Realtime Active Sound Source Localization for Unmanned Ground Robots Using a Self-Rotational Bi-Microphone Array

Deepak Gala, Nathan Lindsay, Liang Sun

This work presents a novel technique that performs both orientation and distance localization of a sound source in a three-dimensional (3D) space using only the interaural time difference (ITD) cue, generated by a newly-developed self-rotational bi-microphone robotic platform. The system dynamics is established in the spherical coordinate frame using a state-space model. The observability analysis of the state-space model shows that the system is unobservable when the sound source is placed with elevation angles of 90° and 0°. The proposed method utilizes the difference between the azimuth estimates resulting from respectively the 3D and the two-dimensional models to check the zero-degree-elevation condition and further estimates the elevation angle using a polynomial curve fitting approach. Also, the proposed method is capable of detecting a 90°-degree elevation by extracting the zero-ITD signal 'buried' in noise. Additionally, a distance localization is performed by first rotating the microphone array to face toward the sound source and then shifting the microphone perpendicular to the source-robot vector by a predefined distance of a fixed number of steps. The integrated rotational and translational motions of the microphone array provide a complete orientation and distance localization using only the ITD cue. A novel robotic platform using a self-rotational bi-microphone array was also developed for unmanned ground robots performing sound source localization. The proposed technique was first tested in simulation and was then verified on the newly-developed robotic platform. Experimental data collected by the microphones installed on a KEMAR dummy head were also used to test the proposed technique. All results show the effectiveness of the proposed technique.

1 INTRODUCTION

The localization problem in the robotic field has been recognized as the most fundamental problem to make robots truly autonomous [10]. Localization techniques are of great importance for autonomous unmanned systems to identify their own locations (i.e., self-localization) and situational awareness (e.g., locations of surrounding objects), especially in an unknown environment. Mainstream technology for localization is based on computer vision, supported by visual sensors (e.g., cameras), which, however, are subject to lighting and line-of-sight conditions and rely on computationally demanding image-processing algorithms. An acoustic sensor (e.g., a microphone), as a complementary component in a robotic sensing system, does not require a line of sight and is able to work under varying light (or completely dark) conditions in an omnidirectional manner. Thanks to the advancement of microelectromechanical technology, microphones become inexpensive and do not require significant power to operate.

Sound-source localization (SSL) techniques have been developed that identify the location of sound sources (e.g., speech and music) in terms of directions and distances. SSL techniques have been widely used in civilian applications, such as intelligent video conferencing [26], [52], environmental monitoring [19], human-robot interaction (HRI) for humanoid robotics [25], and robot motion planning [37], as well as military applications, such as passive sonar for submarine detections, surveillance systems that locate hostile tanks, artillery, incoming
SSL has been achieved by using microphone arrays with more than two microphones. The accuracy of the localization techniques based on microphone arrays is dictated by their physical sizes. Microphone arrays are usually designed using particular (e.g., linear or circular) structures, which result in their relatively large sizes and sophisticated control components for operation. Therefore, it becomes difficult to use them on small robots nor large systems due to the complexity of mounting and maneuvering.

In the past decade, research has been carried out for robots to have auditory behaviors (e.g. getting attention to an event, locating a sound source in potentially dangerous situations, and locating and paying attention to a speaker) by mimicking human auditory systems. Humans perform sound localization with their two ears using integrated three types of cues, i.e., the interaural level difference (ILD), the interaural time difference (ITD), and the spectral information. ILD and ITD cues are usually used respectively to identify the horizontal location (i.e., azimuth angle) of a sound source with higher and lower frequencies. Spectral cues are usually used to identify the vertical location (i.e., elevation angle) of a sound source with higher frequencies. Additionally, acoustic landmarks aid towards bettering the SSL by humans.

To mimic human acoustic systems, researchers have developed sound source localization techniques using only two microphones. All three types of cues have been used by Rodemann et al. in a binaural approach of estimating the azimuth angle of a sound source, while the authors also stated that reliable elevation estimation would need a third microphone. Spectral cues were used by the head-related-transfer-function (HRTF) that was applied to identify both the azimuth and elevation angles of a sound source for binaural sensor platforms. The ITD cues have also been used in binaural sound source localization, where the problem of cone of confusion has been overcome by incorporating head movements, which also enable both azimuth and elevation estimation. Lu et al. used a particle filter for binaural tracking of a mobile sound source on the basis of ITD and motion parallax but the localization was limited in a two-dimensional (2D) plane and was not impressive under static conditions. Pang et al. presented an approach for binaural azimuth estimation based on reverberation weighting and generalized parametric mapping. Lu et al. presented a binaural distance localization approach using the motion-induced rate of intensity change which requires the use of parallax motion and errors up to 3.4 m were observed. Kneip and Baumann established formulae for binaural identification of the azimuth and elevation angles as well as the distance information of a sound source combining the rotational and translational motion of the interaural axis. However, large localization errors were observed and no solution was given to handle sensor noise nor model uncertainty. Rodemann proposed a binaural azimuth and distance localization technique using signal amplitude along with ITD and ILD cues in an indoor environment with a sound source ranging from 0.5 m to 6 m. However, the azimuth estimation degrades with the distance and reduced error with the required calibration was still large.

To the best of our knowledge, the works presented in the literature for SSL using two microphones based on ITD cues mainly provided formulae that calculate the azimuth and elevation angles of a sound source without incorporating sensor noise. The works that use probabilistic recursive filtering techniques (e.g., EKF) for orientation estimation did not conduct any observability analysis on the system dynamics. In other words, no discussion on the limitation of the techniques for orientation estimation was found. In addition, no probabilistic recursive filtering technique was used to acquire distance information of a sound source. This paper aims to address these research gaps.

The contributions of this paper include (1) an observability analysis of the system dynamics for three-dimensional (3D) SSL using two microphones and the ITD cue only; (2) a novel algorithm that provides the estimation of the elevation angle of a sound source when the states are unobservable; and (3) a new EKF-based technique that estimates the robot-sound distance. Both simulations and experiments were conducted to validate the proposed techniques.

The rest of this paper is organized as follows. Section describes the preliminaries. In Section 2D and 3D orientation localization models are presented along with their observability analysis. In Section 4 a novel...
method is proposed to detect non-observability conditions and a solution to the non-observability problem is presented. Section 5 presents a distance localization model with its observability analysis. The EKF algorithm is presented in Section 6. In Sections 7 and 8, the simulation and experimental results are presented respectively, followed by Section 9 which concludes the paper.

2 PRELIMINARIES

2.1 Calculation of ITD

The only cue used for localization in this paper is the ITD, which is the time difference of a sound signal traveling to the two microphones and can be calculated using the cross-correlation technique [30].

Consider a single stationary sound source placed in an environment. Let \( y_1(t) \) and \( y_2(t) \) be the sound signals captured by two spatially separated microphones in the presence of noise, which are given by [30]

\[
y_1(t) = s(t) + n_1(t),
\]
\[
y_2(t) = \delta \cdot s(t + t_d) + n_2(t),
\]

where \( s(t) \) is the sound signal, \( n_1(t) \) and \( n_2(t) \) are real and jointly stationary random processes, \( t_d \) denotes the time difference of \( s(t) \) arriving at the two microphones, and \( \delta \) is the signal attenuation factor due to different traveling distances of the sound signal to the two microphones. It is commonly assumed that \( \delta \) changes slowly and \( s(t) \) is uncorrelated with noises \( n_1(t) \) and \( n_2(t) \) [30]. The cross-correlation function of \( y_1(t) \) and \( y_2(t) \) is given by

\[
R_{y_1,y_2}(\tau) = E[y_1(t) \cdot y_2(t - \tau)],
\]

where \( E[\cdot] \) represents the expectation operator. Figure 1 shows the process of delay estimation between \( y_1(t) \) and \( y_2(t) \), where \( H_1(f) \) and \( H_2(f) \) represent scaling functions or pre-filters [30]. Various techniques can be used to eliminate or reduce the effect of background noise and reverberations [8, 9, 18, 19, 36, 44]. An improved version of the cross-correlation method incorporating \( H_1(f) \) and \( H_2(f) \) is called Generalized Cross-Correlation (GCC) [30], which further improves the estimation of time delay.

The time difference of \( y_1(t) \) and \( y_2(t) \), i.e., the ITD, is given by \( \hat{T} = \arg \max_x R_{y_1,y_2} \). The distance difference of the sound signal traveling to the two microphones is given by \( d = \hat{T} \cdot c_0 \), where \( c_0 \) is the sound speed and is usually selected to be 345 m/s.

2.2 Far-Field Assumption

The area around a sound source can be divided into five different fields: free field, near field, far field, direct field and reverberant field [1, 21]. The region close to a source where the sound pressure and the acoustic particle velocity are not in phase is regarded as the near field. The range of the near field is limited to a distance from the source equal to approximately a wavelength of sound or equal to three times the largest dimension of the sound source, whichever is the larger. The far field of a source begins where the near field ends and extends to infinity. Under the far-field assumption, the acoustic wavefront reaching the microphones is planar and not spherical, in the sense that the waves travel in parallel i.e. the angle of incidence is the same for the two microphones [14].

2.3 Observability Analysis

Consider a nonlinear system described by a state-space model

\[
\dot{x} = f(x),
\]
\[
y = h(x),
\]

where \( x \in \mathbb{R}^n \) and \( y \in \mathbb{R}^m \) are the state and output vectors, respectively, and \( f(\cdot) \) and \( h(\cdot) \) are the process and output functions, respectively. The observability matrix of the system described by (3) and (4) is then given by [23]

\[
\Omega = \left[ \left( \frac{\partial L^0_y h}{\partial x} \right)^T \left( \frac{\partial L^1_y h}{\partial x} \right)^T \cdots \right]^T,
\]

where the Lie derivatives are given by \( L^0_y h = h(x) \) and \( L^1_y h = \frac{\partial L^0_y h}{\partial x} \cdot f \). The system is observable if the observability matrix \( \Omega \) has rank \( n \).

3 Mathematical Models and Observability Analysis for Orientation Localization

The complete localization of a sound source is usually achieved in two stages, the orientation (i.e., azimuth and elevation angles) localization and distance localization. In this section, the methodology of the orientation localization is presented.
As shown in Figures 2 and 3, the acoustic signal generated by the sound source $S$ is collected by the left and right microphones, $L$ and $R$, respectively. Let $O$ be the center of the robot as well as the two microphones. The location of $S$ is represented by $(D, \theta, \phi)$, where $D$ is the distance between the source and the center of the robot, i.e., the length of segment $OS$, $\theta \in [0, \frac{\pi}{2}]$ is the elevation angle defined as the angle between $OS$ and the horizontal plane, and $\phi \in (-\pi, \pi]$ is the azimuth angle defined as the angle measured clockwise from the robot heading vector, $p$, to $OS$. Letting unit vector $q$ be the orientation (heading) of the microphone array, $\beta$ be the angle between $p$ and $q$, and $\psi$ be the angle between $q$ and $OS$, both following a right hand rotation rule, we have

$$\phi = \psi + \beta.$$  \hspace{1cm} (5)

For a clockwise rotation, we have $\beta(t) = \omega t$, where $\omega$ is the rotational speed of the two microphones, and $\psi(t) = \psi - \omega t$. In the shaded triangle, $\triangle SOF$, shown in Figures 3 and 4, define $\alpha \triangleq \angle SOF$ and we have $\alpha + \psi = \frac{\pi}{2}$ and $\cos \alpha = \cos \theta \sin \psi$. Based on the far-field assumption in Section 2.2, we have

$$d = \hat{T} \cdot c_0 = b \cos \alpha = b \cos \theta \sin \psi.$$ \hspace{1cm} (6)

where $b$ is the distance between the two microphones, i.e., the length of the segment $LR$.

To avoid cone of confusion [51] in SSL, the two-microphone array is rotated with a nonzero angular velocity [54]. Without loss of generality, in this paper we assume a clockwise rotation of the microphone array on the horizontal plane while the robot itself does not rotate nor translate throughout the entire estimation process, which implies that $\psi$ is constant.

### 3.2 2D Localization

If the sound source and the robot are on the same horizontal plane, i.e., $\theta = 0$, we have $d = b \sin \psi$. Assume that the microphone array rotates clockwise with a constant angular velocity, $\omega$. Considering the state-space model for 2D localization with the state $x_{2D} \equiv \psi$, and the output as $y_{2D} \equiv d$, we have

$$\dot{x}_{2D} = \dot{\psi} = -\omega,$$ \hspace{1cm} (7)

$$y_{2D} = b \sin \psi.$$ \hspace{1cm} (8)

**Theorem 1** The system described by Equations (7) and (8) is observable if 1) $b \neq 0$ and 2) $\omega \neq 0$ or $\psi \neq 2k\pi \pm \frac{\pi}{2}$, where $k \in \mathbb{Z}$.

**Proof** The observability matrix [23, 24] for the system described by Equations (7) and (8) is given by

$$O_{2D} = \begin{bmatrix} b \cos \psi & b \omega \sin \psi & -b\omega^2 \cos \psi & \cdots \end{bmatrix}^T.$$ \hspace{1cm} (9)

The system is observable if $O_{2D}$ has rank one, which implies $b \neq 0$. If $\omega = 0$, observability requires that $\cos \psi \neq 0$, which implies $\psi \neq 2k\pi \pm \frac{\pi}{2}$. If $\omega \neq 0$, $O_{2D}$ is full rank for all $\psi$.

**Remark 1** Since the two microphones are separated by a non-zero distance, (i.e., $b \neq 0$) and the microphone array rotates with a non-zero constant angular velocity (i.e., $\omega \neq 0$), the system is observable in the domain of definition.
3.3 3D Localization

Considering the state-space model for 3D localization with the state $x_{3D} \triangleq [\theta, \psi]^T$, and the output as $y_{3D} \triangleq d$, we have

$$
\dot{x}_{3D} = \begin{bmatrix} \dot{\theta} \\ \dot{\psi} \end{bmatrix} = \begin{bmatrix} 0 \\ -\omega \end{bmatrix},
$$
\hspace{1cm}
\hspace{1cm}
\hspace{1cm}
(10)

$$
y_{3D} = b \cos \theta \sin \psi.
$$
\hspace{1cm}
\hspace{1cm}
\hspace{1cm}
(11)

**Theorem 2** The system described by Equations (10) and (11) is observable if 1) $b \neq 0$, 2) $\omega \neq 0$, and 3) $\theta \neq 90^\circ$.

*Proof* The observability matrix for (10) and (11) is given by

$$
O_{3D} = \begin{bmatrix}
-b \sin \theta \sin \psi & b \cos \theta \cos \psi \\
-b \omega \sin \theta \cos \psi & b \omega \cos \theta \sin \psi \\
-b \omega^2 \sin \theta \sin \psi & -b \omega^2 \cos \theta \cos \psi \\
-b \omega^3 \sin \theta \cos \psi & -b \omega^3 \cos \theta \sin \psi
\end{bmatrix}.
$$
\hspace{1cm}
\hspace{1cm}
\hspace{1cm}
(12)

The system is observable if

$$
\det \{O_{3D}\} = -b^2 \omega \sin \theta \cos \theta.
$$

It should be noted that higher-order Lie derivatives do not add rank to $O_{3D}$. Consider the squared matrix consisting of the first two rows of $O_{3D}$

$$
O_{3D} = \begin{bmatrix}
-b \sin \theta \sin \psi & b \cos \theta \cos \psi \\
-b \omega \sin \theta \cos \psi & b \omega \cos \theta \sin \psi
\end{bmatrix},
$$
\hspace{1cm}
\hspace{1cm}
\hspace{1cm}
and the determinant of the $O_{3D}$ is

$$
\det \{O_{3D}\} = -b^2 \omega \sin \theta \cos \theta.
$$

Further investigation can be done by selecting two even (or odd) rows from $O_{3D}$ to form a squared matrix, whose determinant is always zero.

**Remark 2** As it is always true that $b \neq 0$ and $\omega \neq 0$ due to Remark 1, the system is observable only when $\theta \neq 0^\circ$ and $\theta \neq 90^\circ$. Experimental results presented by Zhong et al. [54] using a similar model illustrates large estimation error when $\theta$ is close to zero.

To further investigate the system observability, consider the following two special cases: (1) $\theta$ is known and (2) $\psi$ is known.

Assume that $\theta$ is known and consider the following system

$$
\dot{x}_\theta = \dot{\theta} = 0,
$$
\hspace{1cm}
\hspace{1cm}
\hspace{1cm}
(13)

$$
y_\theta = b \cos \theta \sin \psi.
$$
\hspace{1cm}
\hspace{1cm}
\hspace{1cm}
(14)

**Corollary 1** The azimuth angle in the system described by Equations (14) and (15) is observable if 1) $b \neq 0$, 2) $\omega \neq 0$, and 3) $\theta \neq 90^\circ$.

*Proof* The observability matrix associated with (14) and (15) is given by

$$
O_\psi = [b \cos \theta \cos \psi b \omega \cos \theta \sin \psi \cdots]^T.
$$
\hspace{1cm}
\hspace{1cm}
\hspace{1cm}
(15)

This shows that $\psi$ is unobservable when $\theta = 90^\circ$.

Assume that $\psi$ is known and consider the following system

$$
\dot{x}_\psi = \dot{\psi} = -\omega,
$$
\hspace{1cm}
\hspace{1cm}
\hspace{1cm}
(16)

$$
y_\psi = b \cos \theta \sin \psi.
$$
\hspace{1cm}
\hspace{1cm}
\hspace{1cm}
(17)

**Corollary 2** The elevation angle in the system described by Equations (18) and (19) is observable if the following conditions are satisfied: 1) $b \neq 0$, 2) $\omega \neq 0$, and 3) $\theta \neq 0^\circ$.

*Proof* The observability matrix associated with (18) and (19) is given by

$$
O_\theta = [-b \sin \theta \sin \psi 0 \cdots]^T.
$$
\hspace{1cm}
\hspace{1cm}
\hspace{1cm}
(18)

So the system is observable if

$$
b \neq 0, \theta \neq 0^\circ, \text{ and } \psi \neq k \pi.
$$
\hspace{1cm}
\hspace{1cm}
\hspace{1cm}
(19)

As $\psi$ is time-varying, so it won’t stay at $k \pi$. It can be seen that $\theta$ is unobservable when $\theta = 0^\circ$.

4 Complete Orientation Localization

To handle the unobservable situations, i.e., $\theta = 0^\circ$ and $\theta = 90^\circ$, we present a novel algorithm in this section that utilizes both the 2D and 3D localization models to enable the orientation localization of a sound source residing anywhere in the domain of definition, i.e., $\theta \in [0, \pi/2]$ and $\varphi \in (-\pi, \pi]$. 


served when the frequency is at ±85°. Figure 5 shows the resulting magnitude (DFT) of the noised signal. The signals after taking the Discrete Fourier transform (DFT) of the noised signal d(t) with the source located at θ = 50° and θ = 85°, respectively. The two big peaks in the upper figure occur at ±2π/5 rad/sec (i.e., the angular velocity of the rotation of the microphone array) when θ = 50°, whereas small peaks are present in the bottom figure when θ = 85°.

The signals after taking the Discrete Fourier transform (DFT) of the noised signal d(t) with the source located at θ = 50° and θ = 85°, respectively. The two big peaks in the upper figure occur at ±2π/5 rad/sec (i.e., the angular velocity of the rotation of the microphone array) when θ = 50°, whereas small peaks are present in the bottom figure when θ = 85°.

|Frequency (rad/s)| Magnitude (m) |
|-----------------|--------------|
| -250 to 250     | 0.01         |
| -200 to 200     | 0.02         |
| -150 to 150     | 0.03         |

The ITD is considered as zero if the maximum value of the estimated amplitude ̂A_d (when the frequency equals the angular velocity of the rotation of the microphone array) is less than a predefined threshold, d_threshold. The selection of d_threshold determines the accuracy of the estimation when the sound source is around 90° elevation. The value of d_threshold, for example, can be selected as 0.017 m, which corresponds to θ = 85° as in Figure 5, thereby giving an accuracy of 5°.

4.2 Identification of θ = 0°

Theorem 1 guarantees accurate azimuth angle estimation using the 2D model when the sound source is located with zero elevation. We observed that when the elevation of the sound source is not close to zero, the estimation of the azimuth angle provided by the 2D model is far off the real value.

On the other hand, Theorem 2 guarantees that the azimuth angle estimation using the 3D model is accurate for all elevation angles except for θ = 90°, which is detected by the approach in Section 4.1. Therefore, the estimations resulting from both the 2D model 3D models will be identical if the sound source is located at θ = 0°, as shown in Figure 7.
error (RMSE) is used as a measure of the difference between the two azimuth estimations as it includes both mean absolute error (MAE) as well as additional information related to the variance [13]. This error is dependent on the value of elevation angle and it increases as the elevation angle increases, as shown in Figure 8.

![Fig. 8 RMSE between 2D and 3D localization model's estimated azimuth angles.](image)

In order to get an accurate estimate of the elevation angle close to zero, a polynomial curve fitting approach is used to map (in a least-square sense) the RMSE values to the elevation angles. Different RMSE values are collected beforehand in the environment where the localization would be done. The RMSE values associated with the same elevation angle but different azimuth angles express small variations, as seen in Figure 8. Therefore, for a particular elevation angle, the mean of all RMSE values with different azimuth angles will be selected as the RMSE value corresponding to the elevation angle. An example curve is shown in Figure 9.

![Fig. 9 Approximation of the elevation angle from the RMSE data using the least square fitted polynomial.](image)

### Algorithm 0.1 Complete 3D orientation localization

1. Calculate the ITD, $\hat{T}$, from the recorded signals of two microphones.
2. IF $\hat{A}_d < \hat{d}_{\text{threshold}}$ THEN
3. The elevation angle of the sound source is $\theta = 90^\circ$ and the azimuth angle, $\varphi$, is undefined.
4. ELSE
5. Estimate the azimuth $\varphi_{2D}$ and $\varphi_{3D}$ using 2D and 3D localization models, respectively
6. Calculate the RMSE between $\varphi_{2D}$ and $\varphi_{3D}$
7. IF RMSE < RMSE$_{\text{threshold}}$ THEN
8. Use polynomial curve fitting to determine $\theta$ using the calculated RMSE value and estimate $\varphi$ using either 2D or 3D localization model
9. ELSE estimate both $\theta$ and $\varphi$ using the 3D localization model
10. END IF
11. END IF

### 4.3 Complete Orientation Localization Algorithm

Figure 10 illustrates the flowchart of the proposed algorithm for the complete orientation localization. The pseudo code of the proposed complete orientation localization is given in Algorithm 0.1. The RMSE$_{\text{threshold}}$ is the value used to check when the elevation angle is close to $0^\circ$. This threshold value decides the point until which the curve fitting is required, and after which the 3D model can be trusted for elevation estimation.

![Fig. 10 Flowchart addressing the non-observability problems reflected in the 3D localization model.](image)
5 Distance Localization

The novel distance localization approach presented in this section depends on an accurate orientation localization. Assume that the angular location of the sound source has been obtained by using Algorithm 0.1 and the microphone array has been regulated facing toward the sound source, as shown in Figure 11. The proposed distance localization approach requires the microphone array, $LR$, to translate with a distance $\Delta d$ along the line perpendicular to the center-source vector (on the horizontal plane). This translation shifts the center of the microphone array, $O$, to a new point, $O'$, and $\gamma$ is defined as the angle between vectors $OS$ and $O'S$, as shown in Figure 12. Note that the center of the robot, $O$, is unchanged. The objective is to estimate distance $D$ between the center of the robot $O$ and the source $S$.

5.1 Mathematical Model for Distance Localization

Consider the gray triangle shown in Figure 12. Based on the far-field assumption in Section 2.2, the length $d'$ is given by

$$d' = b \sin \gamma.$$  \hspace{1cm} (22)

In triangle $\triangle SOO'$, we have

$$\sin \gamma = \frac{\Delta d}{\sqrt{(\Delta d)^2 + D^2}}.$$  \hspace{1cm} (23)

Defining the state as $x_{dist} = D$ and output as $y_{dist}$, the state-space model is given by

$$\dot{x}_{dist} = 0,$$  \hspace{1cm} (24)

$$y_{dist} = \frac{b \Delta d}{\sqrt{(\Delta d)^2 + D^2}}.$$  \hspace{1cm} (25)

Theorem 3 The system described by Equations (24) and (25) is observable if the following conditions are satisfied:

1) $b \neq 0$, 2) $\Delta d \neq 0$, and 3) $D \neq 0$.

Proof The observability matrix associated with (24) and (25) is given by

$$O_{dist} = \begin{bmatrix} -2 b^2 (\Delta d)^2 & D & \cdots \end{bmatrix}.$$  \hspace{1cm} (26)

So the system is observable if

$$b \neq 0, \ \Delta d \neq 0, \ \text{and} \ D \neq 0.$$  \hspace{1cm} (27)

Remark 3 As the microphones are separated by a non-zero distance, i.e., $b \neq 0$, and the microphone array is being translated by a non-zero distance, i.e., $\Delta d \neq 0$, the system is always observable unless the sound source and the robot are at same location making $D = 0$, which is not in the scope of discussion of this paper.

6 Extended Kalman Filter

The estimation for the angles and distance of the sound source is conducted by extended Kalman filters. Detailed mathematical derivation of the EKF can be found in [4]. Algorithm 0.2 summaries the EKF procedure used in this paper for SSL. The sensor covariance matrix ($R$) is defined as $\sigma_w^2$, and the process covariance matrix ($Q$) is defined as $\sigma_v^2$ for the distance localization, $\sigma^{2}_{\theta_1}$ for the 2D orientation localization and $\text{diag}\{\sigma^{2}_{\theta_1}, \sigma^{2}_{\theta_2}\}$ for the 3D orientation localization, respectively, where $\sigma_v$ is the process noise variance corresponding to the $i^{th}$ state and $\sigma_w$ is the sensor noise variance. Key parameters are listed in Table 1. The complete EKF-based SSL procedure is illustrated in Figure 13.
Table 1 EKF parameters.

| Parameter                                    | Angular localization | Distance localization |
|---------------------------------------------|----------------------|-----------------------|
| Process noise variance ($\sigma_{vi,i}$)    | 0.01                 | 0.1                   |
| Sensor noise variance ($\sigma_{w}$)        | 0.01                 | 0.001                 |
| Initial azimuth angle estimate ($\varphi_{initial}$) | 5°                   | –                     |
| Initial elevation angle estimate ($\theta_{initial}$) | 5°                   | –                     |
| Initial distance estimate ($D_{initial}$)   | –                    | 1 m                   |

Algorithm 0.2 Pseudo code for EKF

1: Initialize: \( \hat{x} \)
2: At each value of sample rate \( T_{out} \),
3: FOR \( i = 1 \) to \( N \) DO
4: \( \dot{x} = x + \left( \frac{T_{out}}{N} \right) f(\dot{x}, u) \)
5: \( A_J = \frac{\partial f}{\partial x}(x, u) \)
6: \( P = P + \left( \frac{T_{out}}{N} \right)(A_J P + PA_J^T + Q) \)
7: \( C_J = \frac{\partial h}{\partial x}(x, u) \)
8: \( K = P C_J^T (R + C_J P C_J^T)^{-1} \)
9: \( P = (I - K C_J) P \)
10: \( \dot{x} = x + K(y[n] - h(\hat{x}, u[n])) \)
11: END FOR

Fig. 13 Block diagram showing the process for the proposed complete angular and distance localization of a sound source using successive rotational and translational motions of a set of two microphones.

7 Simulation Results

In this section, we present the simulation results of the proposed localization technique for both angle and distance localization of a sound source.

7.1 Simulation Environment

The Audio Array Toolbox [16] is used to simulate a rectangular space using the image method described in [2]. The robot was placed in the center (origin) of the room. The two microphones were separated by a distance of 0.18 m from each other which is equal to the approximate distance between human ears. The sound source and the microphones are assumed omnidirectional and the attenuation of the sound is calculated as per the specifications in Table 2.

Table 2 Simulated room specifications

| Parameter                        | Value        |
|----------------------------------|--------------|
| Dimension                        | 20m x 20m x 20m |
| Reflection coefficient of each wall | 0.5            |
| Reflection coefficient of the floor | 0.5            |
| Reflection coefficient of the ceiling | 0.5            |
| Velocity of the sound            | 345 m/s      |
| Temperature                      | 22°C         |
| Static pressure                  | 29.92 mmHg   |
| Relative humidity                | 38 %         |

7.2 Validation of Observability

As discussed earlier, Theorem [1] shows that the 2D model is always observable, however, it does not provide any elevation information of the sound source. On the other hand, Theorem [2] shows that the 3D model is unobservable when the elevation angle of the sound source is 0° or 90°. In order to validate the observability analysis, localization was performed in the simulated environment.

For a sound source located on a 2D plane, Figure [14] shows the average of absolute estimation errors versus different azimuth angles with the sound source at distance of 5 m and 10 m to the robot, respectively. It can be seen that all errors are smaller than 1.8° and the mean of the average of absolute errors is approximately 1° for the two cases.

To verify the observability conditions for the 3D model as described by Equations (10) and (11), the
A number of experiments were performed to validate the performance of the proposed SSL technique for orientation localization, as described in Algorithm 0.1. White noise and speech signals were used as a sound source which was placed individually at different locations in the simulated room with specifications summarized in Table 2. The microphone array was rotated with an angular velocity of $\omega = 2\pi/5$ rad/sec in the clockwise direction for three complete revolutions. The ITD was calculated after every $1^\circ$ rotation followed by the estimation performed using the EKF with parameters given in Table 1. Four different sets of experiments were performed keeping the source at different locations. In first two sets of experiments, the source was placed in all four quadrants including the axes at different distances, keeping the elevation constant at $90^\circ$. To validate the performance of the proposed solution to the non-observability conditions, other two sets experiments were performed by keeping the sound source at different locations. Larger errors were observed when the elevation was close to $90^\circ$, which again echoes Theorem 2.

7.3 Simulation Results for Orientation Localization

A number of experiments were performed to validate the performance of the proposed SSL technique for orientation localization, as described in Algorithm 0.1. White noise and speech signals were used as a sound source which was placed individually at different locations in the simulated room with specifications summarized in Table 2. The microphone array was rotated with an angular velocity of $\omega = 2\pi/5$ rad/sec in the clockwise direction for three complete revolutions. The ITD was calculated after every $1^\circ$ rotation followed by the estimation performed using the EKF with parameters given in Table 1. Four different sets of experiments were performed keeping the source at different locations. In first two sets of experiments, the source was placed in all four quadrants including the axes at different distances, keeping the elevation constant at $20^\circ$ and $60^\circ$. To validate the performance of the proposed solution to the non-observability conditions, other two sets experiments were performed by keeping the sound source at elevation close to $0^\circ$ and $90^\circ$. The results of the localization are presented in Tables 3 and 4. It can be seen that orientation localization is achieved with errors less than $4^\circ$ using speech as well as white noise sound source. Large errors are observed when the elevation of the sound source is around $0^\circ$ and $90^\circ$. Further, the errors with source elevation around $0^\circ$ is less as compared to source elevation around $90^\circ$. This was achieved by using polynomial curve fitting approach mentioned in Section 4.2 with $RMSE_{\text{threshold}} = 1.9^\circ$, which corresponds to $\theta = 15^\circ$ on the fitted curve shown in Figure 9. The value $d_{\text{threshold}}$ was calculated as 0.017 m.
Table 3 Simulation results of orientation localization for speech

| Expt. No. | Act. D(m) | Act. $\varphi(^o)$ | Act. $\varphi(^o)$ | Est. $\psi(^o)$ | Est. $\theta(^o)$ | Avg of abs error ($^o$) | Avg of abs error ($^o$) |
|-----------|-----------|--------------------|--------------------|-----------------|-----------------|-------------------------|-------------------------|
| 1 a       | 5         | 0                  | 0.60              | 0.60            | 20.39           | 0.39                    |                         |
| 1 b       | 5         | 50                 | 51.03             | 1.03            | 21.44           | 1.44                    |                         |
| 1 c       | 7         | 90                 | 91.21             | 0.21            | 20.83           | 0.83                    |                         |
| 1 d       | 7         | 120                | 121.57            | 1.57            | 20.96           | 0.96                    |                         |
| 1 e       | 3         | 180                | 181.03            | 1.03            | 20.16           | 0.16                    |                         |
| 1 f       | 3         | -40                | -39.33            | 0.67            | 19.10           | 0.90                    |                         |
| 1 g       | 10        | -90                | -88.85            | 1.15            | 21.66           | 1.66                    |                         |
| 1 h       | 10        | -140               | -139.52           | 0.48            | 21.48           | 1.18                    |                         |
| 2 a       | 5         | 0                  | 2.31              | 2.31            | 60.68           | 0.68                    |                         |
| 2 b       | 5         | 50                 | 50.65             | 0.65            | 60.53           | 0.53                    |                         |
| 2 c       | 7         | 90                 | 91.79             | 1.79            | 60.70           | 0.70                    |                         |
| 2 d       | 7         | 120                | 121.85            | 1.85            | 60.84           | 0.84                    |                         |
| 2 e       | 3         | 180                | 181.66            | 1.66            | 60.05           | 0.05                    |                         |
| 2 f       | 3         | -40                | -35.66            | 1.34            | 60.38           | 0.38                    |                         |
| 2 g       | 10        | -90                | -89.38            | 0.62            | 59.62           | 0.38                    |                         |
| 2 h       | 10        | -140               | -138.20           | 1.80            | 59.78           | 0.22                    |                         |
| 3 a       | 5         | 50                 | 50.69             | 0.31            | 0               | 3.39                    | 3.39                    |
| 3 b       | 7         | 90                 | 90.41             | 0.41            | 0               | 3.39                    | 3.39                    |
| 3 c       | 3         | 180                | 181.16            | 1.16            | 60.51           | 0.51                    |                         |
| 3 d       | 7         | -120               | -119.00           | 1.00            | 2.40            | 1.60                    |                         |
| 3 e       | 3         | -40                | not def.          | not def.        | 86              | 90.00                   | 4.00                    |
| 3 f       | 7         | 150                | not def.          | not def.        | 89              | 90.00                   | 1.00                    |

Table 4 Simulation results of orientation localization for white noise

| Expt. No. | Act. D(m) | Act. $\varphi(^o)$ | Act. $\varphi(^o)$ | Est. $\psi(^o)$ | Est. $\theta(^o)$ | Avg of abs error ($^o$) | Avg of abs error ($^o$) |
|-----------|-----------|--------------------|--------------------|-----------------|-----------------|-------------------------|-------------------------|
| 1 a       | 5         | 0                  | 1.18              | 1.18            | 19.66           | 0.34                    |                         |
| 1 b       | 5         | 50                 | 51.03             | 1.03            | 20.44           | 0.44                    |                         |
| 1 c       | 7         | 90                 | 90.25             | 0.25            | 20.11           | 0.11                    |                         |
| 1 d       | 7         | 120                | 121.35            | 1.35            | 19.70           | 0.30                    |                         |
| 1 e       | 3         | 180                | 180.41            | 0.41            | 20.48           | 0.48                    |                         |
| 1 f       | 3         | -40                | -39.44            | 0.56            | 19.75           | 0.25                    |                         |
| 1 g       | 10        | -90                | -89.11            | 0.89            | 19.71           | 0.29                    |                         |
| 1 h       | 10        | -140               | -139.67           | 0.35            | 21.18           | 1.18                    |                         |
| 2 a       | 5         | 0                  | 1.31              | 1.31            | 60.38           | 0.38                    |                         |
| 2 b       | 5         | 50                 | 51.59             | 1.59            | 60.39           | 0.39                    |                         |
| 2 c       | 7         | 90                 | 90.74             | 0.74            | 60.87           | 0.57                    |                         |
| 2 d       | 7         | 120                | 121.21            | 1.21            | 60.39           | 0.39                    |                         |
| 2 e       | 3         | 180                | 181.16            | 1.16            | 60.51           | 0.51                    |                         |
| 2 f       | 3         | -40                | -35.66            | 1.34            | 60.41           | 0.41                    |                         |
| 2 g       | 10        | -90                | -88.90            | 1.10            | 60.70           | 0.70                    |                         |
| 2 h       | 10        | -140               | -138.64           | 1.36            | 60.57           | 0.57                    |                         |
| 3 a       | 5         | 50                 | 51.45             | 1.45            | 0               | 1.57                    | 1.57                    |
| 3 b       | 7         | -120               | -118.36           | 1.64            | 0               | 1.57                    | 2.43                    |
| 3 c       | 3         | -40                | not def.          | not def.        | 86              | 90.00                   | 4.00                    |
| 3 d       | 7         | 150                | not def.          | not def.        | 89              | 90.00                   | 1.00                    |

(which corresponds to $\theta = 85^o$, thereby giving an accuracy of $5^o$ when the sound source gets close to $90^o$ elevation) for the simulated environment with specification given in Table 2.

7.4 Simulation Results for Distance Localization

Speech and white-noise sounds were also used to test the performance of the distance localization. A single sound source was placed at different locations and the ITD signal was recorded while the microphone array was continuously shifted for 200 steps each with a distance of $\Delta d = 0.0007$ m. The results are summarized in Tables 5 and 6. The key parameters of the EKF are given in Table 1. The results for the distance localization with a sound source placed at different locations are shown in Figure 18. It is observed that the error in the estimation converges quickly and a total shift of microphone array of approximately 3 cm is sufficient for the estimates to completely converge to and remain in the three standard deviation bounds. The average of absolute error in the estimation is found to be less than
### Table 5 Simulation results of distance localization using speech sound source

| Expt. No. | Act. $\varphi (^\circ)$ | Act. $\theta (^\circ)$ | Act. D(m) | Est. D(m) | Avg of abs error (m) |
|-----------|------------------------|----------------------|-----------|-----------|---------------------|
| 1 a       | 0                      | 5                    | 5.01      | 0.01      |
| 1 b       | 50                     | 5                    | 5.01      | 0.01      |
| 1 c       | 90                     | 7                    | 6.94      | 0.06      |
| 1 d       | 120                    | 7                    | 6.93      | 0.07      |
| 1 e       | 180                    | 3                    | 3.01      | 0.01      |
| 1 f       | -40                    | 3                    | 3.01      | 0.01      |
| 1 g       | -90                    | 10                   | 9.54      | 0.46      |
| 1 h       | -140                   | 10                   | 9.81      | 0.19      |
| 2 a       | 0                      | 5                    | 5.02      | 0.02      |
| 2 b       | 50                     | 5                    | 5.02      | 0.02      |
| 2 c       | 90                     | 7                    | 6.94      | 0.06      |
| 2 d       | 120                    | 7                    | 6.94      | 0.06      |
| 2 e       | 180                    | 3                    | 3.00      | 0.00      |
| 2 f       | -40                    | 3                    | 3.01      | 0.01      |
| 2 g       | -90                    | 10                   | 9.52      | 0.48      |
| 2 h       | -140                   | 10                   | 9.41      | 0.59      |
| 3 a       | 50                     | 0                    | 5.02      | 0.02      |
| 3 b       | -120                   | 4                    | 7.68      | 0.13      |
| 4 a       | -40                    | 86                   | 5         | 0.02      |
| 4 b       | 150                    | 89                   | 7         | 6.83      | 0.17 |

### Table 6 Simulation results of distance localization using white noise sound source

| Expt. No. | Act. $\varphi (^\circ)$ | Act. $\theta (^\circ)$ | Act. D(m) | Est. D(m) | Avg of abs error (m) |
|-----------|------------------------|----------------------|-----------|-----------|---------------------|
| 1 a       | 0                      | 5                    | 5.01      | 0.01      |
| 1 b       | 50                     | 5                    | 5.01      | 0.01      |
| 1 c       | 90                     | 7                    | 6.92      | 0.08      |
| 1 d       | 120                    | 7                    | 6.92      | 0.08      |
| 1 e       | 180                    | 3                    | 3.01      | 0.01      |
| 1 f       | -40                    | 3                    | 3.01      | 0.01      |
| 1 g       | -90                    | 10                   | 9.52      | 0.48      |
| 1 h       | -140                   | 10                   | 9.44      | 0.56      |
| 2 a       | 0                      | 5                    | 5.01      | 0.01      |
| 2 b       | 50                     | 5                    | 5.01      | 0.01      |
| 2 c       | 90                     | 7                    | 6.92      | 0.08      |
| 2 d       | 120                    | 7                    | 6.92      | 0.08      |
| 2 e       | 180                    | 3                    | 3.01      | 0.01      |
| 2 f       | -40                    | 3                    | 3.01      | 0.01      |
| 2 g       | -90                    | 10                   | 9.48      | 0.52      |
| 2 h       | -140                   | 10                   | 9.45      | 0.57      |
| 3 a       | 50                     | 0                    | 5         | 0.01      |
| 3 b       | -120                   | 4                    | 7         | 6.89      | 0.11 |
| 4 a       | -40                    | 86                   | 5         | 0.01      |
| 4 b       | 150                    | 89                   | 7         | 6.90      | 0.10 |

8 Experimental Results

Experiments were conducted using two different hardware platforms: a KEMAR dummy head in a well equipped hearing laboratory and a robotic platform equipped with a set of two rotational microphones. The following subsections discuss the hardware platforms and the results.

#### 8.1 Results using KEMAR Dummy Head

Experiments using the KEMAR dummy head were conducted in a high frequency focused sound treated room with dimension 4.6 m x 3.7 m x 2.7 m as shown in Figure 19. The ITD however is mostly effective for low frequency sounds below 1.5 kHz as a spatial hearing cue. The walls, floor, and ceiling of the room were covered by polyurethane acoustic foam with a thickness of only 5 cm which is relatively low compared to the sound wavelength thereby making a relatively low reduction in low and middle frequencies, thereby making it a challenging acoustic environment. For broad band noise, T60 (i.e., the time required for the sound level to decay 60 dB) was 97 ms. In an octave band centered at 1000 Hz, T60 for the noise was on an average of 324 ms.

The digitally generated audio signals using a MATLAB program and three 12-channel Digital-to-Analog converters running at 44,100 cycles each second per channel were amplified using AudioSource AMP 1200 amplifiers before they were played from an array of 36 loudspeakers. The two microphones were installed on the KEMAR dummy head temporarily mounted on a rotating chair which was rotated at an approximate rate of 32°/s for about two circles in the middle of the room. The data collected in the second rotation was used for the EKF. Motion data was collected by the gyroscope mounted on the top of the dummy head. The audio signals were amplified and collected by a sound card which were then stored on a desktop computer for further processing. The ITD was processed with a generalized cross-correlation model in each time frame.
corresponding to the 120 Hz sampling rate of the gyroscope. The computation was completed by a MATLAB program on a desktop computer. Raw data with a single sound source located at four different locations were collected.

The left two subfigures in Figure 20 are generated when the actual elevation angle is 0°. It can be seen that the azimuth estimations using the 2D and 3D models are very close, which implies the elevation estimates are not reliable (in the top-left figure). When θ = 30°, the azimuth estimates are obviously different (in the top-right figure), which implies reliable elevation estimates using the 3D model (in the bottom-right figure).

8.2 Results using Robotic Platform

Experiments were also performed using a robotic platform shown in Figure 21. In these experiments, two microelectromechanical systems (MEMS) analog/digital microphones were used for recording the sound signal coming from the sound source. Flex adapters were used to hold the microphones. The angular speed of the rotation of the microphone array was controlled by a bipolar stepper motor with a gear ratio adjusted to 0.9° per step. The stepper motor was controlled by an Arduino microprocessor. The distance between the two microphones was kept constant as 0.3 m. An audio (music) was played in a loud speaker which was used as a sound source kept at different locations. The estimation results are shown in Figure 22 and Table 8.
Table 7 Experimental results using KEMAR dummy head: Orientation localization using the 3D model. (RMSE: difference between azimuth estimations using the 2D and 3D models, respectively)

| Expt. No. | Act. \( \varphi \) (°) | Est. \( \varphi \) (°) | Avg of abs error (°) | RMSE \( \varphi \) (°) | Act. \( \theta \) (°) | Est. \( \theta \) (°) | Avg of abs error (°) |
|-----------|-----------------------|-----------------------|----------------------|-----------------------|-----------------------|-----------------------|----------------------|
| 1         | 90                    | 91.21                 | 1.21                 | 1.39                  | 0                     | 13.64                 | 13.64                |
| 2         | -20                   | -21.53                | 1.63                 | 1.16                  | 0                     | 48.14                 | 48.14                |
| 3         | 90                    | 90.40                 | 0.40                 | 79.94                 | 60                    | 59.05                 | 0.95                 |

Table 8 Experimental results using the robotic platform: Orientation localization using 3D model (RMSE: difference between azimuth estimations using the 2D and 3D models, respectively)

| Expt. No. | Act. \( \varphi \) (°) | Est. \( \varphi \) (°) | Avg of abs error (°) | RMSE \( \varphi \) (°) | Act. \( \theta \) (°) | Est. \( \theta \) (°) | Avg of abs error (°) |
|-----------|-----------------------|-----------------------|----------------------|-----------------------|-----------------------|-----------------------|----------------------|
| 1         | -140                  | -140.65               | 0.65                 | 0.72                  | 0                     | 14.96                 | 14.96                |
| 2         | 180                   | 178.71                | 1.29                 | 0.69                  | 5                     | 11.59                 | 6.59                 |
| 3         | 40                    | 39.67                 | 0.33                 | 8.80                  | 55                    | 55.24                 | 0.24                 |
| 4         | 40                    | -38.20                | 1.80                 | 10.96                 | 65                    | 64.67                 | 0.33                 |

Fig. 22 Experimental results for orientation localization using the robotic platform. When \( \theta = 0° \), the azimuth estimates using the 2D and 3D models are very close (in the top-left figure), which implies the elevation estimates are not reliable (in the bottom-left figure). When \( \theta = 55° \), the azimuth estimates are obviously different (in the top-right figure), which implies reliable elevation estimates using the 3D model (in the bottom-right figure).

using the RMSE of the difference between the estimated azimuth values using respectively the 2D and 3D models.

A fitted curve similar to one shown in the Figure 9 can be generated for the environment by keeping the sound source at different elevation angles and recording the RMSE values between \( \varphi_{2D} \) and \( \varphi_{3D} \) estimations. The value of the parameter \( RMSE_{\text{threshold}} \) can be decided, which can be used to check the \( \theta = 0° \) scenario. Further, the generated fitted curve can be used to give a closer estimation of the elevation angle.

9 Conclusion

This paper presents a novel technique that performs a complete localization (i.e., both orientation and distance) of a stationary sound source in a three-dimensional (3D) space. Two singular conditions when unreliable orientation localization (the elevation angle equals 0 or 90°) occurs were found by using the observability theory. The root-mean-squared error (RMSE) value of the difference between the azimuth estimates using respectively the 2D and 3D models was used to check the 0° elevation condition and the elevation was further estimated using a polynomial curve fitting technique. The 90° elevation was detected by checking zero-ITD signal. Based on an accurate orientation localization, the distance localization was done by first rotating the microphone array to face toward the sound source and then shifting the microphones perpendicular to the source-robot vector by a distance of a fixed number of steps. Under challenging acoustic environments with relatively low-energy targets and high-energy noise, high localization accuracy was achieved in both simulations and experiments. The mean of the average of absolute estimation error was less than 4° for angular localization and less than 0.6 m for distance localization in simulation results and techniques to detect \( \theta = 0° \) and 90° are verified in both simulation and experimental results.

Acknowledgements

The authors would like to thank Dr. Xuan Zhong for providing with the experimental raw data using the KEMAR dummy head.

References

1. International Organization for Standardization (ISO), British, European and International Standards (BSEN), Noise emitted by machinery and equipment – Rules for the drafting and presentation of a noise test code. 12001: 1997 Acoustics
39. Pang, C., Liu, H., Zhang, J., Li, X.: Binaural sound localization based on reverberation weighting and generalized parametric mapping. IEEE/ACM Transactions on Audio, Speech, and Language Processing 25(8), 1618–1632 (2017). doi:10.1109/TASLP.2017.2703650
40. Perrett, S., Noble, W.: The effect of head rotations on vertical plane sound localization. The Journal of the Acoustical Society of America 102(4), 2325–2332 (1997).
41. Rodemann, T.: A study on distance estimation in binaural sound localization. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 425–430 (2010). doi:10.1109/IROS.2010.5651455
42. Rodemann, T., Ince, G., Joublin, F., Goerick, C.: Using binaural and spectral cues for azimuth and elevation localization. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2185–2190 (2008). doi:10.1109/IROS.2008.4650667
43. Sabine, W.: Collected Papers on Acoustics. Harvard University Press (1922)
44. Spriet, A., Van Deun, L., Eftaxiadis, K., Laneau, J., Moonen, M., Van Dijk, B., Van Wieringen, A., Wouters, J.: Speech understanding in background noise with the two-microphone adaptive beamformer beam in the nucleus freedom cochlear implant system. Ear and hearing 28(1), 62–72 (2007). doi:10.1097/01.aud.0000252470.54246.54
45. Sturin, D.E., Brandstein, M.S., Silverman, H.F.: Tracking multiple talkers using microphone-array measurements. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 1, pp. 371–374 vol.1 (1997). doi:10.1109/ICASSP.1997.599650
46. Sun, L., Zhong, X., Yost, W.: Dynamic binaural sound source localization with interaural time difference cues: Artificial listeners. The Journal of the Acoustical Society of America 137(4), 2226–2226 (2015). doi:10.1121/1.4920636
47. Tamai, Y., Kagami, S., Amemiya, Y., Sasaki, Y., Mizoguchi, H., Takano, T.: Circular microphone array for robot’s audition. In: Proceedings of IEEE Sensors, 2004., vol. 2, pp. 565–570 (2004). doi:10.1109/ICSENS.2004.1426228
48. Tamai, Y., Sasaki, Y., Kagami, S., Mizoguchi, H.: Three ring microphone array for 3D sound localization and separation for mobile robot audition. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4172–4177 (2005). doi:10.1109/IROS.2005.1545095
49. Tietze, J., Dominguez, F., Silva, B.d., Segers, L., Steenhaut, K., Touhafi, A.: Soundcompass: a distributed MEMS microphone array-based sensor for sound source localization. Sensors 14(2), 1918–1949 (2014). doi:10.3390/s140201918
50. Valin, J.M., Michaud, F., Rouat, J., Letourneau, D.: Robust sound source localization using a microphone array on a mobile robot. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), vol. 2, pp. 1228–1233 (2003). doi:10.1109/IROS.2003.1248813
51. Wallach, H.: On sound localization. The Journal of the Acoustical Society of America 10(4), 270–274 (1939). doi:10.1121/1.1915985
52. Wang, H., Chu, P.: Voice source localization for automatic camera pointing system in videoconferencing. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 1, pp. 187–190 (1997). doi:10.1109/ICASSP.1997.599056
53. Yost, W.A., Zhong, X.: Sound source localization identification accuracy: Bandwidth dependencies. The Journal of the Acoustical Society of America 136(5), 2737–2746 (2014). doi:10.1121/1.4898045
54. Zhong, X., Sun, L., Yost, W.: Active binaural localization of multiple sound sources. Robotics and Autonomous Systems 85, 83–92 (2016). doi:10.1016/j.robot.2016.07.008
55. Zhong, X., Yost, W., Sun, L.: Dynamic binaural sound source localization with ITD cues: Human listeners. The Journal of the Acoustical Society of America 137(4), 2376–2376 (2015). doi:10.1121/1.4920636
56. Zietlow, Y., Hussein, H., Kowerko, D.: Acoustic source localization in home environments-the effect of microphone array geometry. In: 28th Conference on Electronic Speech Signal Processing, pp. 219–226 (2017)