et al. [26].

3. Proposed Node Localization Method

In this section, we explain how the DoA and range estimates taken by the nodes are combined in order to obtain their locations and orientations.

3.1. ML Estimator of Node Locations. Let us consider that a full set of estimations of the range r_{jk} and incidence angle γ_{jk} between node pairs is available, and each estimate has an error with standard deviation $\sigma_r(j, k)$ and $\sigma_\gamma(j, k)$, respectively. The objective is to estimate the position vector \mathbf{p} from the measurements considering the standard deviation of the measurements. Each polar measurement (azimuth and distance pair) is transformed into Cartesian coordinates $\mathbf{d}_{jk} = (v_{jk}, w_{jk})$, where $v_{jk} = r_{jk} \cdot \cos(\gamma_{jk})$ and $w_{jk} = r_{jk} \cdot \sin(\gamma_{jk})$, with $r_{jk} \geq 0$ and $\gamma_{jk} \in [-\pi, \pi]$ for all j and k from 1 to J .

Let us also consider the joint probability density function (PDF) of the measurements in Cartesian coordinates as a

multivariate normal distribution. In Ayllon et al.'s work [24], the next expression for the PDF is proposed

$$f_{jk}(\mathbf{P}) = \frac{1}{2\pi |\mathbf{C}_{jk}|^{1/2}} e^{-(1/2)(\mathbf{p}_j - \mathbf{p}_k - \mathbf{d}_{jk}) \mathbf{C}_{jk}^{-1} (\mathbf{p}_j - \mathbf{p}_k - \mathbf{d}_{jk})^T}, \quad (6)$$

where \mathbf{C}_{jk} is the covariance matrix of the PDF related to the measurement vector of the j th node to the k th node and \mathbf{p}_k is a column vector that contains the coordinates of the latter. It is possible to obtain the most likely node locations using a maximum likelihood estimator, where the log-likelihood L of a given geometry is calculated using the following equation:

$$L = \sum_{j=1}^J \sum_{\substack{k=1 \\ k \neq j}}^J \log(f_{jk}(\mathbf{P})). \quad (7)$$

Plugging (6) into (7) and simplifying, the next expression is obtained:

$$L = b - \frac{1}{2} \sum_{j=1}^J \sum_{k=1}^J (\mathbf{p}_j - \mathbf{p}_k - \mathbf{d}_{jk}) \mathbf{D}_{jk} (\mathbf{p}_j - \mathbf{p}_k - \mathbf{d}_{jk})^T, \quad (8)$$

where $b = -\log(2\pi |\mathbf{C}_{jk}|^{1/2})$ and $\mathbf{D}_{jk} = \mathbf{C}_{jk}^{-1}$.

Assuming that all the covariance matrices are equal and proportional to the identity matrix, so that $\mathbf{D}_{jk} = \rho \mathbf{I}$, with $\rho = 0$ when $j = k$, we can obtain the solution using the following expression:

$$\mathbf{p}_k = \left(\frac{1}{2J} \sum_{j=1}^J v_{kj} - v_{jk}, \frac{1}{2J} \sum_{j=1}^J w_{kj} - w_{jk} \right). \quad (9)$$

This is equivalent to assuming that the variables of the PDF are independent and their standard deviation is constant. This way, every estimation has the same weight and ρ has no effect on the localization result ($\rho = 1$). Please refer to Ayllon et al. [24] for a complete description of the ML location estimator. In this work, we are using the method denominated as ‘‘Naive Covariance Matrix Estimation.’’

Most of self-localization methods (including Jacob et al. [19] and Plinge and Fink [20]) use some kind of iterative optimization algorithm in order to find the node locations. It is common to minimize a pairwise distance error function such as

$$\epsilon = \sum_{j=1}^J \sum_{k=1}^K (r_{jk} - |\mathbf{p}_j - \mathbf{p}_k|)^2, \quad (10)$$

where r_{jk} is the measured distance (range) between nodes j and k (obtained either directly, i.e., ToF, or indirectly, i.e., TDoA triangulation) and $|\mathbf{p}_j - \mathbf{p}_k|$ is the distance between their estimated locations. However, it is important to note that our ML estimator is a closed-form method.

3.2. Orientation Estimation. To obtain γ_{jk} from α_{jk} , first, we must know the orientation of the j th node and solve the DoA uncertainty as shown in (1). Any error in the

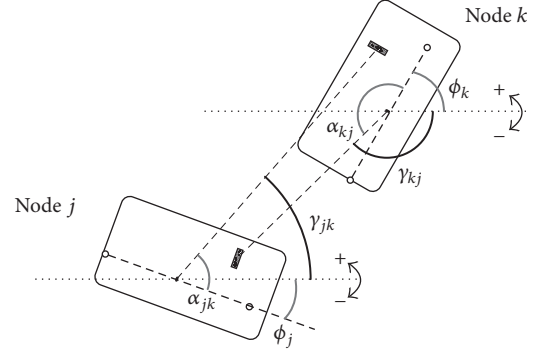


FIGURE 3: Angular relations between node pairs.

orientation estimation is directly added to γ_{jk} , which poses a problem for the estimation of the node locations. Because the digital compass in smartphones is commonly uncalibrated, it introduces a large error that frequently outweighs that of the DoA estimation. Thus, we decided to estimate the orientation of the nodes using the available information instead of relying on an imprecise measurement.

Let us consider that the nodes have their sound source at the center of their microphone array ($d_{sj} = 0$) and we know the value of the true angle between node pairs φ_{jk} (i.e., the actual value without any error). In this scenario, we know that $\varphi_{jk} - \varphi_{kj} = \pm\pi$ rad, for $k \neq j$. Now, if we introduce the approximation from (1), substitute φ_{jk} with γ_{jk} , and substitute the first assumption with $d_{sj} \ll r_{jk}$ (i.e., the distance between the center of the array and the speaker is much smaller than the distance between the nodes), we arrive to $\gamma_{jk} - \gamma_{kj} \approx \pm\pi$, from where the following generalization is obtained:

$$u_{jk}\alpha_{jk} + \phi_j - u_{kj}\alpha_{kj} - \phi_k \approx \begin{cases} \pm\pi, & \text{if } j \neq k \\ 0, & \text{if } j = k. \end{cases} \quad (11)$$

Figure 3 shows the angular relations between node pairs. Notice that when the distance between the nodes is sufficiently large, the error introduced by the speaker not being located at the array center is negligible.

Defining $\mu_{jk} = u_{jk}\alpha_{jk}$ and taking expression (11) into the complex plane, after exponentiation and some operations, it becomes

$$e^{i(\phi_k - \phi_j)} \approx \begin{cases} e^{i(\mu_{jk} - \mu_{kj} - \pi)}, & \text{if } j \neq k \\ 1, & \text{if } k = j. \end{cases} \quad (12)$$

Now, to estimate the orientations, we can force a relative orientation reference, where $\hat{\phi}_1 = 0$, arriving to the following expression:

$$e^{-i\phi_j} \approx \begin{cases} e^{i(\mu_{j1} - \mu_{1j} - \pi)}, & \text{if } j \neq 1 \\ 1, & \text{if } j = 1. \end{cases} \quad (13)$$

Plugging expression (13) into (12), we obtain the final expressions for the orientation estimation:

$$e^{i\phi_k} \simeq \begin{cases} e^{i(\mu_{jk}-\mu_{kj}-\mu_{j1}+\mu_{1j})}, & \text{if } j \neq k, j \neq 1 \\ e^{i(\mu_{1k}+\mu_{k1}-\pi)}, & \text{if } j = 1, k \neq 1 \\ e^{i(\mu_{1k}+\mu_{k1}-\pi)}, & \text{if } j = k, k \neq 1 \\ 1, & \text{if } k = 1. \end{cases} \quad (14)$$

In order to obtain each value of $\hat{\phi}_k$, we have J estimates, the quality of which is directly related to the error in α_{jk} and u_{jk} , and since u_{jk} is an unknown and also has to be estimated, it is the most unreliable. During the optimization process that will be discussed in the next section, orientation estimate $\hat{\phi}_k$ is obtained by taking the 70% trimmed mean of the J available estimates, thus making the results more robust against outliers created by erroneous u_{jk} values.

With the orientation of the first node fixed at zero, we reestablish a relative coordinate system. The points in this space are translated and rotated; it suffices to know the actual position and orientation of one of the nodes (i.e., having a beacon node) to transform the results to a global coordinate system.

3.3. Uncertainty Solution. At this point, we assume that the values of u_{jk} are known; hence, $\hat{\phi} = [0, \hat{\phi}_2, \dots, \hat{\phi}_J]$, and the estimation of the node locations \mathbf{p} depends on a given DoA uncertainty correction matrix \mathbf{U} . However, its actual value is an unknown, and we must work with the estimate $\hat{\mathbf{U}}$ (composed of $J \times J \hat{u}_{jk}$ values). Because the uncertainty correction is a binary variable, there are 2^{J^2} possible values for $\hat{\mathbf{U}}$, which makes it unfeasible to test every single value. Thus, we decided to use a Genetic Algorithm (GA) to find the solution. It is important to highlight that the main diagonal of $\hat{\mathbf{U}}$ is of no interest (the case when $j = k$) and does not need to be estimated, which reduces the maximum number of combinations to $2^{J(J-1)}$.

We have found a clear relation between the log-likelihood for a certain $\hat{\mathbf{U}}$ and the localization error. Thus, we propose using expression (8) as the fitness function. Figure 4 shows the relation between the log-likelihood and the pairwise node distance error for all possible values of $\hat{\mathbf{U}}$ in a network with $J = 4$. Then, the selected fitness metric clearly has a direct relation with the location error.

To improve the convergency of the optimization algorithm with respect to the total number of performance evaluations, instead of using a single GA and several runs (standard scheme), we use an elimination tournament of small GAs. We start with a set of 64 small GAs (denoted stage of the tournament) with a population of $N_p = 10 * J$ individuals and $N_g = J$ generations each. The best solutions of the first round are then paired, generating a new population for every two winners, which are set to compete in the next round. The process is repeated until a global winner is obtained. For illustrative purposes, Figure 5 shows an example of the elimination tournament used in the experiments with a total of $N_r = 3$ rounds. In our case,

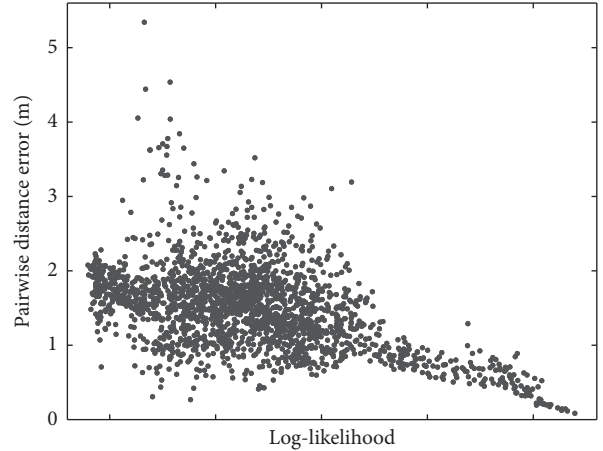


FIGURE 4: Relation between the log-likelihood and the pairwise distance error for all possible values of \mathbf{U} with $J = 4$.

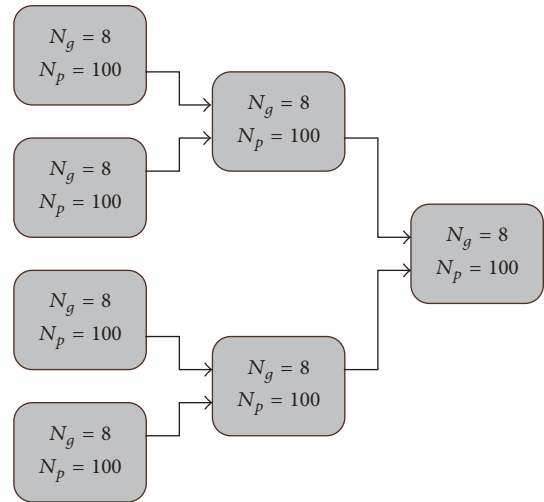


FIGURE 5: Example of elimination tournament with $N_r = 3$ rounds, $N_g = 8$ generations per round, and $N_p = 100$ individuals per stage of the tournament.

we used $N_r = 7$ rounds, since it empirically gave us good convergence results.

The GA algorithm is divided into 7 steps:

- (1) The algorithm is initialized by creating a population of $N_p = 10 * J$ individuals. Each individual ($\hat{\mathbf{U}}_p$) contains $J(J-1)$ genes corresponding to \hat{u}_{jk} . On the first round of the tournament, the genes are randomly selected; for every subsequent round, they are created by reproduction and mutation from the previous stage winners (steps (4) and (5)).
- (2) The population is evaluated. For every $\hat{\mathbf{U}}_p$, node orientations are estimated as described in the previous section, and then the log-likelihood (fitness function) is computed with (8).
- (3) The individuals are sorted according to their fitness level in a descending order. The top performing 10%

is selected to breed a new generation. The remaining 90% of the population is discarded.

- (4) The population is regenerated via the reproduction of successful individuals. For every new individual, two parents are selected at random, each of which randomly provides half of its genes.
- (5) Except for the best performer, the full population is mutated by selecting 1% of their genes at random and inverting their value. Since the probability of a change in u_{jk} involving a change in u_{kj} is very high, 75% of the mutations change the sign of both genes. After mutation takes place, the new generation is complete.
- (6) If the iteration counter is lower than $N_g = J$, the algorithm returns to 2, and the iteration counter is increased; otherwise, it continues to the last step.
- (7) Best \widehat{U}_p is selected as a candidate and is set to compete in the next round of the tournament.

After the GA tournament is completed, the best individual becomes \widehat{U} and is used to estimate the final node positions and orientations.

It is important to highlight that while the computational cost of the optimization algorithm is quite high, the different small GAs can be divided by the total number of nodes of the network, since the parallelization of the elimination tournament is trivial. In a rough approximation, taking the computation time of the closed-form expressions of the ML estimator and the orientation estimator as a single operation, in Big O notation, the parallelized tournament has a complexity $O(J)$. The tournament is composed of 127 GAs divided among J nodes. In the worst scenario, a node has to take care of $N_{ga} = \lceil 127/J \rceil$ GAs. Each GA performs J iterations with a population size of $10 * J$, so, in total, each node needs to compute $N_o = \lceil 127/J \rceil \times J^2$ operations. In average, the computational load of the optimization algorithm (for one node) is around $1300J$ times higher than that of the estimations using the closed-form expressions. Please notice that the need for an iterative algorithm is a direct consequence of the DoA uncertainty. Provided that each node was capable of resolving 360° DoAs (by having 3 or more microphones arranged in a 2D array), the solution to the problem would be found directly.

4. Experiments and Results

To evaluate the proposed algorithm, we generated a realistic database of acoustic signals, which contains 300 different scenarios including both reverberation and background noise. Reverberation was controlled by the absorption coefficient of the walls. Background noise was added as additive white noise controlled by the signal-to-noise ratio (SNR). Each scenario contained 10 randomly located and oriented nodes and was generated with a random combination of the next parameters: room dimensions of 6–12 m long/wide and 2–3 m high, absorption coefficient of 0.5–1, and SNR of 5–20 dB. The positions of the nodes were restricted as if they were on a table of dimensions of 5×2 m (a medium-sized conference table) with a minimum distance between nodes of

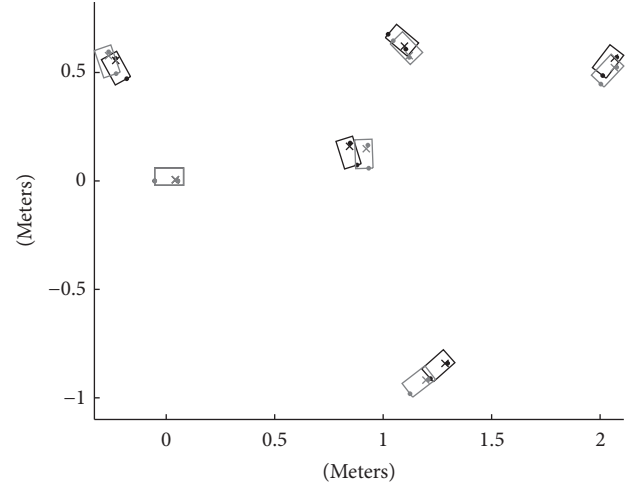


FIGURE 6: Localization results (transposed and rotated) for one example in the database with $J = 6$. The true positions are in black, and the estimates are in grey.

15 cm. The acoustic signals received by the microphones were generated using a room impulse response generator, which was computed using the simple image method described in Allen and Berkley's work [27] at a sampling frequency of 44100 Hz.

The reference acoustic signal emitted by the nodes is a band-limited white noise signal (500 Hz–16 kHz) of length of 4096 samples or 9.29 ms at $f_s = 44100$ Hz. Each device has its unique reference signal, which is known by every node in the network. The selected frequency range is related to the frequency response of typical smartphone speakers, whereas the time duration is a tradeoff between computational complexity and robustness against the SNR. Notice that a short time duration has the added benefit of making the localization process less disturbing to users who are exposed to the reference signals.

Because achieving tight time synchronization between smartphones is not trivial, the synchronicity between nodes was also set at random. All nodes shared an identical sampling frequency f_s , but their clock starting point was biased using a uniform distribution to simulate a loose synchronization between nodes. This clock jitter translates into a range estimation error in meters. For the experiments, the standard deviation of the range estimation was fixed at 3 different values, $\sigma_r = 0.1$ m, $\sigma_r = 0.2$ m, and $\sigma_r = 0.3$ m, depending on the synchronization jitter.

The last consideration is the coordinate system. We have previously mentioned that the origin of coordinates was set at the center of mass of the node locations in the localization process; however, we can assume without loss of generality that the first node is located at the origin of the coordinates. Then, the transposed locations were found by subtracting the coordinates of the first node. Hence, with the condition set for the orientation estimation, the localization results are provided in relation to the first node. With a localization example in Figure 6, we observe that when this reference

TABLE 1: Localization error (centimeters) for different algorithms and network sizes without prior orientation knowledge. DoA estimates with uncertainty have a range of 180° instead of 360° .

σ_r	J	Crocco et al. [22]			Jacob et al. [19]			Proposed			
		Without orientations*			Without uncertainty			With uncertainty			
		Mean	Std.	Trim	Mean	Std.	Trim	Mean	Std.	Trim	
0.1	3	9.6	6.5	8.5	30.3	163.0	11.1	8.3	8.0	7.4	
	4	12.6	8.7	11.1	18.4	46.3	11.8	8.8	11.5	7.4	
	5	12.5	8.7	10.8	15.6	16.8	12.1	8.1	8.1	7.2	
	6	11.3	7.4	9.8	15.0	14.1	12.0	8.0	9.3	6.8	
	7	10.1	6.7	8.9	14.9	15.4	11.8	8.9	11.7	6.9	
	8	9.1	5.6	8.1	15.7	20.8	11.6	9.3	14.3	6.7	
	9	8.7	5.5	7.8	16.4	22.0	11.6	10.0	15.7	6.7	
	10	8.1	4.1	7.3	16.1	22.5	11.2	10.9	16.6	6.9	
	0.2	3	17.5	10.6	15.9	31.5	145.2	12.9	11.6	8.8	10.5
		4	21.7	12.2	19.7	19.8	46.3	13.2	11.7	11.1	10.4
5		23.2	13.0	20.8	16.4	16.8	12.8	10.9	7.9	10.0	
6		22.2	12.1	20.1	15.5	14.2	12.8	10.7	9.4	9.5	
7		20.9	11.9	18.7	15.6	15.5	12.5	11.4	11.6	9.4	
8		20.8	12.5	18.2	16.3	20.4	12.0	11.4	14.2	8.9	
9		19.3	11.5	17.1	17.0	20.9	12.2	11.7	15.6	8.9	
10		18.3	11.3	16.0	16.8	22.5	11.7	12.9	16.8	8.9	
0.3		3	22.9	13.2	21.1	34.7	167.8	15.8	15.4	9.9	14.0
		4	28.7	15.0	26.5	21.8	52.9	14.5	14.8	9.8	13.6
	5	30.9	15.4	28.9	18.1	17.4	14.1	14.0	8.4	13.0	
	6	31.6	16.1	29.1	16.7	15.4	13.5	13.6	8.8	12.3	
	7	32.5	17.4	29.5	17.1	17.4	13.2	14.1	11.1	12.2	
	8	32.7	17.3	30.0	17.1	19.6	12.7	14.0	13.8	11.6	
	9	29.6	14.6	27.1	17.6	20.1	12.8	13.8	14.8	11.1	
	10	30.6	17.2	27.4	17.5	21.4	12.5	15.0	15.8	11.3	

*Node orientations are not considered.

system is used, the estimated and true locations of the first node are identical.

In order to set a comparison with the proposed method, we have implemented 2 of the methods available in the literature, namely, Jacob et al. [19] and Crocco et al. [22].

The method presented in Jacob et al.'s work [19] is based on angle measurements alone. In order to adapt it for the use of range measurements, the solution is scaled to minimize the difference with the measured range values as described in Schmalenstroeyer et al.'s work [28]. It is important to highlight that this method only works without DOA uncertainty (3 or more microphones per node) and so, in order to obtain the results, we assumed that the nodes were capable of measuring 360° DoAs using only 2 microphones, which is physically impossible.

The method described in Crocco et al.'s work [22] only uses range measurements, since it is intended for nodes with a single microphone. This method is not capable of discerning between reflected solutions and so, in order to obtain the results, we considered all the possible reflections. Notice that we obtain the range estimates by averaging the ToAs at both microphones; thus this method is not capable of obtaining orientation estimations. In case the ToAs were obtained at each microphone, it should be possible to also estimate the

orientations by adding some constraints (known distance between same node microphones), although in [22] this is not considered.

4.1. Result Discussion. Table 1 shows the mean, the standard deviation (Std.), and the 25% trimmed mean (trim) of the localization error obtained with the proposed algorithm and those obtained with Jacob et al. [19] and Crocco et al. [22], all of them working without previous knowledge about the node orientations. Please notice that Crocco et al. [22] do not consider node orientations and that Jacob et al. [19] use DoA estimates covering 360° , while the presented method is based on 180° DoA estimates. Of these methods, the proposed method obtains the best overall results except for $\sigma_r = 0.1$, where Crocco et al.'s method [22] is better for large network sizes ($J \geq 8$) due to convergence problems on the DoA uncertainty estimation. This effect can be noticed by looking at the trim and Std. for the proposed method. It is possible to see that while the trimmed mean follows a descending trend when the network size is enlarged, the Std. grows larger.

Crocco et al.'s [22] performance is affected by the range estimate error derived from the synchronization lag. The sensibility of this method to range estimation errors clearly shows that when comparing the results obtained with

TABLE 2: Localization error (centimeters) for ML-DONL algorithm for different network sizes and orientation error. Node orientations are obtained with an electronic compass.

σ_r	J	ML-DONL ($\sigma_\phi = 0^\circ$)			ML-DONL ($\sigma_\phi = 15^\circ$)			
		Mean	Std.	Trim	Mean	Std.	Trim	
0.1	3	7.0	6.9	6.4	24.5	18.8	21.1	
	4	7.5	10.5	6.4	29.3	21.4	25.6	
	5	7.1	8.4	6.2	33.0	22.4	29.1	
	6	6.8	6.6	6.0	34.0	22.1	30.3	
	7	7.0	7.4	5.9	35.8	23.3	31.7	
	8	7.5	10.1	5.8	37.3	23.3	33.3	
	9	7.5	11.5	5.6	38.1	23.9	34.0	
	10	7.5	11.3	5.5	38.7	24.3	34.5	
	0.2	3	10.4	8.1	9.4	25.9	17.6	22.8
		4	10.5	10.5	9.5	30.2	20.5	26.7
5		10.1	8.4	9.1	34.0	22.1	30.0	
6		9.6	6.6	8.7	34.9	21.7	31.2	
7		9.6	7.2	8.6	36.5	22.8	32.4	
8		9.8	10.0	8.1	37.8	23.1	33.9	
9		9.9	11.2	8.0	38.7	23.7	34.6	
10		9.5	11.2	7.7	39.1	23.9	34.9	
0.3		3	14.1	8.5	13.0	28.3	17.3	25.5
		4	14.1	11.5	12.8	32.3	20.4	28.9
	5	13.3	8.9	12.3	35.8	21.5	32.1	
	6	12.6	6.8	11.7	36.5	21.6	32.9	
	7	12.5	7.3	11.4	37.8	22.7	33.7	
	8	12.5	10.1	10.8	38.9	22.9	35.0	
	9	12.2	11.2	10.4	39.7	23.3	35.6	
	10	11.9	11.1	10.1	40.1	23.6	36.1	

increasing σ_r values, Jacob et al.'s method [19] is less affected by the range error, since the geometry is found using the DoAs and the range is only employed to scale the solution. It is worthwhile to highlight again that Jacob et al.'s method [19] is not capable of solving the DoA uncertainty problem. It is not possible to use this method with only 2 microphones per node.

Table 2 shows the mean, Std., and trim of the localization error obtained with the ML-DONL algorithm presented in Ayllon et al.'s work [24] with known orientations (assumed to be obtained with an electronic compass), with and without orientation measurement error. Comparing this table with the previous one, we observe that the error obtained with the presented method using orientation estimates is between those obtained using measured orientations: it is larger than that of the ideal case but much smaller than when the typical error of an uncalibrated compass is introduced.

Table 3 shows the mean, Std., and trim of the orientation estimation error obtained with the proposed algorithm and those obtained by Jacob et al. [19]. The proposed method gets larger errors compared to Jacob et al.'s method [19], although it is worth recalling that the latter does not have to deal with DoA uncertainty. From the table, we can observe that orientation estimation is independent from σ_r , since

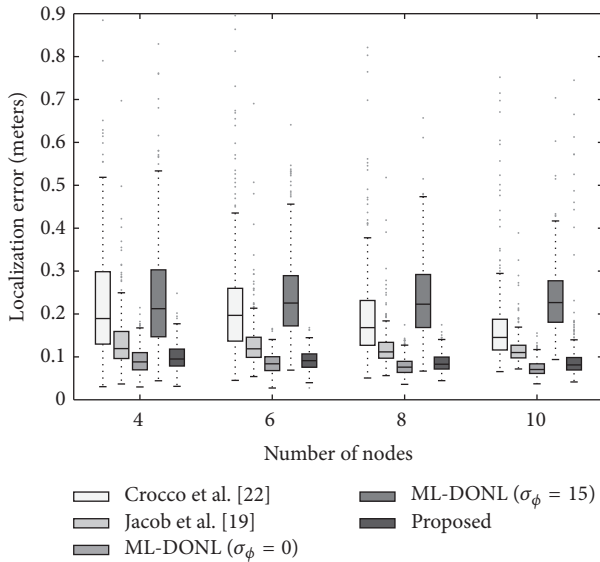
it only uses DoA measurements. With both methods, the orientation estimation error is lower than that of a typical digital compass, rendering them useless for this particular application. Figure 7 shows a box plot of the results obtained for all the tested algorithms with $\sigma_r = 0.2$, from which it is easier to see how the different algorithms perform.

A deeper analysis of the ML location estimation has revealed that large localization errors are associated with large DoA estimation errors, that is, those instances when the largest peak of the correlation corresponds to a reflection instead of the direct signal. Some proposals in the literature use outlier detection techniques to reduce the effect of spurious measurements. Jacob et al. [19] used random sample consensus (RANSAC) for the minimization algorithm (not implemented in our version); in Plinge and Fink's work [20], outliers were detected by applying a threshold to the estimation error. Our current implementation does not contemplate outlier detection; therefore, the obtained errors have large variances.

Regarding the number of nodes, localization accuracy usually increases with larger networks. This result is expected because there is more information available; thus, it is easier to compensate for large local estimation errors (either DoA or range) in one or several nodes. However, due to DoA

TABLE 3: Orientation estimation error (degrees) for different algorithms and network sizes. DoA estimates with uncertainty have a range of 180° instead of 360° .

σ_r	J	Jacob et al. [19]			Proposed		
		Mean	Std.	Trim	Mean	Std.	Trim
0.1	3	2.9°	9.2°	1.4°	4.6°	11.2°	2.0°
	4	2.0°	5.4°	1.3°	2.6°	7.0°	1.7°
	5	1.8°	3.6°	1.3°	2.2°	6.2°	1.5°
	6	1.8°	3.8°	1.2°	2.2°	6.3°	1.3°
	7	2.1°	5.3°	1.2°	2.9°	8.6°	1.3°
	8	2.4°	6.8°	1.2°	3.0°	8.6°	1.2°
	9	2.5°	6.8°	1.1°	4.2°	12.3°	1.2°
	10	2.3°	6.3°	1.1°	6.0°	13.4°	1.6°
0.2	3	3.3°	11.3°	1.4°	4.9°	11.5°	2.1°
	4	2.1°	5.3°	1.4°	2.7°	6.9°	1.7°
	5	2.2°	7.8°	1.3°	2.1°	5.3°	1.5°
	6	1.8°	3.9°	1.2°	2.2°	6.2°	1.3°
	7	2.1°	5.3°	1.2°	2.9°	9.0°	1.3°
	8	2.4°	6.4°	1.2°	2.6°	7.7°	1.2°
	9	2.3°	6.6°	1.1°	3.7°	10.6°	1.2°
	10	2.3°	6.3°	1.1°	6.2°	14.0°	1.5°
0.3	3	2.9°	10.2°	1.4°	4.6°	11.0°	2.1°
	4	2.1°	6.1°	1.4°	2.8°	6.7°	1.7°
	5	1.8°	3.6°	1.3°	2.3°	5.8°	1.5°
	6	1.8°	4.1°	1.2°	2.3°	6.6°	1.4°
	7	2.3°	6.4°	1.2°	2.8°	8.1°	1.4°
	8	2.3°	6.2°	1.2°	2.7°	7.9°	1.3°
	9	2.2°	5.8°	1.1°	3.6°	10.3°	1.2°
	10	2.3°	6.4°	1.1°	5.6°	12.6°	1.5°

FIGURE 7: Box plot of the localization results for the tested algorithms with different number of nodes and $\sigma_r = 0.2$.

uncertainty, the proposed method has some convergence problems with large networks that need to be addressed.

To the best of our knowledge, our proposal is (together with our previous work in Ayllon et al. [24]) the only method capable of 2D DoA-based distributed array configuration calibration using nodes equipped with only 2 microphones.

5. Conclusions

In this paper, we have presented a new self-localization algorithm for wireless smartphone networks composed of commercial off-the-shelf devices that are equipped with two microphones and a speaker. The entire localization process is based on DoA and range estimates between node pairs obtained with acoustic signals. The main novelty of this work is a modification of the previously presented ML-DONL algorithm, which enables us to locate the nodes even without prior knowledge about their orientation. Thus, we eliminate the requirement for an electronic compass. The nodes are located by finding the position of their speaker and estimating their orientation while solving the DoA uncertainty problem, which arises from the use of only 2 microphones per node. The obtained localization error is lower than that obtained when an uncalibrated electronic compass is used, which is the most common scenario for off-the-shelf smartphones. In summary, the proposed algorithm improves the localization accuracy of other methods that

require reference nodes or additional sensors, and it is on the same scale as other DoA-based algorithms without requiring ad hoc hardware. In addition, the computational cost of the algorithm is assumable for current mobile processors. However, the solution of the DoA uncertainty with a GA tournament adds a significative computational load, making it worthy to explore more efficient solutions. Future work will address spurious measurements using outlier detection techniques and will study different approaches to the DoA uncertainty estimation because they are the main sources of error.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work has been funded by the Spanish Ministry of Economy and Competitiveness/FEDER under Project TEC2015-67387-C4-4-R.

References

- [1] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero III, R. L. Moses, and N. S. Correal, "Locating the nodes: cooperative localization in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 54–69, 2005.
- [2] C. Savarese, J. M. Rabaey, and J. Beutel, "Location in distributed ad-hoc wireless sensor networks," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '01)*, vol. 4, pp. 2037–2040, Salt Lake City, Utah, USA, May 2001.
- [3] F. Mondinelli and Z. M. Kovacs-Vajna, "Self-localizing sensor network architectures," *IEEE Transactions on Instrumentation and Measurement*, vol. 53, no. 2, pp. 277–283, 2004.
- [4] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networks Part I: sequential node updating," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5277–5291, 2010.
- [5] E. Robledo-Arnuncio, T. S. Wada, and B.-H. Juang, "On dealing with sampling rate mismatches in blind source separation and acoustic echo cancellation," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 34–37, October 2007.
- [6] P. Aarabi, "The fusion of distributed microphone arrays for sound localization," *Eurasip Journal on Applied Signal Processing*, vol. 2003, no. 4, pp. 338–347, 2003.
- [7] A. Hassani, A. Bertrand, and M. Moonen, "Cooperative integrated noise reduction and node-specific direction-of-arrival estimation in a fully connected wireless acoustic sensor network," *Signal Processing*, vol. 107, pp. 68–81, 2015.
- [8] V. Berisha, H. Kwon, and A. Spanias, "Real-time implementation of a distributed voice activity detector," in *Proceedings of the 4th Workshop on Sensor Array and Multichannel Processing*, pp. 659–662, 2006.
- [9] A. Bertrand and M. Moonen, "Robust distributed noise reduction in hearing aids with external acoustic sensor nodes," *Eurasip Journal on Advances in Signal Processing*, vol. 2009, Article ID 530435, 14 pages, 2009.
- [10] P. Bahl and V. N. Padmanabhan, "Radar: an in-building rf-based user location and tracking system," in *Proceedings of the 9th Annual Joint Conference of the IEEE Computer and Communications Societies, IEEE INFOCOM '00*, vol. 2, pp. 775–784, March 2000.
- [11] P. Pivato, L. Palopoli, and D. Petri, "Accuracy of RSS-based centroid localization algorithms in an indoor environment," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 10, pp. 3451–3460, 2011.
- [12] G. Borriello, A. Liu, T. Offer, C. Palistrant, and R. Sharp, "Walrus: wireless acoustic location with room-level resolution using ultrasound," in *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services (MobiSys '05)*, pp. 191–203, ACM, June 2005.
- [13] S. J. Kim and B. K. Kim, "Accurate hybrid global self-localization algorithm for indoor mobile robots with two-dimensional isotropic ultrasonic receivers," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 10, pp. 3391–3404, 2011.
- [14] S. Ladstätter, P. Luley, A. Almer, and L. Paletta, "Multisensor data fusion for high accuracy positioning on mobile phones," in *Proceedings of the 12th International Conference on Human-Computer Interaction with Mobile Devices and Services, Mobile HCI '10*, pp. 395–396, September 2010.
- [15] M. Klepal, M. Weyn, W. Najib et al., "Ols: opportunistic localization system for smart phones devices," in *Proceedings of the 1st ACM Workshop on Networking, Systems, and Applications for Mobile Handhelds*, pp. 79–80, August 2009.
- [16] F. Höflinger, J. Wendeberg, R. Zhang et al., "Acoustic self-calibrating system for indoor smartphone tracking (ASSIST)," in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '12)*, pp. 1–9, November 2012.
- [17] Z. Li, X. Li, and Y. Wang, "A calibration method for magnetic sensors and accelerometer in tilt-compensated digital compass," in *Proceedings of 9th International Conference on the Electronic Measurement and Instruments, ICEMI '09*, pp. 862–868, 2009.
- [18] A. Plinge, F. Jacob, R. Haeb-Umbach, and G. A. Fink, "Acoustic microphone geometry calibration: an overview and experimental evaluation of state-of-the-art algorithms," *IEEE Signal Processing Magazine*, vol. 33, no. 4, pp. 14–29, 2016.
- [19] F. Jacob, J. Schmalenstroer, and R. Haeb-Umbach, "Microphone array position self-calibration from reverberant speech input," in *Proceedings of the International Workshop on Acoustic Signal Enhancement, IWAENC '12*, pp. 1–4, September 2012.
- [20] A. Plinge and G. A. Fink, "Geometry calibration of multiple microphone arrays in highly reverberant environments," in *Proceedings of the 2014 14th International Workshop on Acoustic Signal Enhancement, IWAENC '14*, pp. 243–247, September 2014.
- [21] M. A. Anwar, H. Hassan, H. Maqbool, A. Rehman, and M. Tahir, "Acoustic sensor network relative self-calibration using joint TDOA and DOA with unknown beacon positions," in *Proceedings of the 2014 IEEE Wireless Communications and Networking Conference, WCNC '14*, pp. 3064–3069, April 2014.
- [22] M. Crocco, A. Del Bue, M. Bustreo, and V. Murino, "A closed form solution to the microphone position self-calibration problem," in *Proceedings of the 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '12*, pp. 2597–2600, March 2012.
- [23] S. Thrun, "Affine structure from sound," in *Proceedings of the Eletronic Proceedings of Neural Information Processing Systems*, pp. 1353–1360, 2005.

- [24] D. Ayllon, H. A. Sanchez-Hevia, R. Gil-Pita, M. U. Manso, and M. R. Zurera, "Indoor blind localization of smartphones by means of sensor data fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 4, pp. 783–794, 2016.
- [25] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays*, pp. 157–180, Springer, Berlin, Germany, 2001.
- [26] S. Sur, T. Wei, and X. Zhang, "Autodirective audio capturing through a synchronized smartphone array," in *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys '14*, pp. 28–41, June 2014.
- [27] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.
- [28] J. Schmalenstroerer, F. Jacob, R. Haeb-Umbach, M. H. Hennecke, and G. A. Fink, "Unsupervised geometry calibration of acoustic sensor networks using source correspondences," in *Proceedings of the 12th Annual Conference of the International Speech Communication Association, Interspeech '11*, pp. 597–600, August 2011.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

