View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Infoscience - École polytechnique fédérale de Lausanne

SOURCE LOCALIZATION AND TRACKING IN NON-CONVEX ROOMS

Orhan Öçal, Ivan Dokmanić and Martin Vetterli

School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
{orhan.ocal, ivan.dokmanic, martin.vetterli}@epfl.ch

ABSTRACT

We consider the estimation of the acoustic source position in a known room from recordings by a microphone array. We propose an algorithm that does not require the room to be convex, nor a line-of-sight path between the microphone array and the source to be present. Times of arrival of early echoes are exploited through the image source model, thereby transforming the indoor localization problem to a problem of localizing multiple sources in the free-field. The localized virtual sources are mirrored into the room using the image source method in the reverse direction. Further, we propose an optimization-based algorithm for improving the estimate of the source position. The algorithm minimizes a cost function derived from the geometry of the localization problem. We apply the designed optimization algorithm to track a moving source, and show through numerical simulations that it improves the tracking accuracy when compared with the naïve approach.

***Index Terms*—** Room impulse response, image source model, indoor localization, tracking, non-convex

1. INTRODUCTION

Outdoor localization is almost "solved" by GPS and the related services. On the contrary, indoor localization is still a challenge, despite numerous attractive applications [1]. For example, automatized inventory management and object tracking rely on indoor localization. A group of applications that recently received substantial attention is the location-aware services. Location-customized information could be valuable for users in administrative buildings, museums or shopping malls. Tracking customers is interesting for management or planning in shopping malls. Another group of important applications of indoor localization is security and rescue operations, for example, tracking the location of firefighters in a burning building [2].

Positioning systems, both indoor and outdoor, can be divided into three main topologies [3]. First, the *self-positioning* system, where the receiver makes measurements from distributed transmitters to determine its own position (e.g., GPS). Second, *remote positioning*, where receivers located at possibly multiple locations measure the signal from an object to find its location. Third, *indirect positioning* where a data link is used to transfer position information from a self-positioning system to a remote site or vice versa.

We propose a novel algorithm for localization in a known room that fits in the *remote positioning* group. The algorithm relies on measuring the times-of-flight from a source to a set of receivers. The signal is arbitrary, but typical examples are ultrasound or ultra-wideband signals (UWB). Room has been used previously in the literature to improve source localization [4]. Differently from earlier approaches, our solution is not limited to convex rooms, nor do we require that the microphones see the source. The main tool to achieve this is *image source mirroring*—a sequence of reflections of the localized virtual sources. We demonstrate through simulations that the proposed algorithm successfully localizes the source in a number of non-convex rooms. Furthermore, we propose an optimization-based technique for refining the position estimate. The refining technique is used as a building block of a tracking algorithm. We show how the optimization-based refinement improves the tracking performance, even when we completely skip the multilateration and mirroring steps. For simplicity, in this proof-of-concept paper, we study the 2D case, but because the image source model holds in all dimensions, extension to 3D is immediate. Nevertheless, we note that there exist "almost-2D" devices [5].

Our algorithm relies on the knowledge of the reflector positions in a room. The room shape can be measured independently and fed as an input into our model, but we can also find the room shape using times of flight, acoustically or through UWB [6, 7, 8, 9].

2. MODELING

We consider the room to be a K-sided polygon given by a $2 \times K$ vertex matrix $\mathbf{P} = [\mathbf{p}_1, \cdots, \mathbf{p}_K]$. Without loss of generality, we choose the vertex $\mathbf{p}_1$ to coincide with the origin, and assume the vertices to be specified in counter-clockwise direction. We define the ith wall of the room as the line segment joining vertices $\mathbf{p}_i$ and $\mathbf{p}_{i+1}$. Each side of the room is associated with a unit outward normal $\mathbf{n}_i$, and we denote the source position by $\mathbf{s}$.

The acoustics of a room can be described by a family of room impulse responses (RIR) which model the channel between a fixed source and a fixed microphone. For the mth microphone the RIR is given by

$$h_m(t) = \sum_i a_{m,i}\delta(t - \tau_{m,i}), \tag{1}$$

where $\delta(t)$ denotes the Dirac delta function. Using the RIR, we can find the signal received by the mth microphone as a convolution between the emitted signal and the RIR, $y_m(t) = (h_m * x)(t) = \int x(s)h_m(t-s)\,ds$. By measuring the RIR between the source and the receiver, one can access the propagation times $\tau_{m,i}$, which can be linked to the source and microphone positions in a known room geometry via the *image source model* [10, 11].

2.1. Image Source Model

The image source model states that reflections can be viewed as direct signals coming from *virtual sources*. The positions of these

This work was supported by an ERC Advanced Grant – Support for Frontier Research – SPARSAM Nr: 247006.

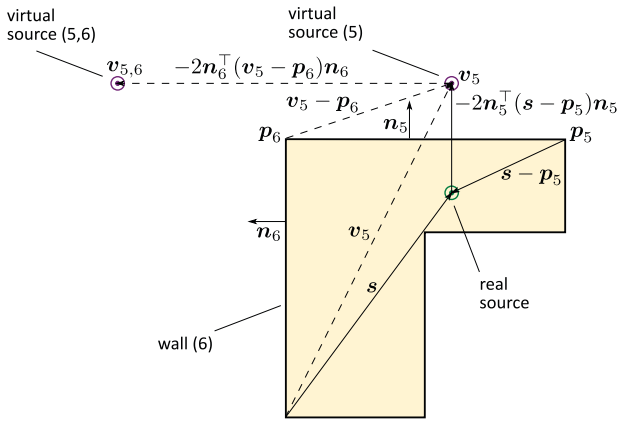

Fig. 1. Finding the locations of virtual sources.

virtual sources are found by mirroring the source across reflective walls as [11]

$$\begin{aligned}\mathbf{v}_i &= \mathbf{s} - 2\mathbf{n}_i^\top(\mathbf{s}-\mathbf{p}_i)\mathbf{n}_i \\ &= \mathbf{s} - 2\mathbf{N}_i(\mathbf{s}-\mathbf{p}_i),\end{aligned} \tag{2}$$

where $\mathbf{N}_i \stackrel{\text{def}}{=} \mathbf{n}_i\mathbf{n}_i^\top$ is the orthogonal projection matrix onto the normal $\mathbf{n}_i$. To find the higher order virtual sources, one can reflect the source across multiple walls, or equivalently, reflect a virtual source across another wall. Thus, for second order reflections we have

$$\mathbf{v}_{i,j} = \mathbf{v}_i - 2\mathbf{N}_j(\mathbf{v}_i - \mathbf{p}_j), \tag{3}$$

and the relation follows the same form for higher order echoes.

If the reflection corresponding to the virtual source is received by the mth microphone positioned at $\mathbf{m}_m$, the quantity $\|\mathbf{v}_{(\cdot)}-\mathbf{m}_m\|/c$, where c is the speed of sound, corresponds to one of the $\tau_{m,i}$ in (1).

3. SOURCE LOCALIZATION

The essential idea behind the algorithm is that we treat all sources as equal whether they be real or virtual, and transform indoor localization to a problem of localizing multiple sources in the free-field. We can localize the source from its distance to three or more microphones at fixed positions. This method is called *multilateration* [12], and the source is localized by intersecting circles with radii equal to the measured distances, as illustrated in Fig. 2. Note that increasing the number of microphones improves the localization.

When the times of arrival (TOA) of echoes are obtained without measurement errors, the circles intersect at a single point that corresponds to the source position. However, if there is jitter in the measurements, the circles will generally not intersect at a single point. In such a case, one can estimate the source position by solving an optimization problem.

Given the distance measurements between the source and the M microphones,

$$r_i = \|\mathbf{s}-\mathbf{m}_i\| + \varepsilon_i, \quad i = 1, 2, \cdots, M,$$

where ε_i is random measurement jitter, and M is the number of microphones, source position can be estimated as the minimizer of

$$\underset{\mathbf{x}\in\mathbb{R}^2}{\text{minimize}} \quad \sum_{i=1}^{M} \left(\|\mathbf{x}-\mathbf{m}_i\| - r_i\right)^2. \tag{4}$$

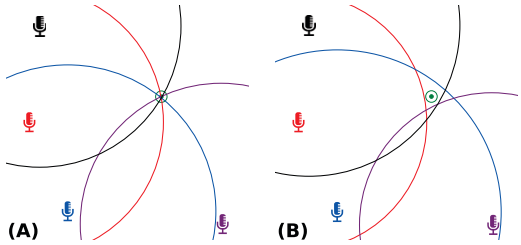

Fig. 2. Multilateration as intersection of circles. (A) Without jitter. (B) With jitter.

If ε_i are i.i.d. centered Gaussian with equal variance, the minimizer of (4) is the maximum likelihood estimator [13]. However, this problem is not convex, and there is no efficient algorithm for finding the globally optimal solution.

Another optimization problem is the 'squared-range-based least squares' [14] obtained by squaring the distances in (4),

$$\underset{\mathbf{x}\in\mathbb{R}^2}{\text{minimize}} \quad \sum_{i=1}^{M} \left(\|\mathbf{x}-\mathbf{m}_i\|^2 - r_i^2\right)^2. \tag{5}$$

Although this too is a non-convex problem, the globally optimal solution can be found efficiently [14, 15].

A challenge when performing multilateration from reverberant recordings is to group the echoes that correspond to a single virtual source. The RIR contains the distances in $\tau_{m,i}$, but echoes coming from different walls can be heard in different orders by the microphones. To solve this problem, we select one echo from each microphone and solve (5) to get a position estimate. If the chosen echo combination does not belong to a single virtual source, there is no point in the plane that yields the chosen distances to the microphones, hence the distances between the estimated position and the microphones will be different than the selected measured distances. We enforce this by evaluating the objective (4) at the optimal solution of (5), and declaring the combination as wrong if the obtained value is above a prescribed threshold. Although this method requires a combinatorial search over the recorded echoes, the number of combinations is small enough to make the method computationally feasible [9].

3.1. Reflecting Localized Sources

After finding the location of the virtual source, we can use the knowledge of the room geometry to reflect it back into the room following the method of images in reverse order. After a sequence of reflections, we find the position of the source that generated the localized virtual source.

We explain the reflecting procedure with reference to Fig. 3. A line is drawn between the virtual source and each of the microphones. Then, we reflect the virtual source across the wall that intersects the drawn lines, and we store the intersection points. If the reflected source is inside the room, we are done. Otherwise, we draw a new set of lines between the previous set of intersection points and the new virtual source, and we reflect the virtual source across the wall that intersects the newly drawn lines. The algorithm is iterated until the reflected source is finally inside the room. In Fig. 3, the procedure is illustrated for a second order virtual source.

A problem that may occur while applying the inverse method of images is that the lines connecting the virtual source with the microphones intersect more than one wall because of the errors in virtual

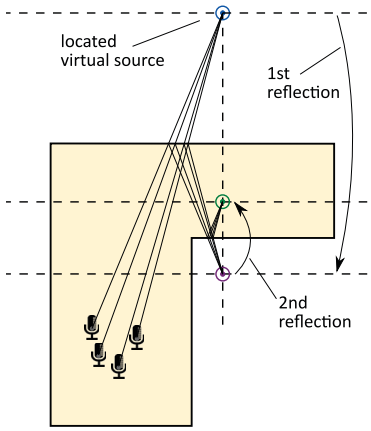

Fig. 3. Reflecting a localized second order virtual source into the room using method of images in reverse order.

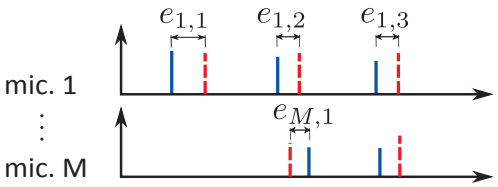

Fig. 4. Notation of G_{RIR}. Solid blue: recorded RIR, $r_{(i,j)}$; dashed red: simulated RIR, $\hat{r}_{(i,j)}$, where i is the microphone index and j is the order of the echo in ith microphone.

source estimation. In such case, we may either discard the problematic virtual source, or we can reflect across the wall with the highest number of intersections.

3.2. Estimating Source Position

Thus far, we have localized multiple virtual sources, and reflected them inside the room. There are different ways to combine multiple reflected sources into a single position estimate. For example, we could use the localization score, G_{LOC}, defined as the optimal value of (4), and choose the reflected virtual source with the smallest G_{LOC}. However, as the measurement jitter increases, it may happen that wrong echo combinations, not corresponding to a real virtual source, yield a better score than the correct echo combinations. Hence, we propose a different scoring, suitable for robust localization with strong measurement jitter.

If the position estimate is close to the source, the simulated RIR from the estimated position is *close* to the recorded RIR. To use this idea to estimate the source position, we define a metric that measures the distance between the two RIRs. For every echo recorded by the microphones, we find the echo closest in time in the simulated RIR, and compute the 2-norm of the time differences. More precisely, we define the cost function between the RIRs as

$$G_{\text{RIR}}(\hat{\mathbf{s}}) = \sum_{i=1}^{M} \sum_{j=1}^{n_i} e_{i,j}^2(\hat{\mathbf{s}}), \tag{6}$$

where $e_{i,j}(\hat{\mathbf{s}}) = \min_k |r_{(i,j)} - \hat{r}_{(i,k)}|$, $r_{(i,j)}$ is the jth echo recorded by the ith microphone, $\hat{r}_{(i,k)}$ is the kth pulse that would have been recorded by the ith microphone if the source was at $\hat{\mathbf{s}}$, and n_i is the number of echoes heard by ith microphone. We select the reflected source $\hat{\mathbf{s}}$ that gives the least G_{RIR} as the estimated source location.

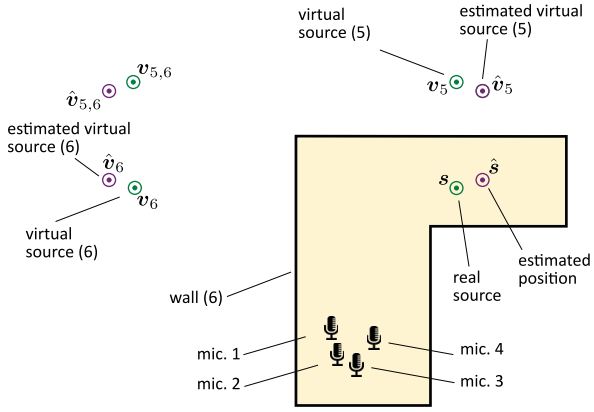

Fig. 5. The notation of location optimization.

3.3. Optimizing the Position Estimate

Observe that we restricted the search space for the minimizer of G_{RIR} to the discrete set of reflected virtual sources. It is possible to further improve the localization by perturbing the estimate, so that G_{RIR} is minimized. Echoes in the simulated RIR are produced by image sources of the source estimate $\hat{\mathbf{s}}$. We perturb $\hat{\mathbf{s}}$ so that the TOA of the echo from each simulated virtual source approaches the TOA of the nearest recorded echo. Virtual source that gives the echo closest to the jth echo recorded by the ith microphone is given as

$$\mathbf{v}_{(i,j)}(\hat{\mathbf{s}}) \stackrel{\text{def}}{=} \operatorname*{argmin}_{\mathbf{v} \in \mathcal{V}(\hat{\mathbf{s}})} |r_{(i,j)} - \|\mathbf{v} - \mathbf{m}_i\||,$$

where by $\mathcal{V}(\hat{\mathbf{s}})$ we denote the set of virtual sources of $\hat{\mathbf{s}}$ of arbitrary order. We can rewrite $e_{i,j}(\hat{\mathbf{s}})$ in (6) as a function of the virtual sources generated by the estimated source position $\hat{\mathbf{s}}$ as $e_{i,j}(\hat{\mathbf{s}}) = |r_{(i,j)} - \|\mathbf{v}_{(i,j)}(\hat{\mathbf{s}}) - \mathbf{m}_i\||$. The source location can be estimated by minimizing G_{RIR} inside the room,

$$\underset{\hat{\mathbf{s}}}{\text{minimize}} \sum_{i=1}^{M} \sum_{j=1}^{n_i} \left(r_{(i,j)} - \|\mathbf{v}_{(i,j)}(\hat{\mathbf{s}}) - \mathbf{m}_i\|\right)^2. \tag{7}$$

Although this again is a non-convex problem, we can find the correct minimum if the initial position estimate is close enough. The local minimum can be computed by iterative optimization methods such as the gradient descent algorithm, which was found to perform successfully. We calculate the gradient as

$$\nabla G_{\text{RIR}}(\hat{\mathbf{s}}) = \sum_{i=1}^{M} \sum_{j=1}^{n_i} 2 \left[\prod_{w \in \text{walls}(\mathbf{v}_{(i,j)})} \left(\mathbf{I}_2 - 2\mathbf{n}_w \mathbf{n}_w^\top\right) \right] \cdot \left(\|\mathbf{v}_{(i,j)}(\hat{\mathbf{s}}) - \mathbf{m}_i\| - r_{i,j}\right) \frac{\mathbf{v}_{(i,j)}(\hat{\mathbf{s}}) - \mathbf{m}_i}{\|\mathbf{v}_{(i,j)}(\hat{\mathbf{s}}) - \mathbf{m}_i\|},$$

where walls $\left(\mathbf{v}_{(i,j)}\right)$ is the sequence of walls that generate $\mathbf{v}_{(i,j)}$.

Minimizing G_{RIR} is motivated by the following proposition.

Proposition 1. *Given the distance measurements between the virtual sources of* $\mathbf{s}$ *and* M *microphones*

$$\rho_{i,\mathbf{v}} = \|\mathbf{v} - \mathbf{m}_i\| + \varepsilon_{i,\mathbf{v}}, \tag{8}$$

where $i = 1, \cdots, M$, $\mathbf{v} \in \mathcal{V}(\mathbf{s})$, *and* $\varepsilon_{i,\mathbf{v}}$ *is i.i.d. measurement jitter*

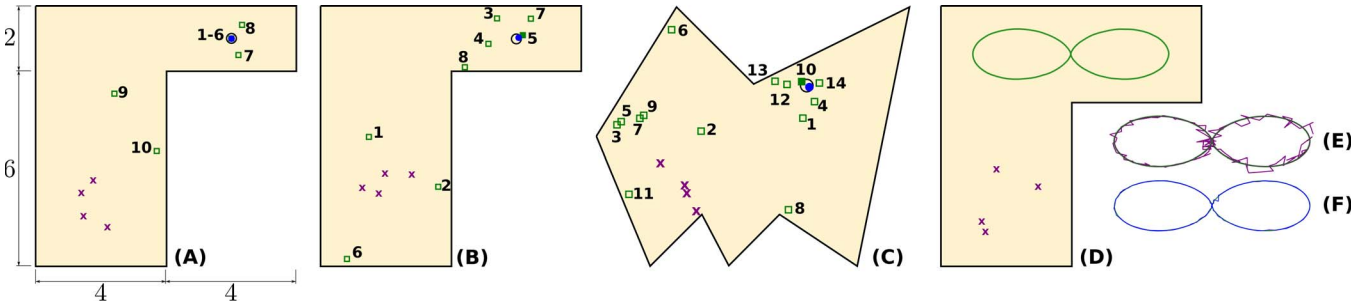

Fig. 6. Localization in L-shaped room (A) without jitter and (B) with jitter having $\sigma = 0.05$. (C) Localization in complex geometry with jitter having $\sigma = 0.1$. (D) The source path and the microphone positions in tracking simulation. (E) Tracking with all the steps of localization algorithm without performing optimization. (F) Tracking by only doing optimization based on the previous estimate.

from $\mathcal{N}\left(0, \sigma^2\right)$*, the solution to*

$$\underset{\hat{\mathbf{s}}}{\text{minimize}} \quad \sum_{i=1}^{M} \sum_{\mathbf{v} \in \mathcal{V}(\mathbf{s})} \left(\rho_{i,\mathbf{v}} - \|\hat{\mathbf{v}}_{\text{walls}(\mathbf{v})} - \mathbf{m}_i\|\right)^2, \qquad (9)$$

where $\hat{\mathbf{v}}_{\text{walls}(\mathbf{v})}$ *is the virtual source obtained by reflecting* $\hat{\mathbf{s}}$ *across the sequence of walls that generate* $\mathbf{v}$*, yields the maximum likelihood estimator of the source location.*

Proof. The likelihood function for obtaining the measured distances is $p\left(\rho \mid \hat{\mathbf{s}}\right) = \prod_{i=1}^{M} \prod_{\mathbf{v}} p\left(\rho_{i,\mathbf{v}} \mid \hat{\mathbf{v}}_{\text{walls}(\mathbf{v})}\right)$. The result is obtained by substituting (8) in the likelihood function and minimizing the negative log-likelihood. □

However, because the echoes are not labeled in the recordings, we do not have access to the distances $\rho_{i,\mathbf{v}}$ between virtual sources and the microphones. Hence, we cannot calculate $\rho_{i,\mathbf{v}} - \|\hat{\mathbf{v}}_{\text{walls}(\mathbf{v})} - \mathbf{m}_i\|$. Solving (7) can be viewed as a heuristic for solving (9).

3.4. Tracking

Source tracking can be performed by repeated localization; the source can be localized independently at each time instant using the described algorithm. However, because the current position of the source depends on previous locations, we can leverage previous estimates to improve the performance.

We propose the following method: For the initial position there are no prior estimates, so we localize the source using the algorithm described in Section 2. This means that we do 1) echo sorting, 2) virtual source localization, 3) virtual source reflection, and 4) minimization of G_{RIR}. For the remaining time instances, we assume that the source position did not change significantly (by choosing the time interval appropriately), and we localize the source only by solving (7) by the gradient descent initialized at the previous estimate. Numerical experiments are described in the next section.

4. NUMERICAL SIMULATIONS

In all figures, the purple 'x' denotes the microphone position and the black circle depicts the source location. Green squares are the reflected virtual sources ordered by G_{LOC}, where smaller indices denote better scores. The solid square depicts the reflected virtual source with the best G_{RIR}, and the blue dot is obtained by solving (7) by gradient descent initialized at the position of the solid square.

We test the localization algorithm in an L-shaped room shown in Fig. 6A. The coordinates of the source are $(6, 7)$, and we use four microphones positioned uniformly at random over the square with corners at $(1, 1)$, $(1, 3)$, $(3, 3)$ and $(1, 3)$. We stop the simulation after the third order echoes. Note that there is no line-of-sight path between the source and any microphone.

Fig. 6A shows an outcome of the localization from jitter-free measurements. It can be seen that G_{LOC} prefers positions close to the source, and that G_{RIR} chooses the best position among the reflected virtual sources. The resulting solid green square overlaps with the blue dot, as in this case both G_{RIR} and the solution to (7) give perfect localization.

Fig. 6B shows localization with the measurement jitter drawn i.i.d. from a centered Gaussian with $\sigma = 0.05$. Although there are reflected sources in the vicinity of the true source position, the ones giving the best G_{LOC} are further away. However, G_{RIR} successfully discriminated the *correct* reflected source (closest to the true position). We observe that solving (7) further improves the position.

We tested the algorithm in a more complex room, as show in Fig. 6C, with $\sigma = 0.1$. The reflected virtual sources still concentrate around the real source position. Although the positions with the best localization scores are everywhere in the room, G_{RIR} selects the one that is closest to the source. Again, solving (7) improves the estimate.

Fig. 6E and Fig. 6F show the result of tracking a source that was moving along the curve $[s_1(t)\ s_2(t)]^\top : [0, \infty) \to \mathbb{R}^2$ with $s_1(t) = 4 + 3\cos^3(\pi t/60)$ [m] and $s_2(t) = 6.5 + 2\sin^3(\pi t/60)$ [m], with the jitter variance of $\sigma = 0.05$. Fig. 6E was obtained by going through all of the steps of the localization algorithm at each time instance (echo sorting, multilateration, image source reflecting), but without performing the optimization (7). In Fig. 6F we used *only* the optimization, as described in Section 3.4. As can be seen, the second, simpler approach performs significantly better.

5. CONCLUSION

We presented an algorithm for indoor localization in a known room with a general (possibly non-convex) geometry, bounded by planar walls. Our algorithm uses the early reflections to localize the source, even without the line-of-sight path. Using the received echoes we first locate the virtual sources and then find the position inside the room that generates them. We then optimize the location estimate based on the simulated room impulse response. We apply the algorithm to track a moving source, and we demonstrate how the refinement technique based on the previous position estimate improves the tracking performance. Ongoing work includes performance analysis with missing or erroneous echoes, as well as experiments in a real room with an embedded ultrasonic device.

6. REFERENCES

[1] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 37, no. 6, pp. 1067–1080, 2007.

[2] M. Harris, "The way through the flames," *IEEE Spectr.*, vol. 50, no. 9, pp. 30–35, 2013.

[3] C. Drane, M. Macnaughtan, and C. Scott, "Positioning GSM telephones," *IEEE Commun. Mag.*, vol. 36, no. 4, pp. 46–54, 59, 1998.

[4] F. Ribeiro, D. Ba, C. Zhang, and D. Florêncio, "Turning enemies into friends: Using reflections to improve sound source localization," in *IEEE Int. Conf. Multimedia and Expo*, Singapore, 2010, pp. 731–736.

[5] M. Toda, "Cylindrical PVDF film transmitters and receivers for air ultrasound," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 49, no. 5, pp. 626–634, 2002.

[6] F. Ribeiro, D. A. Florencio, D. E. Ba, and C. Zhang, "Geometrically constrained room modeling with compact microphone arrays," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 20, no. 5, pp. 1449–1460, 2012.

[7] F. Antonacci, J. Filos, M. R. P. Thomas, E. A. P. Habets, A. Sarti, P. A. Naylor, and S. Tubaro, "Inference of room geometry from acoustic impulse responses," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 20, no. 10, pp. 2683–2695, 2012.

[8] I. Dokmanic, Y. M. Lu, and M. Vetterli, "Can one hear the shape of a room: The 2-D polygonal case," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, Prague, 2011, pp. 321–324.

[9] I. Dokmanic, R. Parhizkar, A. Walther, Y. M. Lu, and M. Vetterli, "Acoustic echoes reveal room shape," *Proc. Natl. Acad. Sci. USA*, vol. 110, no. 30, pp. 12186–12191, 2013.

[10] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, pp. 943, 1979.

[11] J. Borish, "Extension of the image model to arbitrary polyhedra," *J. Acoust. Soc. Am.*, vol. 75, no. 6, pp. 1827–1836, 1984.

[12] D. Manolakis, "Efficient solution and performance analysis of 3-D position estimation by trilateration," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 4, pp. 1239–1248, 1996.

[13] K. Cheung, W.-K. Ma, and H. So, "Accurate approximation algorithm for TOA-based maximum likelihood mobile location using semidefinite programming," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, Montreal, 2004, vol. 2, pp. 145–148.

[14] A. Beck, P. Stoica, and J. Li, "Exact and approximate solutions of source localization problems," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1770–1778, 2008.

[15] K. Cheung, H. C. So, W.-K. Ma, and Y. T. Chan, "Least squares algorithms for time-of-arrival-based mobile location," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 1121–1130, 2004.