

Received April 17, 2020, accepted May 3, 2020, date of publication May 7, 2020, date of current version May 21, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2993076

Free-Field TDOA-AOA Sound Source Localization Using Three Soundfield Microphones

LUKA KRALJEVIĆ ^{ID}, (Member, IEEE), MLADEN RUSSO ^{ID}, (Member, IEEE),
MAJA STELLA ^{ID}, (Member, IEEE), AND MARJAN SIKORA ^{ID}, (Member, IEEE)

Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, 21000 Split, Croatia

Corresponding author: Luka Kraljević (lkraljev@fesb.hr)

This work was supported by the Croatian Science Foundation under Project UIP- 2014-09-3875.

ABSTRACT Accurate source localization is an important problem in many research areas as well as practical applications in wireless communications and acoustic signal processing. This paper presents a passive three-dimensional sound source localization (SSL) method that employs a geometric configuration of three soundfield microphones. Two methods for estimating the angle of arrival (AOA) and time difference of arrival (TDOA) are proposed based on Ambisonics A and B format signals. The closed-form solution for sound source location estimation based on two TDOAs and three AOAs is derived. The proposed method is evaluated by simulations and physical experiments in our anechoic chamber. Simulations demonstrate that the estimation method can theoretically obtain Cramér-Rao lower bound for a small Gaussian noise present in AOA and TDOA observations. Investigation on the uncertainty of TDOA and AOA measurements depending on the length of measurement interval is also conducted. Experimental results in terms of RMSE indicate that the proposed solution can be used to accurately find a 3D position of the sound source in free-field environment. Performance evaluation regarding the number of estimation steps shows that higher accuracy can be achieved by longer observations of stationary sound source.

INDEX TERMS 3D sound source localization, angle of arrival, Cramér-Rao bound, soundfield microphone, time difference of arrival.

I. INTRODUCTION

The sound source localization (SSL) is an essential step in a wide range of audio/acoustic-based applications. Nowadays-concerned research topics on SSL are ranging from detection of the speaker position in human-computer interaction [1] or smart video conferencing [2], robot movement in an unknown environment [3], [4], search and rescue [5] to advance military applications such as localization of a sniper [6] and medium-range aircraft localization [7]. Besides, SSL is usually used as a necessary preprocessing step before the enhancement of an acoustic signal from a particular location [8].

In general, the task of source localization can be viewed either as an active localization scenario, where transmitters actively emit signals to illuminate the target of interest while the target location is inferred by collecting reflections, wherein the scenario of passive localization, receiving sensors can only collect signals emitted from the source. In many

areas, only a passive source location is considered, where the signals usually do not carry the time information about their transmission. The sound localization as a type of passive localization refers to the problem of estimating the position from which a sound signal originates concerning the microphone array geometry. In this case, a localization system is unable to directly measure the time of arrival (TOA) between the source location and receiving sensor, but instead, only the difference between times when the different sensors receive the signals can be measured.

Various methods for SSL have been proposed where all methods can be grouped by their efforts to detect sound source either in 2D space [9], [10] or in 3D space [3], [11]. Fundamentally, there are two main approaches to finding a source with respect to recorded audio signals. Both approaches are mainly based on estimating the time difference of arrival (TDOA) obtained by using various configurations of microphone arrays, such as linear array [12], circular array [13], or distributed array [14] and different cross-correlation algorithm to estimate time lag between microphones. The first approach aims to maximize the

The associate editor coordinating the review of this manuscript and approving it for publication was Manuel Rosa-Zurera.

steered response power (SRP) of the output of a delay and sum beamformer [15]. This direct approach performs an exhaustive search in the whole SRP space to find a sound location, which is found to be computationally expensive. In contrast, the indirect approach is based on using estimated TDOAs measurements, where the sound source is ascertained by addressing criteria such as the hypercone fitting problem [16]. Although results obtained by the indirect method are more prone to error as they are more sensitive to background noise, and reflections oppose to direct methods, the main advantage of indirect TDOA based approach is that it can be effectively used in distributed microphone network since it only requires for TDOA values to be transmitted and not raw sound signal data [17].

A. RELATED WORK

In many research papers, TDOA is a standard measurement used for passive localization [18], [19]. Source localization based solely on TDOA measurements demanded a specific number of measuring devices, i.e., a minimum number of three and four sensors are required to locate an unknown target respectively in 2D and 3D space. The principle of TDOA based approach is to estimate source position from the intersection of hyperbolic arcs and surfaces respectively for 2D and 3D cases. In the case of near-field applications where the source range to the sensor array baseline ratio is not large, the resulting intersection can be obtained by solving a set of nonlinear equations. On the contrary, in the far-field applications, resulting intersection produces a low location estimate since the hyperbolic arcs/surfaces become almost parallel to one another. Given the evidence that the accuracy of the source position estimate degrades when the source moves sufficiently far away from the sensors resulted in two different localization problems. Recently, unified near/far-field TDOA based localization was proposed in [20]. The proposed approach consisted of two formulations for the unified localization problem in two different coordinate systems. The first formulation is of a nonlinear non-convex weighted least squares optimization based on the modified polar representation for the source position, and the other is the non-convex fractional programming formulation using the conventional Cartesian coordinates of the source position as the optimization variable. Besides the TDOA-based approach, a source location can also be calculated from the AOA measurements and its derivative. Obtaining AOA often involves a sensor that is equipped with an array of receivers; thus, it elevates a requirement of synchronization between different sensors since each produces angle by itself. In [21], the authors proposed a solution that can attain Cramér-Rao lower bound under mild conditions. Another 3D bearing-only localization is proposed by [22], where authors achieved a significant reduction in bias and root-mean-square error using a pseudo-linear estimator. In general, the passive source localization problem is not trivial since direct relationships between the position of a source and the measurements are

complex, and procedures for solving equations for the TDOA and AOA methods are hard because of nonlinearity.

One of the more recent research directions is in combining TDOA and AOA measurements, where algebraic manipulations allow transforming the relationships to the linear form and lowering energy consumption [23]. There are several advantages with the hybrid TDOA-AOA approach, such as improved localization performance [24], [25], reduced number of sensors required [26] and it can minimize the occurrence of ghost targets which is typical for localization approach with individual TDOA measurements [27]. Many studies have been carried out proposing a different solution to source localization [26], [28]–[31]. The most straightforward method to source localization is an exhaustive search in a feasible solution region, which is a time-consuming and inefficient solution for real-time application. In general, the maximum likelihood (ML) estimator is introduced to estimate the location since it is asymptotically efficient.

However, the aftermentioned approaches are computationally expensive, and it is hard to find a closed-form solution, or a closed-form solution does not exist at all. One of the solutions is to linearize equations with a recursive approach such as Taylor-series. Nevertheless, these numerical search techniques can converge to an optimal solution only if the ML function is convex. The numerical methods are prone to error since they depend on the right initial position guess, and thus it is difficult to guarantee its global convergence and calculating time. Their iterative nature does not make them very suitable for real-time applications. To improve the robustness and reduce the complexity, a closed-form solution is required. A linear least-squares estimator with the closed-form solution, called pseudo linear estimator (PLE), was proposed in [22]. Although this approach is less computationally demanding, the estimated source position is biased because of the correlation between system and measurement. Recently, authors [32] proposed a new method for localization. The proposed method represents a simple algebraic solution that does not suffer from the local convergence problem. However, this method also has a larger bias since localization accuracy is affected by the deviation of the noise correlation matrix. Another fine localization solution, based on generalized trust-region subproblem technique is proposed by [33], where authors analyzed the necessary optimal conditions of squared range difference least square cost function. To reduce the bias of the estimator, a few different methods were proposed. In [34] BR-PLE method is proposed for reducing bias. Authors in [35] proposed a solution to find a rotation angle that could reduce the bias of the estimator since they demonstrated that estimator performance was sensitive to origin rotation. [36] proposed another solution for reducing estimator bias in the presence of sensors error. The proposed method introduces a quadratic constraint so that estimator expectation cost function can attain the minimum value at a true position and thus to achieve Cramer-Rao lower bound. By analyzing different approaches, one could conclude that for the most methods,

an estimation bias arises from the least-squares techniques. One of the solutions is to construct a new pseudo linear system such as [27], where a new weight least-square method is proposed. Authors claimed that the novel structured total least squares method could reduce estimation bias when the target is outside the convex hull formed by measurement sensors. The total least squares estimator was also considered in [37], where it has demonstrated improved accuracy over least-square solutions. Another practical solution to position estimation was proposed by [38]. The proposed closed-form method was based on converting time-delay measurements to angular information, but it couldn't achieve the CRLB performance. Recently, an efficient closed-form solution has been proposed for passive source location using only two stations [26]. In their work, a new relationship between hybrid TDOA and AOA measurements and unknown source positions was constructed. The theoretical simulation shows that the proposed solution can achieve CRLB for Gaussian noise, where bias compared to variance can be ignored.

By analyzing the aforementioned different source localization approaches, some general remarks regarding localization performance can be made. Investigating different types of input data for source estimation, one can observe that using hybrid TDOA and AOA measurements reduces the number of sensors required, which is especially important for a real-life scenario where the basic assumption about line-of-sight (LOS) between unknown source position and sensor may not be respected. In many the non-line-of-sight (NLOS) environments, signals emitted from the source are often inaccessible by all measurement sensors, and one of the solutions is to use wireless sensor networks (WSN) in which received signals by each sensor are transmitted to the fusion center for source localization. Authors in [39] proposed a distributed NLOS cooperative localization algorithm. The proposed localization algorithm employs the multiplicative convex model based on the physical mechanism of the NLOS propagation to achieve robustness in changing environments. A TDOA-based cooperative localization approach for mixed LOS/NLOS conditions is proposed by [40]. For the location of the multiple stationary target nodes, the authors formulated a non-convex robust weighted least squares problem (RWLS). To efficiently solve RWLS, the semidefinite relaxation technique is used to transform RWLS into a convex mixed semidefinite and second-order cone programming problem. Authors in [41] presented the energy-based localization solution in WSNs using received signal strength (RSS) and received signal strength difference (RSSD). In the proposed solution, RSSD is based on transforming the nonlinear and non-convex objective functions into a convex optimization problem via relaxation and semidefinite programming. Another mixed semidefinite and second-order cone relaxation for source localization in 3D WSN was proposed by [42]. The proposed target node localization is based on hybrid RSS-AOA measurements in both noncooperative and cooperative WSNs, where the authors proposed new LS estimator to reduce the implementation costs. From conducted analysis authors

concluded that proposed RSS-AOA based estimator is more suitable for large-scale cooperative WSNs compared with AOA-TDOA based estimators. In general, centralized algorithms suffer from high computation complexity, network traffic bottleneck, and as such, are not recommended in scenarios where each sensor node cannot get raw measurements directly. To address such problem authors in [43] proposed a completely decentralized localization approach based on augmented Lagrangian methods and alternating direction method of multipliers (ADMM). Discussing performance comparison in terms of computational efficiency, it is visible that the unknown source position can be directly calculated from different sets of geometric relationships. Although computational complexity for this kind of approach is low, it frequently does not perform sufficiently when measurement error exists. Since TDOA and AOA observations have measurement error, the location of a source is often estimated by propagating these errors through the computation, where iterative nonlinear minimization is required for an optimal solution. The most straightforward approaches for handling measurement noise are iterative algorithms based on initial position estimate, obtained, for example, by the Gauss-Newton method, which has high computational requirements. The maximum likelihood estimator is asymptotically efficient, but it requires a good initial guess. To avoid the need for good initial position guess characteristic for iterative ML approach, different closed-form source estimation methods are proposed. A linear LS approach is an alternative approach, which can achieve CRLB but reports large estimation bias. In general, various closed-form solutions have been proposed, each designed to reduce biases, or to work with a different number of sensors. To assess the performance of localization estimation, most state-of-the-art studies perform benchmark in terms of CRLB. Analyzing literature, one can see that the weighted least square estimator is superior in comparison to LS solutions in terms of lower bias. Also, it can be concluded that even both WSL and ML approaches can achieve CRLB, the WSL approach is more computationally attractive and suitable for real-time application and does not really on the initial guess. From the performance comparison between different methods, one can conclude that while most state-of-the-art closed-form solutions achieve CRLB, each approach is designed to work under restricted measurement error.

In the context of room acoustics, it is difficult to establish measurement error as the measurement microphone captures not only the direct-path component of the source signal but also the multipath component caused by reflections. The multipath component, together with the background noise, can lead to distortion of the time delay information from received signals, and thus it can degrade the localization performance. To address key challenges of the realistic environment such as room reverberation, background noise, and sound interference, different methods to compute the TDOAs across various combinations of pairs of spatially separated microphones were proposed [44]–[47]. Recent work also

suggested that deep learning can successfully be applied for modeling rooms acoustic [48]. For instance, in [49] authors employed deep model for SSL, where it was shown that deep learning-based system achieved higher accuracy under low SNR conditions in comparison with cross-correlation phase transform (GCC-PHAT) method. Authors in [17] proposed a novel learning approach for SSL based on TDOA estimation, where coordinates of a sound source were defined as functions of TDOA. In their work, pre-recorded sound measurements and their corresponding source locations were used to train the multilevel B-Splines based learning model. A new dataset for learning-based SSL was proposed by [50] which contained different acoustic events recorded in the anechoic chamber, where the anechoic chamber environment was used to verify the feasibility of the proposed baseline model. Authors in [51] addressed SSL for indoor environments with high reverberation and low signal-to-noise ratio. The authors proposed a novel sound source localization method using a probabilistic neural network for the classification of 3D space clusters. Another learning-based approach was demonstrated in [52], where authors evaluated the Structural Sparse Bayesian Learning model with signals recorded in an anechoic chamber with one reflective plate. Moreover, it can be seen that many researchers validated their SSL methods by using recorded audio signals simulating free field conditions or experimental analysis was carried out inside a semi-anechoic and anechoic chamber. The anechoic chamber provides a good simulation of the outdoor conditions due to the low level of reflections. This environment provides the possibility to precisely control the conditions and to measure the levels of sound events and noise, which is substantial for many experiments. Kotus *et al.* [53] performed multiple sound sources localization in the anechoic chamber, where different methods for obtaining the direction of arrival were tested. In the work [54], the authors proposed a modified cross-correlation algorithm to obtain a more reliable measurement of time difference of arrival in the reverberation environment. They performed a triangulation procedure using calculated TDOA values to obtain a sound source position in 3D space. Another three-dimensional method for SSL was proposed by Ding *et al.* [55]. The authors performed theoretical simulation accompanied by experimental results in the anechoic chamber. They proposed the use of a planar microphone array combined with a beamforming technique to obtain the location of the point sound source.

B. OUTLINE

This paper presents a passive sound source localization method using three soundfield microphone stations. The geometric configuration of three soundfield microphones can be employed to obtain two TDOA and three AOA measurements concerning unknown source position. A closed-form mathematical solution for SSL estimation is presented. Results are given in terms of RMSE, where it was shown that simulation of the sound source estimation algorithm can reach Cramér-Rao lower bound for small Gaussian noise presented in mea-

surements. Theoretical simulation is supported by the experimental analysis conducted in our anechoic chamber [56]. In this work, TDOA and AOA measurements are directly obtained by exploiting the A and B format of the soundfield microphones, which can achieve small measurement deviations. This paper also demonstrates the value of using a soundfield microphone for the SSL task because AOA can be easily obtained due to the configuration of the soundfield microphone capsules.

The rest of the paper is organized as follows. Section II describes the mathematical formulation of the SSL algorithm. Section III presents our TDOA and AOA estimation methods using soundfield microphones. The simulation and experimental results in the anechoic chamber are given in Section IV, and Section V concludes the paper.

II. 3D SOURCE LOCATION ESTIMATION

This section explains a novel sound source localization method based on three soundfiled microphone stations. First, we establish a geometrical relationship between an unknown position of the sound source and known positions of sensors, over obtained AOA and TDOA observations. Second, we define a weighted least squares estimator, where the sum of the squared residuals is minimized with respect to the error measurement vector. Measurement error is modeled as a covariance matrix containing measurement uncertainties. In experimental phase those measurement error parameters are obtained through testing in the anechoic chamber [56].

Here we presented the theoretical formulation of the localization scenario where a geometric configuration of three stations $\mathbf{m}_i = [m_{x,i}, m_{y,i}, m_{z,i}]^T \in \mathbb{R}^3$ is used to estimate the position of the single sound source $\mathbf{s} = [s_x, s_y, s_z]^T \in \mathbb{R}^3$ in 3D space. Assuming that each station on position \mathbf{m}_i can determine bearing angles of received sound wave transmitted from the unknown source position \mathbf{s} , geometric relationship between \mathbf{s} and \mathbf{m}_i can be expressed through observed AOAs by nonlinear equation:

$$\begin{bmatrix} \theta_i \\ \phi_i \end{bmatrix} = \begin{bmatrix} \arctan \left(\frac{s_y - m_{y,i}}{s_x - m_{x,i}} \right) \\ \arctan \left(\frac{s_z - m_{z,i}}{\sqrt{(s_x - m_{x,i})^2 + (s_y - m_{y,i})^2}} \right) \end{bmatrix}, \quad i = 1, 2, 3 \quad (1)$$

where $\theta_i \in (-\pi, \pi)$ and $\phi_i \in (0, \pi/2)$ form an AOA pair corresponding to azimuth and elevation angles in the right-hand coordinate system.

Let r_i be the Euclidean distance between the source and microphone m_i :

$$r_i = \|\mathbf{s} - \mathbf{m}_i\| = \sqrt{(\mathbf{s} - \mathbf{m}_i)^T (\mathbf{s} - \mathbf{m}_i)}, \quad i = 1, 2, 3. \quad (2)$$

and r_1 is the true distance from \mathbf{m}_1 to the unknown source, then all TDOA observations with respect to \mathbf{m}_i are given as:

$$\tau_{i1} = r_i - r_1 = \Delta t_{i1} * v, \quad i = 2, 3. \quad (3)$$

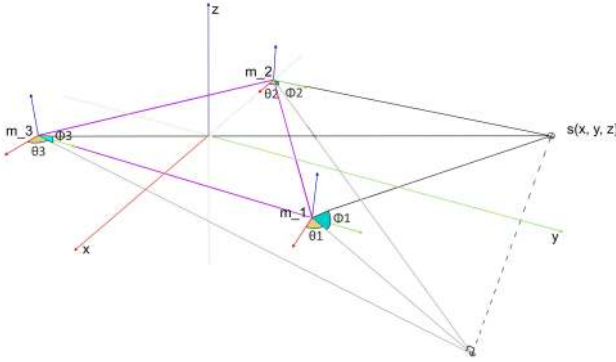


FIGURE 1. Localization geometry.

where τ_{i1} is the range difference, $\Delta_{t_{i1}}$ is the actual TDOA measurement obtained as the propagation lag between m_i and m_1 and v is the speed of signal propagation.

To estimate the unknown source position s using true TDOA and AOA values, we define vector $\boldsymbol{\mu} = [\tau_{21}, \tau_{31}, \kappa_1^T, \kappa_2^T, \kappa_3^T]^T \in \mathbb{R}^8$ containing two TDOA observations with respect to m_1 and three AOA observations produced from positions m_1, m_2 and m_3 . In practice, the AOA and TDOA are noisy observations where we assume that true values are influenced by additive Gaussian noise as:

$$\begin{aligned} \hat{\kappa}_i &= \kappa_i + \boldsymbol{\epsilon}_i = \begin{bmatrix} \theta_i + n_i \\ \phi_i + m_i \end{bmatrix}, \\ \hat{\tau}_{i1} &= \tau_{i1} + \varepsilon_{i1} \end{aligned} \quad (4)$$

When expressing noisy TDOA and AOA in vector form results in:

$$\hat{\boldsymbol{\mu}} = \boldsymbol{\mu} + \boldsymbol{\varepsilon} \quad (5)$$

where $\hat{\boldsymbol{\mu}}$ is actual noisy measurement vector $\boldsymbol{\mu}$ and $\boldsymbol{\varepsilon} = [\varepsilon_{21}, \varepsilon_{31}, \boldsymbol{\epsilon}_1^T, \boldsymbol{\epsilon}_2^T, \boldsymbol{\epsilon}_3^T]^T \in \mathbb{R}^8$ is zero-mean Gaussian with covariance matrix $\boldsymbol{Q} \in \mathbb{R}^{8 \times 8}$

For localization scenario shown in Fig 1. we construct a unit norm vector $\mathbf{b}_i = [\cos \phi_i \cos \theta_i, \cos \phi_i \sin \theta_i, \sin \phi_i]^T \in \mathbb{R}^3$ so that $\mathbf{b}_i^T \mathbf{b}_i = 1$ and to satisfy the given geometric relation:

$$\mathbf{s} - \mathbf{m}_i = r_i \mathbf{b}_i, \quad i = 1, 2, 3. \quad (6)$$

By constructing the matrix \mathbf{G}_i in order to have columns that are orthonormal basis of the plane orthogonal to \mathbf{b}_i , it follows that $\mathbf{G}_i^T \mathbf{G}_i = \mathbf{I}_{2 \times 2}$ and $\mathbf{G}_i^T \mathbf{b}_i = \mathbf{0}_2$ where $\mathbf{I}_{2 \times 2}$ and $\mathbf{0}_2$ are identity matrix and zero vector. Let $\mathbf{G}_i \in \mathbb{R}^{3 \times 2}$ be:

$$\mathbf{G}_i = \begin{bmatrix} \sin \theta_i & \sin \phi_i \cos \theta_i \\ -\cos \theta_i & \sin \phi_i \sin \theta_i \\ 0 & -\cos \phi_i \end{bmatrix}, \quad i = 1, 2, 3. \quad (7)$$

By performing left multiplication on (6) with \mathbf{G}_i^T we get the following expression:

$$\mathbf{G}_i^T \mathbf{u} = \mathbf{G}_i^T \mathbf{m}_i, \quad i = 1, 2, 3. \quad (8)$$

By realizing that $\mathbf{b}_i^T \mathbf{b}_j = \mathbf{b}_j^T \mathbf{b}_i$ and $\mathbf{b}_i^T \mathbf{b}_i = 1$ we construct identity:

$$\mathbf{b}_i^T \mathbf{b}_i + \mathbf{b}_i^T \mathbf{b}_j - \mathbf{b}_j^T \mathbf{b}_i + \mathbf{b}_j^T \mathbf{b}_j = 0, \quad i \neq j = 1, 2, 3. \quad (9)$$

Equation (9) can be rewritten with respect to m_1 as

$$r_i (\mathbf{b}_i - \mathbf{b}_1)^T (\mathbf{b}_i + \mathbf{b}_1) = 0, \quad i = 2, 3. \quad (10)$$

where r_i is the geometric distance from m_i to s . Substituting (3) as r_i in (10) we have $r_i (\mathbf{b}_i + \mathbf{b}_1) = r_i \mathbf{b}_i + (r_1 + \tau_{i1}) \mathbf{b}_1$ and after using (6) as $r_i = \mathbf{b}_i^T (\mathbf{u} - \mathbf{m}_i)$ equation (10) becomes:

$$2(\mathbf{b}_i - \mathbf{b}_1)^T \mathbf{s} = (\mathbf{b}_i - \mathbf{b}_1)^T (\mathbf{m}_1 + \mathbf{m}_i - \tau_{i1} \mathbf{b}_1), \quad i = 2, 3. \quad (11)$$

A closed form solution was obtained in [26] for two stations and one source, while in this work, we expand the solution to work with three stations. Using (8) and (11) together in matrix form gives:

$$\mathbf{h} = \mathbf{G}^T \mathbf{s} \quad (12)$$

$$\mathbf{h} = [\beta_{21}, \beta_{31}, s_1^T \mathbf{G}_1, s_2^T \mathbf{G}_2, s_3^T \mathbf{G}_3]^T \in \mathbb{R}^8 \quad (13)$$

$$\beta_{i1} = (\mathbf{b}_i - \mathbf{b}_1)^T (\mathbf{m}_1 + \mathbf{m}_i - \tau_{i1} \mathbf{b}_1), \quad i = 2, 3 \quad (14)$$

$$\mathbf{G} = [2(\mathbf{b}_2 - \mathbf{b}_1), 2(\mathbf{b}_3 - \mathbf{b}_1), \mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3] \in \mathbb{R}^{3 \times 8} \quad (15)$$

Equation (12) represents an ideal condition which doesn't hold in practice since there is a measurement error $\boldsymbol{\varepsilon}$ in \mathbf{h} and \mathbf{G} matrix. To analyze the influence of the error $\boldsymbol{\varepsilon}$ for a given measurement vector $\hat{\boldsymbol{\mu}}$ in (5) on localization accuracy, $\boldsymbol{\varepsilon}$ will be introduced in (1) and (3) where the true geometric values of TDOA and AOA will be expressed in terms of their noisy observations (4). We can approximate (12) up to first order noise terms which gives:

$$\hat{\mathbf{h}} \approx \hat{\mathbf{G}}^T \mathbf{s} + \mathbf{T} \boldsymbol{\varepsilon} \quad (16)$$

where $\hat{\mathbf{h}}$ and $\hat{\mathbf{G}}$ are matrices containing noisy measurements instead of true TDOA and AOA values as their counterparts \mathbf{h} and \mathbf{G} . The formulation of \mathbf{T} matrix is shown in (18) - (20), as shown at the bottom of the next page. From (5) it follows that $\mathbf{T} \boldsymbol{\varepsilon}$ in (16) is zero-mean Gaussian with covariance matrix \boldsymbol{Q} . To calculate a weighted Least Square estimate of \mathbf{s} from (16), the sum of squared residuals is minimized with the respect to error measurement vector $\boldsymbol{\varepsilon}$.

$$\begin{aligned} \hat{\mathbf{s}} &= \arg \min \|\hat{\mathbf{h}} - \hat{\mathbf{G}}^T \mathbf{s}\|^2 \\ &= \arg \min (\hat{\mathbf{h}} - \hat{\mathbf{G}}^T \mathbf{s})^T \mathbf{W}^{-1} (\hat{\mathbf{h}} - \hat{\mathbf{G}}^T \mathbf{s}) \\ &= (\hat{\mathbf{G}} \mathbf{W}^{-1} \hat{\mathbf{G}}^T)^{-1} \hat{\mathbf{G}}^{-1} \hat{\mathbf{h}} \end{aligned} \quad (17)$$

where $\mathbf{W} = \mathbf{T} \boldsymbol{Q} \mathbf{T}^T$ is weight matrix. The expression (17) is based on the implicit assumption that the measurement errors are uncorrelated with each other and that TDOA and AOA observations have corresponding σ_{RD} and σ_{AOA} uncertainties. This is insured by modeling \boldsymbol{Q} as a matrix, where the diagonal elements are given in form of two TDOA and three AOA variances, setting the rest of off-diagonal entries to null.

III. TDOA-AOA ESTIMATION METHOD

In this section, we explain the proposed approach for estimating TDOA and AOA values, computed from the real sound measurements collected by three soundfield microphone stations. Given that the solution (17) performs minimization of the sum of squared residuals with respect to measurement error, it was essential to determine observation uncertainties σ_{RD} and σ_{AOA} of the proposed TDOA-AOA estimation method in free field conditions. In this work, to ensure universal reference regarding methods uncertainties, all measurements are made in a small anechoic chamber [56] which facilitates free-field conditions where no reverberation of the sound source is present. Fig. 2 displays our measurement setup with three soundfield microphones and a loudspeaker representing a single sound source inside an anechoic chamber.

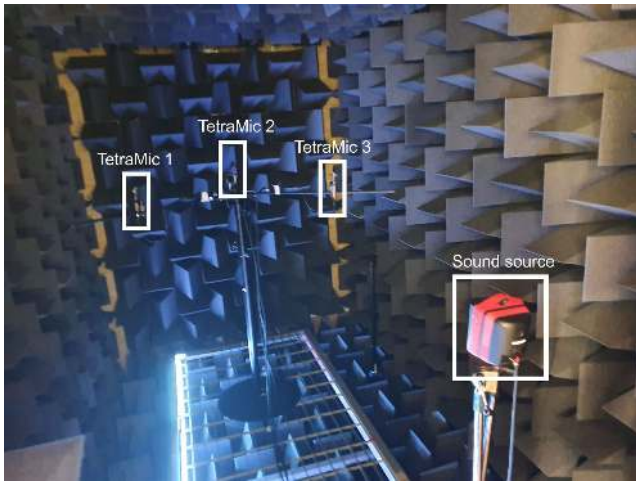


FIGURE 2. Anechoic chamber with triangular stand counting three soundfield microphones and speaker representing sound source.

Soundfield microphone uses 4 subcardioid capsules mounted as close as possible to form a tetrahedron. Each soundfield microphone can be viewed as four symmetrical receivers positioned on the surface of the sphere where it can produce two distinct sets of audio signals called A-format and B-format. A-format consists of 4 signals coming from each microphone capsule arranged as shown in Fig. 3.

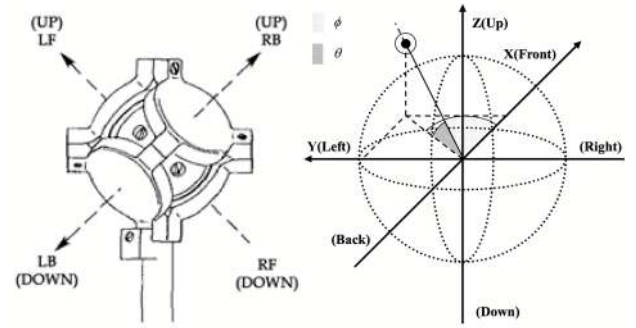


FIGURE 3. Soundfield microphone.

The B-format signals comprise a truncated spherical harmonic decomposition of the sound field. They correspond to the sound pressure and the three components of the pressure gradient at a point in space. The transformation from A to B-format can be easily performed by knowing the measurement values of the individual capsules in A-format. A linear system of equations displayed below can be used for format conversion:

$$\begin{aligned}
 p_w &= p_{LF} + p_{RB} + p_{RF} + p_{LB}, \\
 p_x &= p_{LF} - p_{RB} + p_{RF} - p_{LB}, \\
 p_y &= p_{LF} - p_{RB} - p_{RF} + p_{LB}, \\
 p_z &= p_{LF} - p_{RB} - p_{RF} - p_{LB},
 \end{aligned} \tag{21}$$

where p_w is sound pressure signal at the microphone position, p_x is the sound velocity in the direction back and forth, p_y is the sound velocity information in the direction left to right, and p_z is the sound velocity information in the direction up and down. Additionally signal filtering can be used to compensate for inequalities between individual capsules.

To validate estimation algorithm (17) using real measurements, this paper proposes two methods for calculating TDOA and AOA values from observed sound signal emitted from unknown source location s .

A. TDOA ESTIMATION

The system is composed of an unknown sound source and three soundfield microphones. If the sound source transmits a signal at time $T = 0$, the microphones will sense the

$$T = \begin{bmatrix} -(b_2 - b_1)^T b_1 & 0 & r_1 b_2^T L_1 & -r_2 b_1^T L_2 & \mathbf{0}_{1 \times 2} \\ 0 & -(b_3 - b_1)^T b_1 & r_1 b_3^T L_1 & \mathbf{0}_{1 \times 2} & -r_3 b_1^T L_3 \\ \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & T_1 & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & T_2 & \mathbf{0}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & T_3 \end{bmatrix} \tag{18}$$

$$L_i = \begin{bmatrix} -\cos \phi_i \sin \theta_i & -\sin \phi_i \cos \theta_i \\ \cos \phi_i \cos \theta_i & -\sin \phi_i \sin \theta_i \\ 0 & \cos \phi_i \end{bmatrix} \tag{19}$$

$$T_i = -r_i \begin{bmatrix} \cos \phi_i & 0 \\ 0 & 1 \end{bmatrix}, \quad i = 1, 2, 3. \tag{20}$$

signals at the unknown times T1, T2, and T3. Following (3), the difference between TOAs can be measured as follows:

$$\begin{aligned} \Delta T_{21} &= T_2 - T_1, \\ \Delta T_{31} &= T_2 - T_1, \end{aligned} \quad (22)$$

where the microphone at position m_1 is taken as a reference. Note that for the case of three microphones ΔT_{32} value could be additionally calculated to introduce redundancy in the system. To calculate the time lag between two microphones, in this work we used the generalized cross-correlation phase transform (GCC-PHAT) algorithm as follows:

$$\begin{aligned} \hat{G}_{PHAT}(f) &= \frac{X_1(f)[X_i(f)]^*}{|X_1(f)[X_i(f)]^*|}, \\ \hat{d}_{PHAT}(1, i) &= \arg \max(\hat{R}_{PATH}(d)) \end{aligned} \quad (23)$$

where x_i is a sound signal received by the microphone at position m_i , X_i is Fourier transform and $\hat{R}_{PATH}(d)$ is the inverse Fourier transform of i -th signal and $[\]^*$ denotes the complex conjugate. The term \hat{d}_{PHAT} corresponds to the estimated time difference between m_1 and m_i . To estimate TDOA using soundfield microphones, in this work, A-format is used which enables to calculate TDOA as average time lag between corresponding four capsules of the observed microphones i.e. $\hat{d}_{PHAT}(1_{LF}, i_{LF})$, $\hat{d}_{PHAT}(1_{LB}, i_{LB})$, $\hat{d}_{PHAT}(1_{RB}, i_{RB})$ and $\hat{d}_{PHAT}(1_{RF}, i_{RF})$. Note that for this pairwise TDOA calculation, each soundfield microphone should be aligned with global coordinate axes. Calculating the range difference from (3) is straightforward, where $v = 331.57 + 0.607\lambda$ [m/s] is speed of sound in the air and λ is air temperature in C° . As already mentioned, algebraic solution (17) assumes that the measurement vector is corrupted with small Gaussian noise defined by covariance matrix \mathcal{Q} , pointing to the need of investigating the uncertainty of our TDOA estimation method. In order to analyze our TDOA based estimation performance given measurement duration t , we examined uncertainty reduction when the length of the sound signal used for estimation increases. From the results shown in Fig. 4 it is visible that by choosing measurement duration $t > 300$ ms will provide stable TDOA observations ($RD < 5$ cm). Fig. 4 illustrates deviation of time difference regarding m_1 and m_2 for a given sound source location $s = [160, -30, -20]^T$.

B. AOA ESTIMATION

In this work, we propose a method for calculating AOA estimation by exploiting B-format signals of the soundfield microphone. The AOA estimation technique is based on obtaining directivity vector $d = [d_x, d_y, d_z]^T \in \mathbb{R}^3$. The procedure is performed by calculating the power of each velocity signal for a given integration interval t , where each velocity signals correspondent to X, Y, Z component of B-format. Next, normalization is applied by dividing each p component with $R = \sqrt{p_x^2 + p_y^2 + p_z^2}$ where p_x, p_y and p_z represents the total power over t for each discrete axis. To produce a pair of AOA observations from the calculated directional vector d , similar

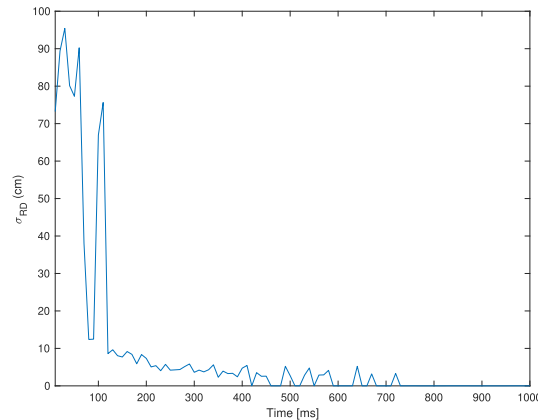


FIGURE 4. Dependency between standard deviation of TDOA estimation and measurement length.

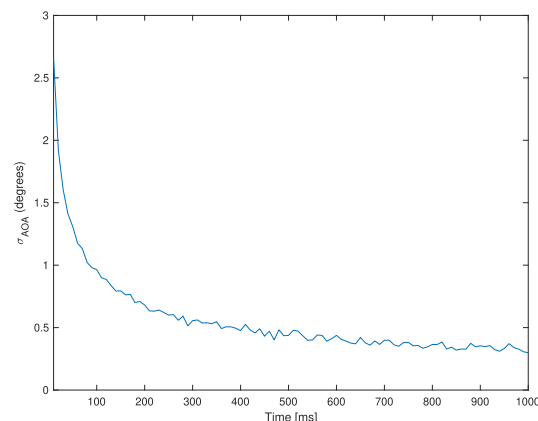


FIGURE 5. Dependency between standard deviation of AOA estimation and measurement length.

expression as (1) is used:

$$\begin{bmatrix} \theta_i \\ \phi_i \end{bmatrix} = \begin{bmatrix} \arctan\left(\frac{d_{y,i}}{d_{x,i}}\right) \\ \arctan\left(\frac{d_{z,i}}{\sqrt{d_{x,i}^2 + d_{y,i}^2}}\right) \end{bmatrix}, \quad i = 1, 2, 3 \quad (24)$$

where θ_i and ϕ_i are angles concerning microphone internal origin. It should be noted here that (24) is prone to the error if the axis of each soundfield microphone used for AOA estimation are not perfectly aligned with the axis of global geometry.

Following similar logic as in previous subsection III-A, we investigated the dependency between the directivity error and integration time i.e. duration of the sound signal used for calculating AOA observations. From the Fig.5 it is visible that for $t > 400$ ms our methods starts to stabilize, providing estimation results with standard deviation of $\sigma_{AOA} < 0.5^\circ$. Fig. 5 illustrates standard deviation of the estimated angle of directional vectors, measured by m_1 given the position of the sound source $s = [160, -30, -20]^T$.

Based on our findings in terms of dependency between σ_{AOA} and t and by following the conclusion regarding σ_{RD} and t from subsection III-A, we have chosen duration of the signal measurement to be $t = 500\text{ms}$ in all our subsequent experiments.

IV. LOCALIZATION RESULTS AND PERFORMANCE EVALUATION

This section presents the performance study regarding the localization accuracy of the proposed SSL approach in the free field conditions. First, a simulation is made to demonstrate a margin of localization error regarding uncertainty in TDOA-AOA observations. To alleviate the dependency on the particular geometry, TDOA-AOA values are calculated by sampling the angles between three microphone stations from a uniform distribution. In total, three scenarios are conducted to analyze how measurement error affects the deviation of localization results compared to theoretical CRLB. Second, the experimental validation of the simulation results is made in the anechoic chamber containing one sound source and three soundfield microphones. Anechoic environment, however far from being realistic, gives the possibility to precisely control the environmental conditions and to accurately determine the measurement deviations of the proposed TDOA-AOA estimation method, which is substantial in this experiment. If the experiment was carried out in a reverberant room, the room acoustics would influence the estimation of TDOA-AOA values and thus obtained measurement uncertainties would not make a universal reference. Identifying measurement uncertainties σ_{RD} and σ_{AOA} , proposed solution (17) can be tested using estimated TDOA-AOA values calculated from real sound source.

A. SSL SIMULATION UNDER GAUSSIAN NOISE

This subsection presents performance analysis of the proposed estimation method where it will be shown that for the given localization scenario, the proposed method can reach CRLB for Gaussian noise present in TDOA and AOA measurements. For the known source position s , CRLB matrix is defined by [57]:

$$CRLB(s) = FIM^{-1}(s) \tag{25}$$

$$FIM(s) = \left(\frac{\partial \boldsymbol{\mu}}{\partial \mathbf{s}^T}\right)^T \mathbf{Q}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \mathbf{s}^T} \tag{26}$$

where FIM is the Fisher information matrix given as (26) for zero-mean Gaussian with a covariance matrix $\mathbf{Q} \in \mathbb{R}^{8 \times 8}$. Taking partial derivative of $\boldsymbol{\mu}$ with respect to s_x , s_y and s_z yields:

$$\frac{\partial \boldsymbol{\mu}}{\partial \mathbf{s}^T} = [c_{21}, c_{31}, \mathbf{D}_1^T, \mathbf{D}_2^T, \mathbf{D}_3^T]^T \in \mathbb{R}^{8 \times 3} \tag{27}$$

where:

$$c_{i1} = \frac{s - m_i}{r_i} - \frac{s - m_1}{r_1}, \quad i = 2, 3. \tag{28}$$

$$\mathbf{D}_m = \begin{bmatrix} \frac{-d_{y,i}}{l_i^2} & \frac{d_{x,i}}{l_i^2} & 0 \\ \frac{-d_{x,i}d_{z,i}}{r_i^2 l_i} & \frac{-d_{y,i}d_{z,i}}{r_i^2 l_i} & \frac{l_i}{r_i^2} \end{bmatrix},$$

$$l_i = \sqrt{d_{x,i}^2 + d_{y,i}^2}, \quad i = 1, 2, 3 \tag{29}$$

values $d_{x,i} = s_x - m_{x,i}$ and respectively $d_{y,i}$ and $d_{z,i}$ are the difference from source position s and m_i for a given direction.

The theoretical value of CRLB presents the best possible accuracy that estimator can achieve given small measurement error. To adequately analyze the performance of the proposed method, evaluation is made in terms of the difference between the theoretical value of square root CRLB and root mean square error (RMSE) of the estimation algorithm for a given measurement error variance $\sigma_{AOA}^2, \sigma_{RD}^2$. In this work, measurement error covariance matrix \mathbf{Q} is modeled as:

$$\mathbf{Q} = \begin{bmatrix} \sigma_{RD}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{RD}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{AOA}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{AOA}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{AOA}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{AOA}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{AOA}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{AOA}^2 \end{bmatrix} \tag{30}$$

In total, three different simulations for a sound source localization scenario is made. Firstly, we inspect the performance of the proposed estimation method when the uncertainty of TDOA measurement increases. Secondly, we perform a similar test by increasing the standard deviation σ_{AOA} . And for the third simulation, a difference of RMSE from theoretical root CRLB will be analyzed in a scenario where the range of the source s is increasing from the origin.

For an input simulation data, TDOA and AOA observations are modeled as (4), where τ_{i1} and κ_i correspond to true geometrical values calculated from (1) and (2), while ε_{i1} and ϵ_i are corresponding zero-mean Gaussian noise defined by error variance $\sigma_{AOA}^2, \sigma_{RD}^2$. We should also note that measurement noise between each TDOA and AOA observation is uncorrelated.

Each simulation results are reported in terms of RMSE which is defined as $RMSE(s) = \sqrt{\sum_{l=1}^L \|\check{s}_l - s\|^2 / L}$, where \check{s}_l is the estimated source position after l -th iteration and s is true position of the source. $L = 500$ is the number of estimation runs for a given $\sigma_{RD}, \sigma_{AOA}$ and s . Note that weight matrix \mathbf{W} in (17) depends on actual source position via (18) which is not know in advance. We can define \mathbf{W} as an identity matrix $\mathbf{I} \in \mathbb{R}^{8 \times 8}$ to obtain an initial position estimates \check{s}_1 from (17) which in this case reduces (17) back to LS estimation. For the rest of the simulation, the matrix \mathbf{W} is computed using previously estimated source position \check{s}_{l-1} together with the current AOA observation. Simulation procedure respects geometric configuration shown in Fig. 1 where three stations are placed 1 m apart forming vertices of an equilateral triangle

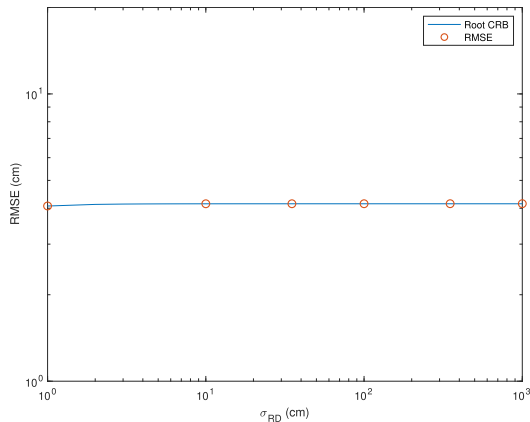


FIGURE 6. Simulation 1: RMSE dependency on σ_{RD} .

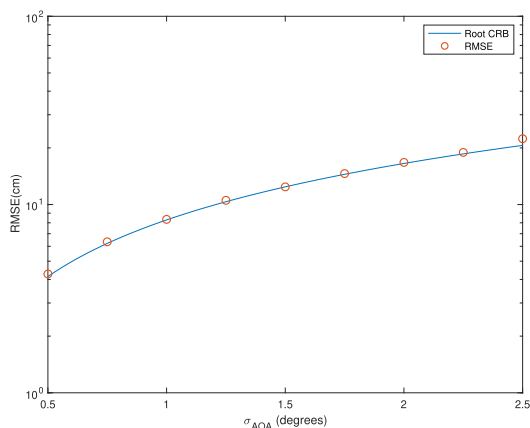


FIGURE 7. Simulation 2: RMSE dependency on σ_{AOA} .

with the center at the origin and lying on the XY plane. For each simulation, the calculated RMSE value is given by averaging results obtained for ten different rotations of the equilateral triangle around Z-axis.

Results of the first simulation are shown in Fig. 6 For a fixed parameters $\sigma_{AOA} = 0.5^\circ$ and $s = [160, 40, 40]^T$ cm, reported RMSE closely follows theoretical CRLB given increase in σ_{RD} . Fig. 7 illustrates the performance of the second simulation where σ_{AOA} varies from $[0.5^\circ - 2.5^\circ]$ while parameter σ_{RD} is fixed at 10 cm leaving source at position $s = [160, 40, 40]^T$ cm. The third simulation shows the performance evaluation depending on the distance from the source for the fixed $\sigma_{RD} = 10$ cm and $\sigma_{AOA} = 0.5^\circ$. For this case the source moves away from the origin in accordance with the expression $s = a[50, 50, 50]^T$ where a ranges from 1 to 5. From Fig. 8 it is clear that proposed method provides near CRLB performance.

B. EXPERIMENTAL EVALUATION

In this section, we present experimental results on sound source localization in 3D space. Following theoretical simulation given in subsection IV-A, three TetraMic soundfield microphones were placed 1 m apart forming vertices of an

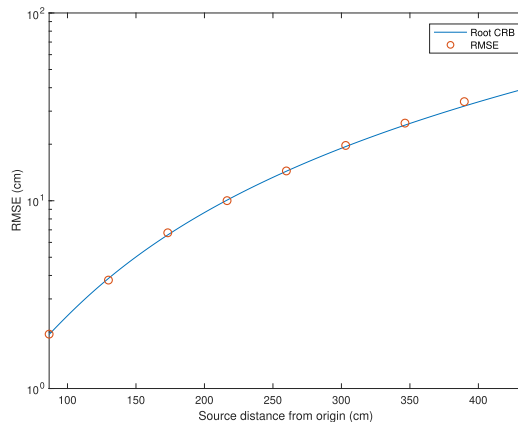


FIGURE 8. Simulation 3: RMSE dependency on the distance of the source.

equilateral triangle. Arta measurement software was used as the white noise generator, which was then converted to the analog audio signal by MOTU STAGE B-16 audio interface, amplified with Pioneer SA-8500 II sound amplifier and then played through Magnat 145 801 12 loudspeaker, which represented the source. The TetraMic is connected to audio interface via four XLR-M connectors coming from the PPA receiver. For each soundfield microphone, the four-channel signal is digitized with the B-16 audio interface and recorded with Reaper software in A-format. To obtain B-format from four-channel A-format in this work, we used an official software VVMic which also performs the corrections using the calibration file, provided with the specific microphone.

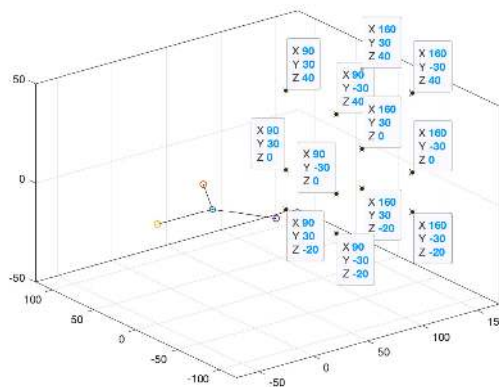


FIGURE 9. Measurement plan.

In the experiment we used 48 different sound source locations. Due to space limitations of our anechoic chamber all measurements were conducted regarding $s_x > 0$ by following measurement plan shown in Fig. 9, where after performing measurements for each 12 different sound source locations, an equilateral triangle holding TetraMic microphones was rotated around Z-axis for 90° .

For each source location s_i white noise signal of 18s duration was recorded. As concluded in the previous section, $t = 500$ ms interval was chosen for measurements' duration which resulted in 36 observations for each source position s_i .

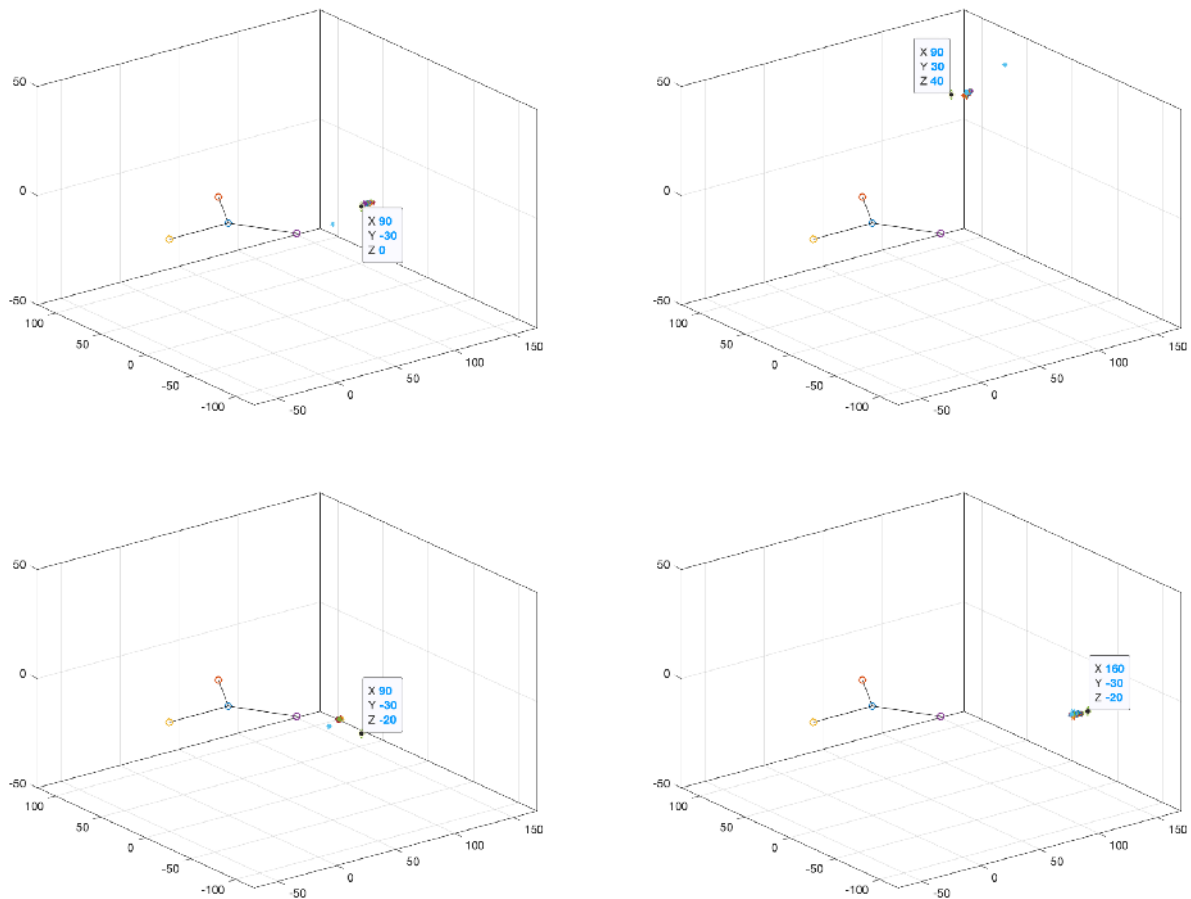


FIGURE 10. Sound source position estimation for 4 different locations in anechoic chamber.

Following the procedure explained in section III, for each segment we calculated two TDOAs from A-format and three AOA from B-format. Measured observations were first used in (17) to get the initial position estimate $\hat{\mathbf{s}}_1$ of the source which is needed to calculate the weight matrix \mathbf{W} . After obtaining an initial position estimate, method (17) was used 35 more times where for each position update, source estimation from the previous step was used to calculate \mathbf{W} which was used in conjunction with new TDOA and AOA observations to estimate new sound source position in 3D space. Since the computational step (17) is performed for each position update, it is interesting to analyze the efficiency of (17) in terms of computation time given its corresponding TDOA-AOA values. In this work, all processing procedures were carried out on the platform of Matlab 2019 on a 64-bit PC with the computational capability of Intel(R) Core(TM) i5 CPU @ 3.80 GHz and memory of 32.00 GB 400 MHz DDR4. In the average running time of our algorithm (17) for initial position estimation is 25.4ms and 3.9 ms for each additional position update.

To show the effectiveness of the proposed approach, Fig. 10 illustrates the estimation results for 4 different sound source locations, where reported RMSE results were 11.17,

15.56 18.90, and 10.47 cm respectively for each location. It is interesting to note that the estimate (shown in cyan) distant from the rest of the estimation group corresponds to the initial position guess, where the true source location is indicated by the coordinates label. Narrow estimation spread is clearly visible.

It was also interesting to analyze the dependency between reported RMSE and the number of iterations. From Fig. 11 it is visible that by using more iterations, (17) will provide better results. For longer measurement intervals, we can achieve even higher accuracy. The obtained result indicates that the proposed procedure can be used to locate the stationary sound source with excellent precision in the free-field. We plot the experimental cumulative distribution function (CDF) of positioning error RMSE in Fig. 12 to show the localization performance of the proposed weighted least square estimator (17). CDF in Fig. 12 is calculated for 48 source positions with a median value of 21.58 cm. Given a measurement interval of 18 sec, i.e., 36 estimations per point, the worst reported RMSE value reaches 38.90 cm while the best point has RMSE of 10.47 cm. From the CDF graph, it is visible that for 90% source positions, localization error falls within the margin of 28.38 cm.

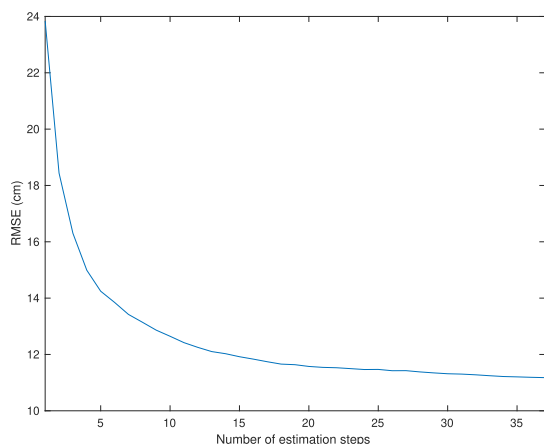


FIGURE 11. Dependency between reported RMSE and number of estimation steps.

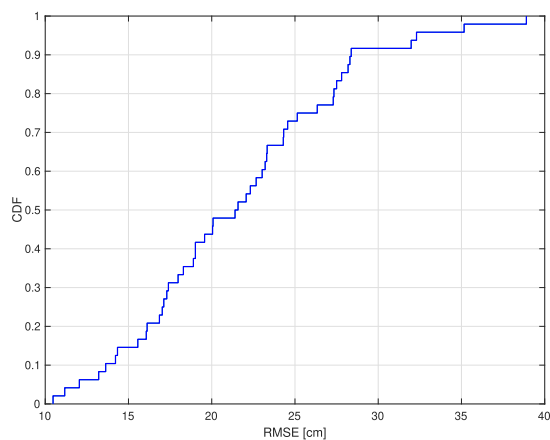


FIGURE 12. The CDF of the RMSE for 48 measurement points.

From the presented results, it is evident that the proposed measurement method provides accurate estimations of azimuth and elevation angles of the unknown sound source in the anechoic chamber. In particular, free field conditions play a major role in the acoustic measurements and sound perception experiments, as the results are influenced only by the direct-path component of the sound source and not by multipath component caused by room reflections. Given that the experimental evaluation in the SSL task was performed using real sound measurements, the above results indicate the potential for further research, where it can be expected that by employing the more robust method for estimating TDOA, such as method in [54], the results can translate to real-world cases. Although it is undeniable that the room acoustics would influence the SSL recognition results, conducting experimental evaluation in a controlled environment provides universal reference to localization performance. Since most state-of-the-art TDOA-based SSL either do not have experimental analysis (where results are given through simulation) or experimental evaluation is performed in non-reproducible environmental conditions, a direct comparison is often inaccessible. Regardless of the results obtained under certain conditions, our comparison with state-of-the-art is

made to demonstrate the performance of the proposed SSL approach. For instance, the state-of-the-art method in [17] reported a localization performance of 0.3m in terms of RMSE concerning the typical room experiment. Authors also compared themselves with commonly used SRP-HAT method, and they showed that their approach outperformed the SRP-HAT method, which reported a localization error of 0.76m. It is to expect that evaluation results obtained in the free-field will outperform results obtained in a real-life environment [17], and from comparison, it can be seen that the proposed TDOA-AOA approach, based on three soundfield microphones, achieves 0.09 m lower localization error compared to Multilevel B-Splines-Based Learning Approach and 0.55 m compared to SRP-HAT method [17]. Given the fact that the SSL results presented are 30% and 72% lower in terms of RMSE indicate the effectiveness of the proposed approach in free-filed with great potential to operate under realistic conditions.

V. CONCLUSION

This paper proposes a sound source localization method in 3D space using a geometric configuration of three microphone stations. The closed-form solution for estimating a sound source location based on two TDOAs and three AOAs is presented. In this work, a soundfield microphone is used as a measurement station. By exploiting A-format of the soundfield microphone, a pair-wise TDOA estimation using a general cross-correlation is obtained as the time lag between the signals from two microphones. The method for obtaining AOA measurements is proposed, based on a calculating directional vector derived from B-format. The investigation of the impact of the signal measurement interval length on TDOA and AOA estimation performance is made. The proposed method was evaluated by simulations and physical experiments in our anechoic chamber. The simulation results of the sound source localization reported that (17) reaches the theoretical CRLB performance regarding a small Gaussian noise presented in measurements. Since (17) can be viewed as iterative algorithm, an investigation between accuracy and number of iteration steps was made, concluding that with a longer recording of the analyzed signals from unknown source position, we can achieve even higher accuracy. The obtained experimental results indicate that the proposed procedure can be used to locate the stationary sound source with good precision in free-field environment. In our future work, it is planned to extend the proposed method for multiple source localization and tracking, first in free-field anechoic chamber conditions, and later in different SNR and reverberation conditions. Special efforts will be put on problems tackling real-life scenarios, like detecting speech in noisy environments when distant recording is required, in order to increase speech recognition performance.

REFERENCES

- [1] X. Li and H. Liu, "Sound source localization for HRI using FOC-based time difference feature and spatial grid matching," *IEEE Trans. Cybern.*, vol. 43, no. 4, pp. 1199–1212, Aug. 2013.

- [2] Q. Yan, J. Chen, G. Ottoy, and L. De Strycker, "Robust AOA based acoustic source localization method with unreliable measurements," *Signal Process.*, vol. 152, pp. 13–21, Nov. 2018.
- [3] D. GalaNathan Lindsay, N. Lindsay, and L. Sun, "Three-dimensional sound source localization for unmanned ground vehicles with a self-rotational two-microphone array," in *Proc. 5th Int. Conf. Control, Dyn. Syst., Robot. (CDSR)*, Niagara Falls, ON, Canada, Jun. 2018, pp. 7–9.
- [4] C. Rascon and I. Meza, "Localization of sound sources in robotics: A review," *Robot. Auto. Syst.*, vol. 96, pp. 184–210, Oct. 2017.
- [5] L. Wang and A. Cavallaro, "Acoustic sensing from a multi-rotor drone," *IEEE Sensors J.*, vol. 18, no. 11, pp. 4570–4582, Jun. 2018.
- [6] T. Damarla, L. M. Kaplan, and G. T. Whipps, "Sniper localization using acoustic asynchronous sensors," *IEEE Sensors J.*, vol. 10, no. 9, pp. 1469–1478, Sep. 2010.
- [7] R. Blumrich and J. Altmann, "Medium-range localisation of aircraft via triangulation," *Appl. Acoust.*, vol. 61, no. 1, pp. 65–82, Sep. 2000.
- [8] W. Ma, H. Bao, C. Zhang, and X. Liu, "Beamforming of phased microphone array for rotating sound source localization," *Journal Sound Vib.*, vol. 467, Feb. 2020, Art. no. 115064.
- [9] L. Battista, E. Schena, G. Schiavone, S. A. Sciuto, and S. Silvestri, "Calibration and uncertainty evaluation using Monte Carlo method of a simple 2D sound localization system," *IEEE Sensors J.*, vol. 13, no. 9, pp. 3312–3318, Sep. 2013.
- [10] C. Pang, H. Liu, J. Zhang, and X. Li, "Binaural sound localization based on reverberation weighting and generalized parametric mapping," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 8, pp. 1618–1632, Aug. 2017.
- [11] X. Zhong, L. Sun, and W. Yost, "Active binaural localization of multiple sound sources," *Robot. Auto. Syst.*, vol. 85, pp. 83–92, Nov. 2016.
- [12] H. He, L. Wu, J. Lu, X. Qiu, and J. Chen, "Time difference of arrival estimation exploiting multichannel spatio-temporal prediction," *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 3, pp. 463–475, Mar. 2013.
- [13] D. Pavlidi, A. Griffin, M. Puigt, and A. Mouchtaris, "Real-time multiple sound source localization and counting using a circular microphone array," *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 10, pp. 2193–2206, Oct. 2013.
- [14] A. Canclini, F. Antonacci, A. Sarti, and S. Tubaro, "Acoustic source localization with distributed asynchronous microphone networks," *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 2, pp. 439–443, Feb. 2013.
- [15] K. Noh and J.-H. Chang, "Joint optimization of deep neural network-based dereverberation and beamforming for sound event detection in multi-channel environments," *Sensors*, vol. 20, no. 7, p. 1883, Apr. 2020.
- [16] A. Canclini, P. Bestagini, F. Antonacci, M. Compagnoni, A. Sarti, and S. Tubaro, "A robust and low-complexity source localization algorithm for asynchronous distributed microphone networks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 10, pp. 1563–1575, Oct. 2015.
- [17] L. Nguyen, J. Valls Miro, and X. Qiu, "Multilevel B-splines-based learning approach for sound source localization," *IEEE Sensors J.*, vol. 19, no. 10, pp. 3871–3881, May 2019.
- [18] K. C. Ho, "Bias reduction for an explicit solution of source localization using TDOA," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2101–2114, May 2012.
- [19] L. Yang and K. C. Ho, "An approximately efficient TDOA localization algorithm in closed-form for locating multiple disjoint sources with erroneous sensor positions," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4598–4615, Dec. 2009.
- [20] G. Wang and K. C. Ho, "Convex relaxation methods for unified near-field and far-field TDOA-based localization," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2346–2360, Apr. 2019.
- [21] K. C. Ho and W. Xu, "An accurate algebraic solution for moving source location using TDOA and FDOA measurements," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2453–2463, Sep. 2004.
- [22] K. Dogancay and G. Ibal, "Instrumental variable estimator for 3D bearings-only emitter localization," in *Proc. Int. Conf. Intell. Sensors, Sensor Netw. Inf. Process.*, 2005, pp. 63–68.
- [23] A. Yassin, Y. Nasser, M. Awad, A. Al-Dubai, R. Liu, C. Yuen, R. Raulefs, and E. Aboutanios, "Recent advances in indoor localization: A survey on theoretical approaches and applications," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1327–1346, 2nd Quart., 2017.
- [24] L. Cong and W. Zhuang, "Hybrid TDOA/AOA mobile user location for wideband CDMA cellular systems," *IEEE Trans. Wireless Commun.*, vol. 1, no. 3, pp. 439–447, Jul. 2002.
- [25] A. N. Bishop, B. Fidan, K. Do ançay, B. D. O. Anderson, and P. N. Pathirana, "Exploiting geometry for improved hybrid AOA/TDOA-based localization," *Signal Process.*, vol. 88, no. 7, pp. 1775–1791, Jul. 2008.
- [26] J. Yin, Q. Wan, S. Yang, and K. C. Ho, "A simple and accurate TDOA-AOA localization method using two stations," *IEEE Signal Process. Lett.*, vol. 23, no. 1, pp. 144–148, Jan. 2016.
- [27] T. Jia, H. Wang, X. Shen, Z. Jiang, and K. He, "Target localization based on structured total least squares with hybrid TDOA-AOA measurements," *Signal Process.*, vol. 143, pp. 211–221, Feb. 2018.
- [28] K. Dogancay, "3D pseudolinear target motion analysis from angle measurements," *IEEE Trans. Signal Process.*, vol. 63, no. 6, pp. 1570–1580, Mar. 2015.
- [29] C. Zhou, G. Huang, H. Shan, and J. Gao, "Bias compensation algorithm based on maximum likelihood estimation for passive localization using TDOA and FDOA measurements," *Acta Aeronaut. Astronaut. Sin.*, vol. 36, pp. 979–986, Mar. 2015. [Online]. Available: http://hkxb.buaa.edu.cn/EN/volumn/volumn_1443.shtml
- [30] X. Qu, L. Xie, and W. Tan, "Iterative constrained weighted least squares source localization using TDOA and FDOA measurements," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 3990–4003, Aug. 2017.
- [31] Z. Liu, R. Wang, and Y. Zhao, "A bias compensation method for distributed moving source localization using TDOA and FDOA with sensor location errors," *Sensors*, vol. 18, no. 11, p. 3747, Nov. 2018.
- [32] M. Sun, L. Yang, and D. K. C. Ho, "Efficient joint source and sensor localization in closed-form," *IEEE Signal Process. Lett.*, vol. 19, no. 7, pp. 399–402, Jul. 2012.
- [33] 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, May 2008.
- [34] Y. Wang and K. C. Ho, "An asymptotically efficient estimator in closed-form for 3-D AOA localization using a sensor network," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6524–6535, Dec. 2015.
- [35] L. Badriasi, H. Kennedy, and A. Finn, "Effects of coordinate system rotation on two novel closed-form localization estimators using azimuth/elevation," in *Proc. 16th Int. Conf. Inf. Fusion*, Jul. 2013, pp. 1797–1804.
- [36] Y. Zhao, Z. Li, B. Hao, J. Si, and P. Wan, "Bias reduced method for TDOA and AOA localization in the presence of sensor errors," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.
- [37] K. Dogancay, "Relationship between geometric translations and TLS estimation bias in bearings-only target localization," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1005–1017, Mar. 2008.
- [38] J. Li, Y. Zhao, and D. Li, "Accurate single-observer passive coherent location estimation based on TDOA and DOA," *Chin. J. Aeronaut.*, vol. 27, no. 4, pp. 913–923, Aug. 2014.
- [39] S. Chen, J. Zhang, and C. Xu, "Robust distributed cooperative localization with NLOS mitigation based on multiplicative convex model," *IEEE Access*, vol. 7, pp. 112907–112920, 2019.
- [40] H. Chen, G. Wang, and X. Wu, "Cooperative multiple target nodes localization using TOA in mixed LOS/NLOS environments," *IEEE Sensors J.*, vol. 20, no. 3, pp. 1473–1484, Feb. 2020.
- [41] H. Lohrasbipeydeh and T. A. Gulliver, "Unknown RSSD-based localization CRLB analysis with semidefinite programming," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3791–3805, May 2019.
- [42] Q. Qi, Y. Li, Y. Wu, Y. Wang, Y. Yue, and X. Wang, "RSS-AOA-based localization via mixed semi-definite and second-order cone relaxation in 3-D wireless sensor networks," *IEEE Access*, vol. 7, pp. 117768–117779, 2019.
- [43] C. Zhang and Y. Wang, "Sensor network event localization via nonconvex nonsmooth ADMM and augmented lagrangian methods," *IEEE Trans. Control Netw. Syst.*, vol. 6, no. 4, pp. 1473–1485, Dec. 2019.
- [44] T. Padois, O. Doutres, and F. Sgard, "On the use of modified phase transform weighting functions for acoustic imaging with the generalized cross correlation," *The J. Acoust. Soc. Amer.*, vol. 145, no. 3, pp. 1546–1555, Mar. 2019.
- [45] T. Padois, O. Doutres, F. Sgard, and A. Berry, "Optimization of a spherical microphone array geometry for localizing acoustic sources using the generalized cross-correlation technique," *Mech. Syst. Signal Process.*, vol. 132, pp. 546–559, Oct. 2019.
- [46] T. Padois, "Acoustic source localization based on the generalized cross-correlation and the generalized mean with few microphones," *J. Acoust. Soc. Amer.*, vol. 143, no. 5, pp. EL393–EL398, May 2018.

- [47] A. K. Zad Tehrani, B. Makkiabadi, A. Pourmohammad, and S. H. Hozhabr, "Sound source localization using time differences of arrival; Euclidean distance matrices based approach," in *Proc. 9th Int. Symp. Telecommun. (IST)*, Dec. 2018, pp. 91–95.
- [48] J. Vera-Diaz, D. Pizarro, and J. Macias-Guarasa, "Towards end-to-end acoustic localization using deep learning: From audio signals to source position coordinates," *Sensors*, vol. 18, no. 10, p. 3418, Oct. 2018.
- [49] J. Ding, B. Ren, and N. Zheng, "Microphone array acoustic source localization system based on deep learning," in *Proc. 11th Int. Symp. Chin. Spoken Lang. Process. (ISCSLP)*, Nov. 2018, pp. 409–413.
- [50] T. Wu, Y. Jiang, N. Li, and T. Feng, "An indoor sound source localization dataset for machine learning," in *Proc. 2nd Int. Conf. Comput. Sci. Artif. Intell. (CSAI)*, 2018, pp. 28–32.
- [51] Y. Sun, J. Chen, C. Yuen, and S. Rahardja, "Indoor sound source localization with probabilistic neural network," *IEEE Trans. Ind. Electron.*, vol. 65, no. 8, pp. 6403–6413, Aug. 2018.
- [52] Y. Liu, L. Wang, X. Zeng, and H. Wang, "Sound source localization in reverberant environments based on structural sparse Bayesian learning," *Acta Acustica United Acustica*, vol. 104, no. 3, pp. 528–541, May 2018.
- [53] J. Kotus, "Multiple sound sources localization in free field using acoustic vector sensor," *Multimedia Tools Appl.*, vol. 74, no. 12, pp. 4235–4251, Jun. 2015.
- [54] N. Zhu and T. Reza, "A modified cross-correlation algorithm to achieve the time difference of arrival in sound source localization," *Meas. Control*, vol. 52, nos. 3–4, pp. 212–221, Mar. 2019.
- [55] H. Ding, Y. Bao, Q. Huang, C. Li, and G. Chai, "Three-dimensional localization of point acoustic sources using a planar microphone array combined with beamforming," *Roy. Soc. Open Sci.*, vol. 5, no. 12, Dec. 2018, Art. no. 181407.
- [56] M. Russo, L. Kraljević, M. Stella, and M. Sikora, "Acoustic performance analysis of anechoic chambers based on ISO 3745 and ISO 26101: Standards comparison and performance analysis of the anechoic chamber at the University of Split," in *Proc. Euronoise*, 2018, pp. 1–6.
- [57] S. M. Kay, *Fundamentals of statistical Signal Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.



LUKA KRALJEVIĆ (Member, IEEE) received the M.S. degree in electronics and computer engineering from the University of Split, Croatia, in 2014, where he is currently pursuing the Ph.D. degree on a project funded by Croatian Science Foundation project. He has been with the Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture (FESB), University of Split, since 2016. His research interests include deep learning, signal processing, and advanced human–computer interfaces, such as speech recognition, gesture recognition, and affective computing.



MLADEN RUSSO (Member, IEEE) received the B.S. degree (*cum laude*) in electrical engineering and the M.S. and Ph.D. degrees from the University of Split, Croatia, in 2001, 2006, and 2010, respectively. He has been with the Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture (FESB), University of Split, since 2001. He is currently involved as Associate Professor and is leading a Research Team and Laboratory on smart environment technologies. He has participated and led several scientific and professional research projects. His research interests include signal processing and machine learning technologies, affective computing, human–computer interfaces (speech recognition, emotion recognition, and virtual and augmented reality technologies), wireless positioning, and ambient energy harvesting. He is a member of CCIS and an Associate Editor of the *Journal of Communications Software and Systems*.



MAJA STELLA (Member, IEEE) received the B.S. degree (*cum laude*) in electrical engineering and the M.S. and Ph.D. degrees from the University of Split, Croatia, in 1999, 2007, and 2011, respectively. She has been with the Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture (FESB), University of Split, since 2001. She is currently involved as an Assistant Professor. Her research interests include wireless positioning, signal processing, pattern recognition, affective computing, and ambient energy harvesting. She is a member of CCIS.



MARJAN SIKORA (Member, IEEE) received the B.S. degree in electrical engineering and the M.S. and Ph.D. degrees from the University of Zagreb, Croatia, in 1995, 2000, and 2010, respectively. He has been with the Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture (FESB), University of Split, since 2006. He is currently involved as an Associate Professor and is leading the Laboratory of Electroacoustics. He has participated in several scientific and professional projects. His research interests include programming, programming languages, electroacoustics, geographic information systems, and simulations and virtual and augmented reality technologies. He is a member of the Association for Computing Machinery, National Geographic Society, and Croatian Acoustical Society.

...