Saqib *et al.*
EURASIP Journal on Audio, Speech, and Music Processing (2024) 2024:22
https://doi.org/10.1186/s13636-024-00340-y

EURASIP Journal on Audio, Speech, and Music Processing

EMPIRICAL RESEARCH

Open Access

Robust acoustic reflector localization using a modified EM algorithm

Usama Saqib[1*†], Mads Græsbøll Christensen[1†] and Jesper Rindom Jensen[1†]

Abstract

In robotics, echolocation has been used to detect acoustic reflectors, e.g., walls, as it aids the robotic platform to navigate in darkness and also helps detect transparent surfaces. However, the transfer function or response of an acoustic system, e.g., loudspeakers/emitters, contributes to non-ideal behavior within the acoustic systems that can contribute to a phase lag due to propagation delay. This non-ideal response can hinder the performance of a time-of-arrival (TOA) estimator intended for acoustic reflector localization especially when the estimation of multiple reflections is required. In this paper, we, therefore, propose a robust expectation-maximization (EM) algorithm that takes into account the response of acoustic systems to enhance the TOA estimation accuracy when estimating multiple reflections when the robot is placed in a corner of a room. A non-ideal transfer function is built with two parameters, which are estimated recursively within the estimator. To test the proposed method, a hardware proof-of-concept setup was built with two different designs. The experimental results show that the proposed method could detect an acoustic reflector up to a distance of 1.6 m with 60% accuracy under the signal-to-noise ratio (SNR) of 0 dB. Compared to the state-of-the-art EM algorithm, our proposed method provides improved performance when estimating TOA by 10% under a low SNR value.

Keywords TOA estimation, DOA estimation, Expectation-maximization, Active source localization, Robot/drone audition, Prewhitening

1 Introduction

Within the context of robot audition, the use of echolocation for acoustic reflector localization and estimation has been proposed by various researchers in the past [1–3]. Within this domain, researchers are utilizing acoustic signal processing techniques and propose combining echolocation with state-of-the-art technologies, e.g., laser- and camera-based technologies to aid a robot in constructing a spatial map of an indoor environment. This can be accomplished by a collocated microphone-loudspeaker combination. One major disadvantage of the camera and laser-based technologies is that they cannot work in complete darkness and cannot detect transparent surfaces that are typically found in an office environment, This makes accurate construction of a spatial map of an environment a difficult process.

The process involved in the aforementioned echolocation techniques is to probe the environment with a known sound so that the reflected signal acquired by a microphone can be processed to estimate the time of arrival (TOA) of the acoustic echo that aids a robot to estimate the distance between the acoustic reflector. Traditionally, TOA information is extracted from room impulse response (RIR) estimates (Fig. 1) which is normally done using a peak-picking approach [2–6]. This model is broadly divided into two distinct parts: the direct path including early reflections and late reflections which are comprised of a stochastic dense tail [7]. The direct-path component is the shortest distance a sound can take, i.e., it provides information about the distance between the transmitter and receiver while

†Usama Saqib, Mads Græsbøll Christensen and Jesper Rindom Jensen contributed equally to this work.

*Correspondence:
Usama Saqib
ussa@create.aau.dk
[1] Department of Electronic Systems, Audio Analysis Lab, Aalborg University, Fredrik Bajers Vej 7K, Aalborg 9220, Denmark

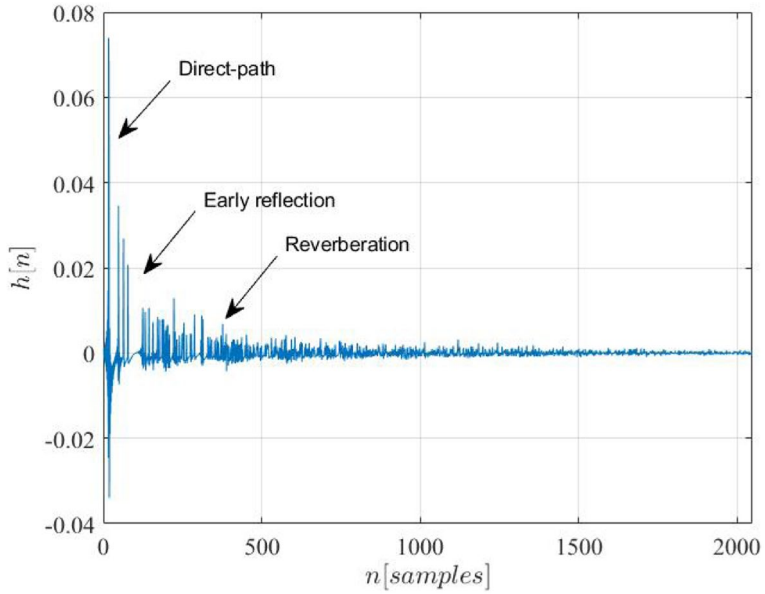

Fig. 1 Transfer function of the room between source and microphone, RIR. The direct path contains the highest energy followed by the early reflection and reverberation which is represented by a dense tail

early reflections help in inferring the distance of the closest acoustic reflector [2, 3, 8]. While TOA estimation enables a robot to determine the distance of an acoustic reflector, the direction-of-arrival (DOA) of an acoustic source is required to determine the location of an acoustic source. This is done by incorporating multiple receivers attached to a robot [9–11]. Recent advancement in machine learning techniques has also enabled robotic platform to incorporate echolocation for terrain classification and detecting echoes from noisy data. For example, in [12], the author proposed training using advanced signal filtering and machine learning techniques which could be used to accurately classify terrain types for a small mobile robot. One potential for such a method is to help robot navigation, i.e., detecting roads from other surfaces. Moreover, echolocation is used to map a spatial map of an indoor environment. For example, in [13], the authors propose training a neural network to predict depth maps and gray-scale images from sound alone. The work presented in [13] was later improved in [14] by improving the neural network and reducing the computation time needed to run the model. The contribution of the paper was a full 360^o $3D$ depth reconstruction with 4 microphones and a lidar-based SLAM for training a model. One notable difference between model-based approaches and data-driven approaches is the availability of large data sets required to train a neural network. Comparatively, the model-based approach finds the feature of interest directly from the signal model.

While ultrasonic sensors are popular within robotics to detect obstacles, these require specialized hardware to transmit/receive acoustic echoes and could potentially increase the overall cost of a robotic platform. However, most robots intended for human-robot interaction (HRI) consist of a collocated microphone-loudspeaker setup, e.g., Softbank's NAO robot. In our previous work, we proposed a TOA/DOA estimator based on the expectation-maximization (EM) framework [8] but with crude assumptions about the acoustic properties of the acoustic reflectors (point source, ideal reflectors, etc.) and the hardware (ideal response, omnidirectionality). However, these assumptions lead to a detrimental model mismatch in practical settings, e.g., since loudspeakers/microphones contribute to a phase lag due to propagation delay [15], which deteriorates the performance of the TOA/DOA estimator in [8, 16], particularly in the presence of multiple acoustic reflections. This causes a severe problem when using the TOA/DOA estimates in robots for generating a spatial map of an indoor environment using acoustic echoes. Therefore, we propose an algorithm that utilizes the previously proposed loudspeaker-microphone setup to estimate the distance of an acoustic reflector, while estimating the response of the acoustic systems, which may facilitate simultaneous estimation of multiple acoustic echoes impinging at different TOAs and/or from different DOAs.

Traditionally, estimating the transfer function of the loudspeaker is usually done using a loudspeaker-enclosed microphone (LEM) setup which involves placing the setup within an anechoic environment. However, in [17], the researchers proposed a method to measure the

transfer function of the loudspeaker within an echoic environment. This is done by utilizing two loudspeakers, one of them calibrated and its transfer function already estimated within an anechoic chamber. The loudspeaker is placed in a fixed location within the environment. The process involves transmitting a white noise signal through the calibrated loudspeaker to measure its impulse response (IR) and later replacing the loudspeaker with the uncalibrated loudspeaker and repeating the IR measurement. The transfer function of the uncalibrated loudspeaker is estimated using least squares. Furthermore, TOA estimation can also be influenced by the materials that acoustic reflectors are composed of, e.g., concrete, glass, and cardboard. This is because some materials absorb certain sound frequencies that could lead to non-ideal characteristics of the observed signals [18]. The aforementioned method requires access to an anechoic chamber which is a time-consuming process, hence, there is a need to estimate the response of the acoustic system directly from the model.

In this paper, we, therefore, extend the model-based method originally proposed in [19] and later used in our previous work [8] to accommodate the non-ideal transfer function of an acoustic system, i.e., the loudspeaker, the microphone, and the reflecting materials. We take a model-based approach to TOA estimation where the model of the early reflections is used to derive a statistically optimal estimator. More specifically, we include an unknown filter to model the uncertainties of the acoustic system which may alleviate the need to estimate loudspeaker IR measurement suggested in [17]. Moreover, to test the proposed method, a proof-of-concept setup is built to conduct experiments using real data.[1]

The remaining part of this paper is organized as follows: Section 2 introduces the problem formulation, and Section 3 proposes the TOA estimation method based on EM. Finally, the experimental results followed by discussion and conclusion can be found in Sections 4, 5, and 6, respectively.

2 Problem formulation

Consider the scenario where a loudspeaker is emitting a known probe signal, which is then propagating an acoustic environment, and recorded by a microphone. This can be mathematically modeled as

$$\begin{aligned} y(n) &= h(n) * s(n) + w(n) \\ &= x(n) + w(n), \end{aligned} \tag{1}$$

where $h(n)$ is the acoustic impulse response from the loudspeaker to the microphone, $s(n)$ is the known probe signal, and $w(n)$ is additive background noise while $x(n) = h(n) * s(n)$. The acoustic impulse response can be further modeled by decomposing the reverberation into early and late reverberation components. The early reflections are modeled as time-delayed and filtered versions of the known probe signal, where the filter represents the responses of the loudspeaker, microphone, and acoustic reflectors. Mathematically, we formulate this as

$$y(n) = \sum_{r=1}^{R} g_r * s(n - \tau_r) + v(n), \tag{2}$$

where R is the number of early reflections, g_r is the filter pertaining to the r^{th} reflection, τ_r is the delay of the r^{th} reflection, and $v(n)$ is a noise term embracing both the additive background noise and the late reflections. In the special case where $M = 1$ for all $r = 1, \ldots, R$, we get the ideal model used in [8], which does not account for the non-ideal hardware responses that are inevitable in real scenarios. We then assume stationarity and that we have N observations following this model, i.e.,

$$\mathbf{y}(n) = \sum_{r=1}^{R} \mathbf{G}_r \mathbf{s}(n - \tau_r) + \mathbf{v}(n), \tag{3}$$

$$= \sum_{r=1}^{R} \mathbf{S}_r(n - \tau_r)\mathbf{g}_r + \mathbf{v}(n), \tag{4}$$

$$\mathbf{G}_r = \left[\mathbf{D}_0\mathbf{g}_r, \mathbf{D}_1\mathbf{g}_r, \cdots, \mathbf{D}_{M-N}\mathbf{g}_r \right]^T \tag{5}$$

$$\mathbf{g}_r = [g_{0,r}, g_{1,r}, \cdots, g_{M-1,r}]^T. \tag{6}$$

$$\mathbf{S}(n-\tau) = \begin{bmatrix} s(n-\tau+M-1) & \cdots & s(n-\tau+N-M) \\ s(n-\tau+M) & \cdots & s(n-\tau+N-M+1) \\ \vdots & & \vdots \\ s(n-\tau+N-1) & \cdots & s(n-\tau) \end{bmatrix} \tag{7}$$

$$\begin{aligned} \mathbf{s}(n-\tau) = &[s(n-\tau), s(n-\tau+1), \cdots \\ &, s(n-\tau+N-1)]^T, \end{aligned} \tag{8}$$

Here, $\mathbf{D}$ is a cyclic shift register that delays filter gain $\mathbf{g}_r$. The matrix $\mathbf{G}_r$ has a dimension of $(N - M + 1) \times N$ while $\mathbf{S}$ has a dimension of $(N - M + 1) \times M$, where N is the length of the signal while M is the filter length. The filter $\mathbf{g}_r$ is a $1 \times M$ vector of the r-th reflection. If we assume that the noise term is white Gaussian noise, the maximum likelihood estimator for the unknown filters, $\mathbf{g}_r$, and delays, τ_r, for $r = 1, \ldots, R$, is given by

[1] The dataset and code for this work can be found here: https://doi.org/10.5281/zenodo.5082224

$$\{\widehat{\boldsymbol{\tau}}, \widehat{\mathbf{g}}\} = \arg \min_{\tau_r, g_r \forall r \in [1;R]} \left\| \mathbf{y}(n) - \sum_{r=1}^{R} \mathbf{S}(n-\tau_r)\mathbf{g}_r \right\|^2. \tag{9}$$

Compared to [19], we do not assume that the gain or filter $\mathbf{g}_r$ is set to 1. Hence, the problem at hand is to estimate the delay τ_r and the filter parameters $\mathbf{g}_r$. Moreover, in this paper, we are interested in estimating these parameters to localize the position of an acoustic reflector using echolocation which was not addressed in [19]. Furthermore, resolving (9) to estimate τ_r and $\mathbf{g}_r$ clearly, leaves us with a computationally complex and multidimensional task. However, as we shall see next, this can be solved by incorporating iterative procedures such as expectation-maximization (EM).

3 Robust EM-based acoustic reflector localization

The EM algorithm developed in [20] is a general method intended to solve maximum-likelihood (ML) estimation problem given incomplete data [19]. It is intended to alleviate the complexity of parameter estimation. The EM algorithm requires that the complete data be specified. Here, we may define our complete data as all the observations of the individual reflections, each defined as

$$\mathbf{x}_r(n) = \mathbf{S}(n-\tau_r)\mathbf{g}_r + \mathbf{v}_r(n), \tag{10}$$

for, $r = 1, \ldots, R$, where $\mathbf{v}_r(n)$ are individual noise terms obtained by arbitrarily decomposing the noise term $\mathbf{v}(n)$ into R components, such that

$$\sum_{r=1}^{R} \mathbf{v}_r(n) = \mathbf{v}(n). \tag{11}$$

Moreover, we can write the observed signal as the sum of the individual observed reflections, i.e.,

$$\mathbf{y}(n) = \sum_{r=1}^{R} \mathbf{x}_r(n). \tag{12}$$

We let the individual noise terms be independent, zero-mean, white Gaussian and distributed as $\mathcal{N}(\mathbf{0}, \beta_r \mathbf{C})$, where $\mathbf{0}$ is a vector of zeros and $\mathbf{C} = \mathrm{E}[\mathbf{v}(\mathrm{n})\mathbf{v}^{\mathrm{T}}(\mathrm{n})] = \sigma_{\mathrm{v}}^2 \mathbf{I}_{\mathrm{N}}$ is an $N \times N$ matrix of $\mathbf{v}(n)$, σ_v^2 is the variance. E[.] is the mathematical expectation. Moreover, the scaling factors, β_r, are non-negative, real-valued scalars that satisfy the following:

$$\sum_{r=1}^{R} \beta_r = 1. \tag{13}$$

Here, the β_r must satisfy the condition above but it is an arbitrary free variable and could be used to control the rate of convergence. The choice of β could be resort to more investigation as noted by [19] but here we choose the $\beta = 1/R$. The EM algorithm for the problem at hand is given by

E-step:

$$\begin{aligned} \widehat{\mathbf{x}}_r^{(i)}(n) =& \mathbf{S}(n-\widehat{\tau}_r^{(i)})\widehat{\mathbf{g}}_r^{(i)} \\ &+ \beta_r \left[\mathbf{y} - \sum_{r=1}^{R} \mathbf{S}(n-\widehat{\tau}_r^{(i)})\widehat{\mathbf{g}}_r^{(i)}. \right] \end{aligned} \tag{14}$$

M-step:

$$\{\widehat{\mathbf{g}}_r, \widehat{\tau}_r\}^{(i+1)} = \arg \min_{\mathbf{g},\tau} \left\| \mathbf{x}_r^{(i)}(n) - \mathbf{S}(n-\tau)\mathbf{g} \right\|^2, \tag{15}$$

where $^{(i)}$ denotes the iteration index. The M-step can be simplified since the estimator is linear with respect to the unknown filter coefficients. Moreover, under white Gaussian conditions, the estimator in (15) becomes a maximum likelihood estimator. We can thus solve for these first, which yields

$$\widehat{\mathbf{g}}_r^{(i+1)} = \left[\mathbf{S}^T(n-\tau_r)\mathbf{S}(n-\tau_r) \right]^{-1} \mathbf{S}^T(n-\tau_r)\mathbf{x}_r^{(i)}(n), \tag{16}$$

If we insert this back into (15), we get

$$\begin{aligned} \widehat{\tau}_r^{(i+1)} =& \arg \max_{\tau} \mathbf{x}_r^{(i)}\mathbf{S}(n-\tau) \left[\mathbf{S}^T(n-\tau)\mathbf{S}(n-\tau) \right]^{-1} \cdot \\ & \mathbf{S}^T(n-\tau)\mathbf{x}_r^{(i)}(n), \end{aligned} \tag{17}$$

A potential problem with these estimators is that the filter estimates $\widehat{\mathbf{g}}_r$ are unconstrained, which may lead to unreasonably large filter coefficients, since the reflections may partly cancel each other out. One way of addressing such problems is by introducing a constraint on the white noise gain of the filter:

$$\begin{aligned} \{\widehat{\mathbf{g}}_r, \widehat{\tau}_r\}^{(i+1)} = \arg \min_{\mathbf{g},\tau} & \left\| \mathbf{x}_r^{(i)}(n) - \mathbf{S}(n-\tau)\mathbf{g} \right\|^2 \\ & \text{s.t.} \quad \|\mathbf{g}\| < \epsilon. \end{aligned} \tag{18}$$

This can be solved using the method of Lagrange multipliers, i.e., to solve for the constrained filter, we write

$$\begin{aligned} \{\widehat{\mathbf{g}}_r, \widehat{\tau}_r\} &= \arg \min_{\mathbf{g},\tau} -2\mathbf{x}_r^T(n)\mathbf{S}(n-\tau)\mathbf{g} + \\ & \mathbf{g}^T\mathbf{S}^T(n-\tau)\mathbf{S}(n-\tau)\mathbf{g} + \lambda(\mathbf{g}^T\mathbf{g} - \epsilon) \\ &= \arg \min_{\mathbf{g},\tau} J(\mathbf{g}, \tau) \end{aligned} \tag{19}$$

By taking the partial derivative with respect to the filter, we get

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{g}_r} =& -\mathbf{S}^T(n-\tau_r)\mathbf{x}_r(n) + \mathbf{S}^T(n-\tau_r)\mathbf{S}(n-\tau_r)\mathbf{g}_r \\ & + \lambda \mathbf{g}_r = 0. \end{aligned} \tag{20}$$

That is, the filter estimate becomes

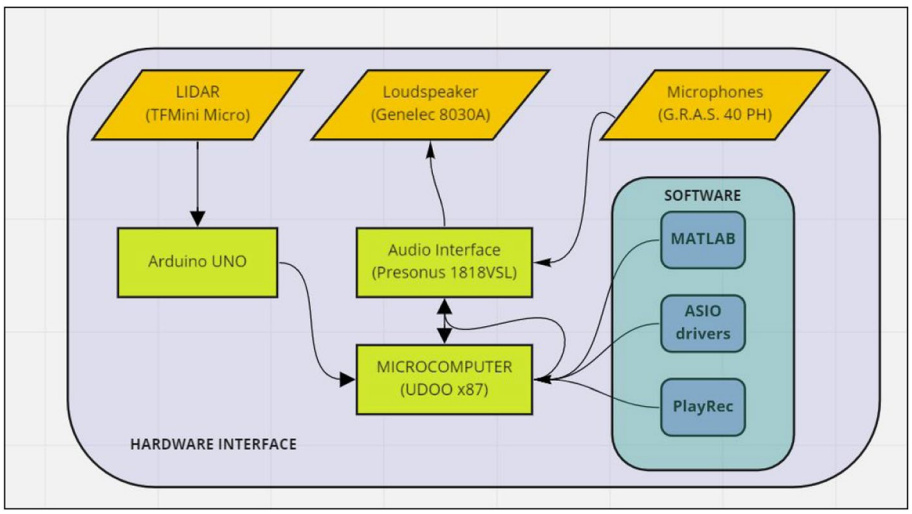

Fig. 2 An overview of the hardware required to design the platform used in this research

$$\widehat{\mathbf{g}}_r = \left[\mathbf{S}^T(n-\tau_r)\mathbf{S}(n-\tau_r) + \lambda\mathbf{I}\right]^{-1} \mathbf{S}^T(n-\tau_r)\mathbf{x}_r(n). \tag{21}$$

where λ is the tuning parameter that is empirically set while the $\mathbf{I}$ is the identity matrix. The estimated τ_r of an acoustic reflector could be converted into a distance estimate if we assume that the speed of sound is known for the given environment and that we are interested in estimating only the first-order early reflection. This simple conversion can be done as follows:

$$d = c \times \tau, \tag{22}$$

where c is the speed of sound and d is the distance of an acoustic reflector with respect to a source.

However, by taking the acoustic response within the model, we can estimate multiple reflections originating from two acoustic reflectors, i.e., first-order and second-order reflection. By combining the proposed method with eco-labeling [21–23], we can estimate the position of multiple acoustic echoes.

4 Experimental results

In this section, we investigate two issues, the performance of the proposed method under different conditions, and the benefit of estimating multiple acoustic echoes. In the first experiment, the proposed method was tested using signals that are synthesized using the room impulse response generator [24] with the following setup. The synthetic room has a dimension of $6.38 \times 5.4 \times 4.05$ m. The analysis window considered was set to $\tau_{\min}$ and $\tau_{\max}$ samples corresponding to a distance of 0.5 m to 3 m similar to the computation time to run performed in [25]. This analysis window also helps in estimating the first-order early reflection and prevents the direct-path component from being estimated. Moreover, the probe signal $s(n)$ is a broadband signal of length 2000 samples drawn from a Gaussian burst with zero padding to form a signal of length 20,000 samples.

4.1 Proof-of-concept

The experimental platform is used to evaluate the performance of the proposed method. The overall system architecture is shown in Fig. 2. Two design variations are proposed to test the proposed method for the acoustic reflector's position and distance estimation. One variation consists of a loudspeaker (Genelec 8030A) with a microphone (G.R.A.S 40 PH) attached to the top of the loudspeaker. The distance between the acoustic center of a loudspeaker and the center of a microphone is 0.15 m. This is shown in Fig. 3. The second variation consists of a 6 microphone arranged in a uniform circular array (UCA) of radius 0.2 m with a loudspeaker placed at the center of the UCA. This is shown in Fig. 4. The loudspeaker-microphone was placed 1.5 m above the floor inside Aalborg University's `Sound Lab` that has a dimension of $6.38 \times 5.4 \times 4.05$ m. Furthermore, both the loudspeaker and microphones are connected to an audio interface (Presonus 1818VSL). A Lidar sensor (TFMini Micro) is used to measure the distance between the wall and the platform and is used as a ground truth for further analysis. The audio interface is subsequently connected to a laptop via a USB port. To ensure low latency from hardware, ASIO driver[2] is installed from the internet.

[2] https://www.asio4all.org/.

Fig. 3 Hardware setup for experiments with single channel microphone-loudspeaker

Fig. 4 Hardware setup for experiments with multi-channel microphones organized in a uniform circular array with a loudspeaker placed at the center of the array

Moreover, MATLAB is used as a data acquisition software tool to record and save the observed signals and for statistical analysis of the proposed method. Furthermore, for multichannel data acquisition, PlayRec [26] is used to transmit and record sound simultaneously. The sampling frequency is set to 48, 000 Hz while the speed of sound is assumed as 343 m/s

4.2 Simulated and real results

In the first experiment, the non-ideal characteristic of acoustic systems is modeled by filtering the room impulse response, h_{RIR} using a bandpass filter with the impulse response, h_{BP}, to obtain our non-ideal impulse response, h_{NI}, i.e.,

$$h_{\text{NI}} = h_{\text{RIR}} * h_{\text{BP}}. \tag{23}$$

The bandpass filter was a second-order Butterworth filter with cutoff frequencies, $\boldsymbol{\omega} = [0.2\pi, 0.6\pi]$. The non-ideal room impulse response was then applied to a known probe signal, $s(n)$, to generate the observation used for the experiment. Here, the search interval for the delays, or TOAs, was chosen as $\tau \in [1, 80]$ samples, and therefore we set N to 2, 080. The number of reflections was set to $R = 3$ because this number gives us better estimates of 2 acoustic reflectors, the number of EM iterations was set to 100, and $\beta_r = 1/R$. Furthermore, the direct-path component was removed from the observed signal using an RIR generator. Using this setup, we ran the Ideal-EM (EMI) method with a filter length $M = 1$ as proposed in [19], and the presented robust-EM method (EMR) with filter length $M = 5$ and $\lambda = 100$. The resulting cost functions, $J(\mathbf{g}, \tau)$ from (19), are depicted in Figs. 5 and 6, respectively. Here, J_1, J_2, and J_3 represent the cost function with $M = 1, \lambda = 0$, $M = 5, \lambda = 100$, and $M = 15, \lambda = 500$, respectively. From the results, we can first see how the ideal impulse responses are affected by the bandpass filter applied to it, which smears out the peaks. When applying the EMI method, we therefore also do not see two clearly defined peaks around the time-of-arrivals of the two components. If we instead use the EMR method, we can model the effects of the bandpass filter, which results in two broader, but clearly defined peaks at the TOA.

Furthermore, we repeat the simulated experiment in a practical setting using the hardware platform in Fig. 3. The platform was placed at a corner of a room with a distance to the walls, 1 m and 0.65 m, respectively. The collocated microphone-loudspeaker setup probes the environment with a known sound, and the received echoes are recorded by the microphone. The observed signal was later used to estimate the RIR of the environment using the dual-channel method [27]. This is done by computing $\widehat{H}(f) = Y(f)/S(f)$ and then taking the inverse DFT to get $\widehat{h} = \mathcal{F}^{-1}\{\widehat{H}(f)\}$. The EMR's filter length was set to $M = 15$, $\lambda = 500$, and $R = 3$. As seen in Fig. 7, the EMR method successfully estimates all the peaks corresponding to an individual acoustic reflector. In this experiment, both M and λ are set empirically. However, in the future iteration of this work, we can adaptively select these parameters.

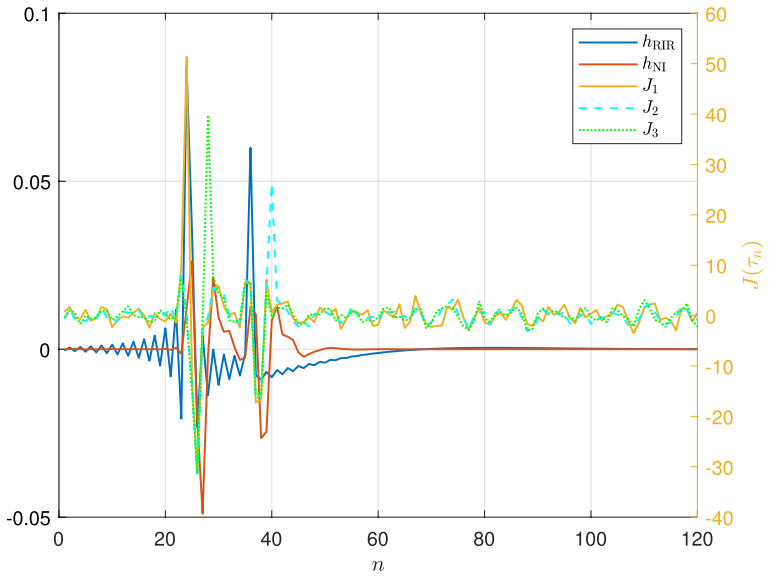

Fig. 5 Cost functions of the M-step for $M = 1$ using the EMI method in [19]

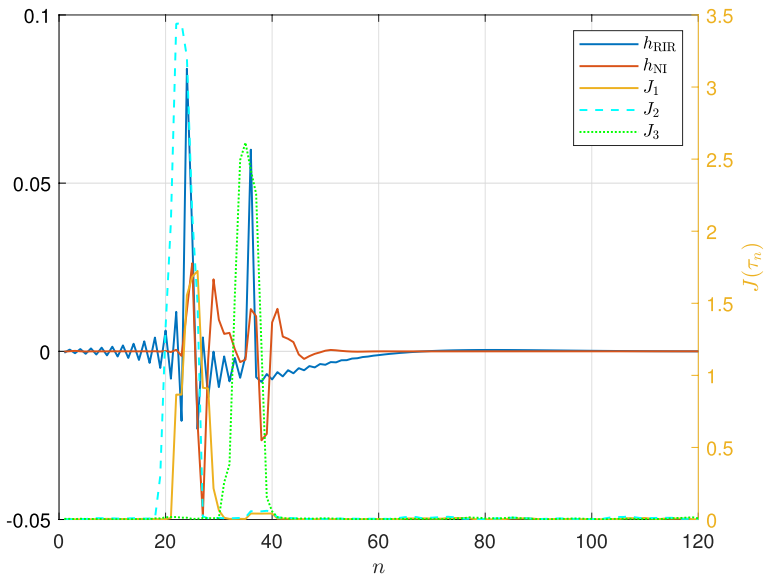

Fig. 6 Cost functions of the M-step for $M = 5$ and $\lambda = 100$ using the proposed method (EMR)

4.3 Impact of distances and background noises

In this experiment, we evaluate the performance of the proposed TOA estimator and compare it against varying distances. The setup was placed at a distance of [0.8, 1.0, 1.5, 2.0, 2.5] m, and 100 acoustic echoes were recorded at each interval. The data was collected using the single channel setup shown in Fig. 3. Accuracy is defined as the percentage of TOA that is within ±10% of the ground truth value obtained from the lidar. The proposed method (EMR) is compared with the previous method (EMI) proposed by [19] and single-channel localization and mapping (ScLAM) [28]. These results are shown in Fig. 8. The data obtained from this experiment is also summarized in Table 1.

Additionally, a comparison of the proposed method against different background noise was also performed. To simulate different noise levels, a separate loudspeaker was placed at a distance of 6.4 m away from the setup within the lab. This separate loudspeaker was used to simulate a low signal-to-noise ratio (SNR). The separate

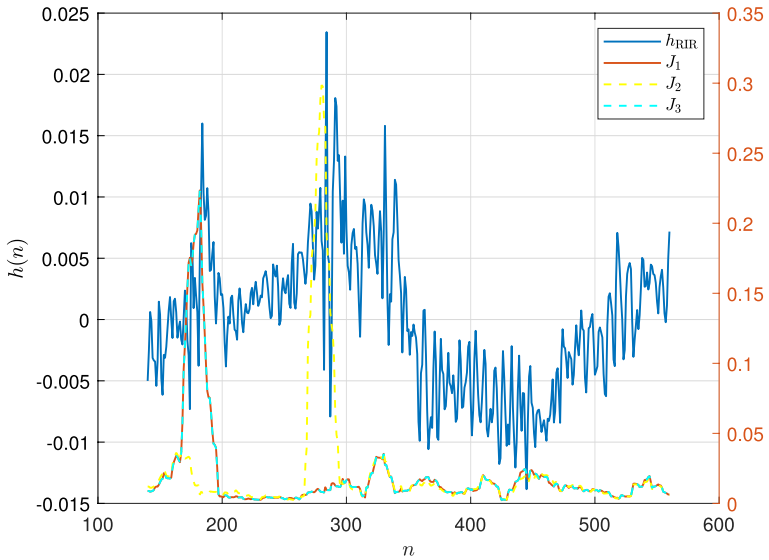

Fig. 7 Estimating multiple acoustic echoes using real data obtained from hardware platform in Fig. 3a

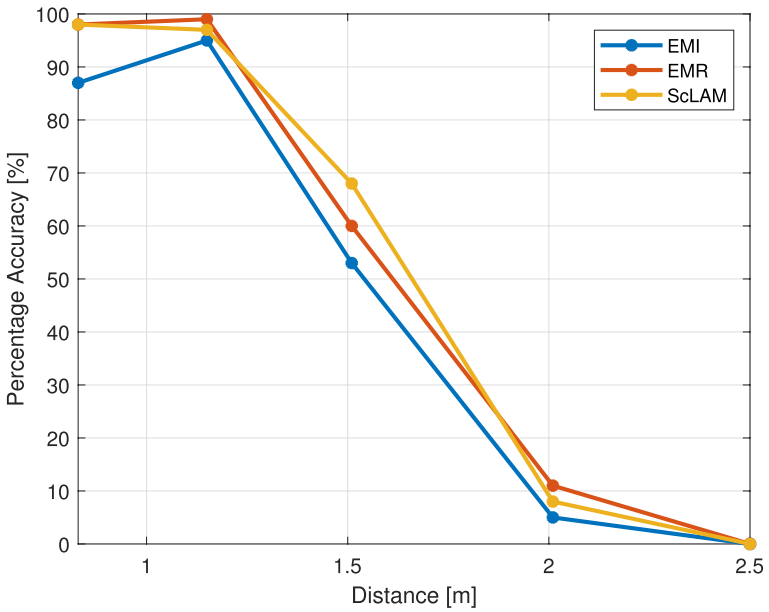

Fig. 8 Comparison of the proposed method robust EM with $M = 5$ and $\lambda = 100$ against ideal EM $M = 1$ for acoustic reflector estimation at varying distances

loudspeaker is playing an audio clip from YouTube called cocktail party[3]. The SNR is defined as the variance of the observed signal, $\mathbf{x}(n)$, against the variance of the background noise, $\mathbf{v}(n)$.

$$\mathrm{SNR} = \frac{\sigma_x^2}{\sigma_v^2}, \tag{24}$$

where $\sigma_x^2 = E[\|\mathbf{x}(n)\|^2]$ and $\sigma_v^2 = E[\|\mathbf{v}(n)\|^2]$. Both the observed signal and the background noise are recorded for 1 s. The background noise was recorded before the system probed the environment with a known signal. Based on this configuration, 4 SNRs were selected by adjusting the loudness of the separate speaker, [0, 10, 20, 30] dB. Furthermore, 100 audio recordings were obtained at each SNR to evaluate the proposed method (EMR). The evaluation results are shown in

[3] https://youtu.be/IKB3Qiglyro.

Fig. 9. According to Table 1, both the standard deviation σ and root mean square error (RMSE) of the EMI and EMR increases when the distance between the acoustic reflector and the platform increases while the mean value μ is close to the ground truth for a distance up to 1.5 and for all SNRs.

4.4 Evaluation of robust EM using multilateration technique

In this experiment, we test the performance of the proposed method using multilateration technique. In this way, we can estimate the DOA of the acoustic echoes which can aid robotic platforms to locate the source of

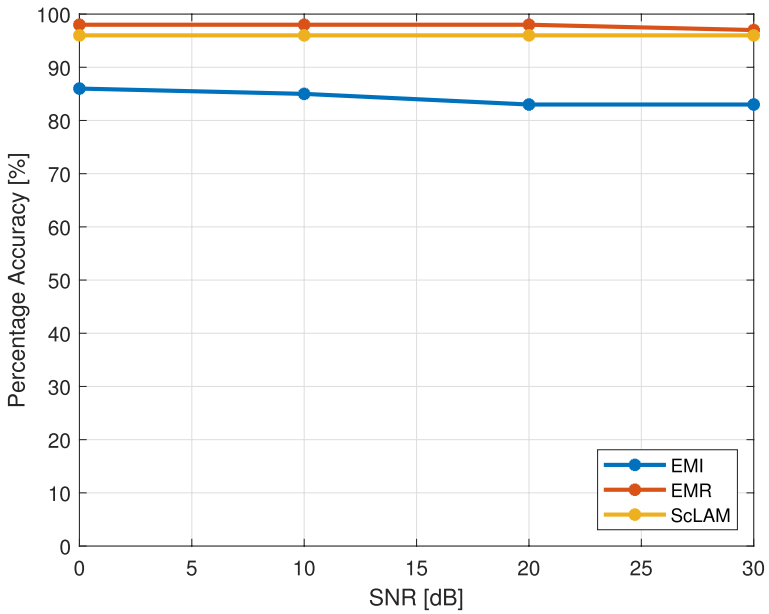

Fig. 9 Comparison of the proposed method robust EM $M = 5$ and $\lambda = 100$ against ideal EM with $M = 1$ for acoustic reflector estimation against different background noise

Table 1 Comparison of EMI against the other TOA estimation methods under different distances and background noise

	EMI SNR = 30 dB			**EMI SNR = 0 dB**		
Lidar data [m]	**μ [m]**	**σ [m]**	**RMSE [m]**	**μ [m]**	**σ [m]**	**RMSE [m]**
0.83	**0.8886**	**0.0403**	**0.0710**	**0.8856**	**0.0436**	**0.0704**
1.15	1.1306	0.1274	0.1282	1.1151	0.1108	0.1156
1.51	1.4185	0.2522	0.2671	1.4288	0.2739	0.2844
2.01	1.2356	0.2772	0.8221	1.2348	0.2689	0.8201
	EMR $M = 5$ $\lambda = 100$ **SNR = 30 dB**			**EMR** $M = 5$ $\lambda = 100$ **SNR = 0 dB**		
Lidar data/m	**μ [m]**	**σ [m]**	**RMSE [m]**	**μ [m]**	**σ [m]**	**RMSE [m]**
0.83	**0.8734**	**0.0105**	**0.0447**	**0.8703**	**0.0233**	**0.0464**
1.15	1.0772	0.0252	0.0769	1.0705	0.0246	0.0831
1.51	1.4370	0.2585	0.2674	1.4541	0.2549	0.2597
2.01	1.2379	0.3434	0.8443	1.2837	0.3531	0.8067
	ScLAM = 30dB			**ScLAM = 0dB**		
Lidar data [m]	**μ [m]**	**σ [m]**	**RMSE [m]**	**μ [m]**	**σ [m]**	**RMSE [m]**
0.83	**0.8826**	**0.0059**	**0.0709**	**0.8796**	**0.0214**	**0.0704**
1.15	1.0977	0.0871	0.1281	1.0776	0.0395	0.1156
1.51	1.4789	0.2301	0.2670	1.5312	0.2245	0.2843
2.01	1.2658	0.3276	0.8221	1.2648	0.3197	0.8200

the acoustic echoes. The idea here is that the proposed method will estimate TOAs from each of the microphone-loudspeaker combinations, which will then be used with a multilateration technique. Multilateration is a localization technique popularly used in telecommunication to estimate the direction and distance of a transmitter/source [29–31]. Moreover, multilateration was also used to estimate the robot's position in 3D space as proposed in [32]. Within the context of this paper, multilateration is used to estimate the location of the acoustic reflector. Multilateration techniques rely on the TOAs' knowledge of the acoustic reflections and also assume that the locations of the sensor nodes are known with respect to the same coordinate system. To locate an acoustic reflector, we need to set a reference with respect to a coordinate system. This information could be known from the robot's motor encoder or from an inertial measurement unit (IMU) but this aspect of robot navigation is beyond the scope of this paper. More specifically, let us assume that we have P microphones and the source is placed on the same xy-plane. Using (17), we can estimate the TOA and (22), the range value vector, $\mathbf{d}$. If the microphones are located on the xy-plane or 2D plane, at positions, $[\mathbf{x}_p, \mathbf{y}_p] = [(x_1, y_1), (x_2, y_2), \ldots, (x_P, y_P)]$, where P are the number of microphones, then based on the range data $\mathbf{d}_p$ a circle can be drawn from each microphone. The point of intersection of these individual circles would yield the location of the acoustic reflector as seen in Fig. 10. The true acoustic reflector position (x, y) is at the intersection of all the circles and satisfies the following equations:

$$(x - x_p)^2 = d_p^2, \quad p = 1, \cdots, P. \tag{25}$$

In the presence of noise, the estimations of $\mathbf{d}$, the circles will not intersect at a single point. Therefore, a least-square fit can be used to obtain the acoustic reflector location estimate [33], i.e.,

$$\mathbf{r}_s = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}, \tag{26}$$

where

$$\mathbf{A} = \begin{bmatrix} 2(x_1 - x_P) & 2(y_1 - y_P) \\ \vdots & \\ 2(x_{P-1} - x_P) & 2(y_{P-1} - y_P) \end{bmatrix} \tag{27}$$

$$\mathbf{b} = \begin{bmatrix} x_1^2 - x_P^2 + y_1^2 - y_P^2 + d_P^2 - d_1^2 \\ \vdots \\ x_{P-1}^2 - x_P^2 + y_{P-1}^2 - y_P^2 + d_P^2 - d_{P-1}^2 \end{bmatrix} \tag{28}$$

The setup used for this experiment is shown in Fig. 4. Here, the setup was fixed at distances [0.7, 1.1, 1.5] m against an acoustic reflector. Furthermore, 50 recordings were made at each distance which was later evaluated. The results are depicted in Fig. 11 and listed in Table 2. According to Table 2, the σ and RMSE values of the proposed method increase as the platform's distance with respect to the wall also increases while μ value is close to 0.7 m at an SNR of 30.

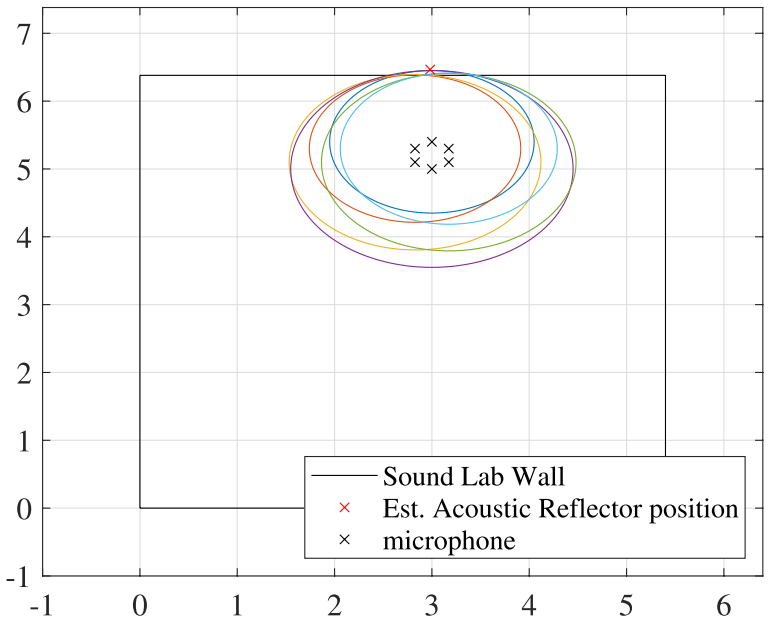

Fig. 10 EMR and multilateration technique to localize an acoustic echo situated at a distance of 0.7 m. The convergence of the individual circles indicates the location of the acoustic reflectors

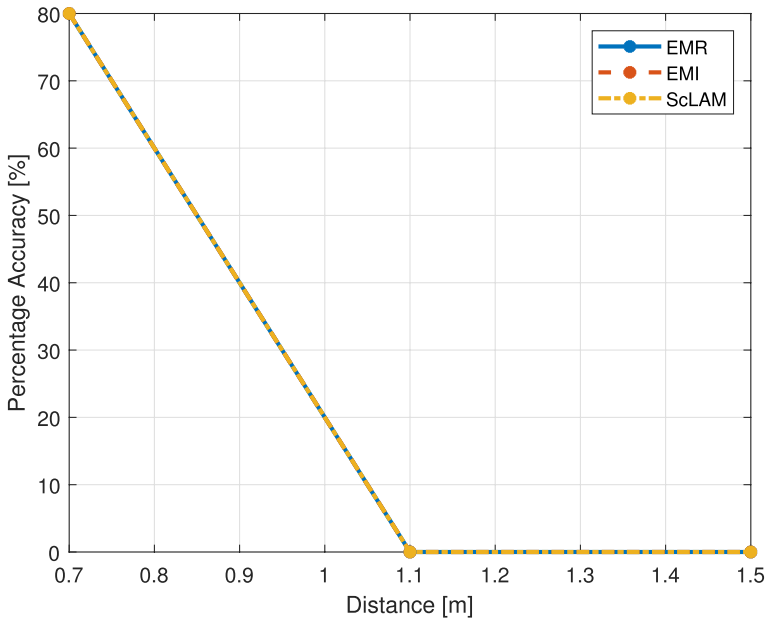

Fig. 11 Evaluation of the proposed method with multilateration to detect a single acoustic reflector

Table 2 Performance of the proposed method using multilateration technique evaluated over distances

	EMI SNR = 30			EMR SNR = 30			ScLAM SNR = 30		
Lidar data [m]	μ[m]	σ[m]	**RMSE [m]**	μ[m]	σ[m]	**RMSE [m]**	μ[m]	σ[m]	**RMSE [m]**
0.7	**0.6240**	**0.1442**	**0.1617**	**0.6154**	**0.15383**	**0.16176**	**0.65628**	**0.072963**	**0.08443**
1.1	0.8428	0.0689	0.2660	0.77155	0.058971	0.26605	0.77155	0.058971	0.3336
1.5	1.1686	0.3247	0.4617	3.1354	0.18567	1.9132	1.6851	1.5701e-15	0.18509

5 Discussion and limitations

Two platform designs were proposed to test the algorithm: A collocated microphone-loudspeaker as seen in Fig. 3 and a uniform circular microphone array with a loudspeaker positioned at the center of the array as seen in Fig. 4. The results obtained from the first experiment revealed that the proposed method can be used to estimate multiple acoustic reflections as EMR can account for the acoustic system's response which can hinder the estimation accuracy of multiple acoustic reflections. As seen in Fig. 6, EMR estimates multiple peaks that correspond to an acoustic reflectorm, while EMI (Fig. 5) estimates a single acoustic reflector. Therefore, estimating multiple acoustic reflectors using the proposed method is beneficial for spatial map construction in an indoor environment.

In the second experiment, the performance of EMR and EMI are evaluated using the proof-of-concept setup described in Section 4.1. The results in Fig. 8 reveal that EMR provides significant improvements in estimating the acoustic reflector as it can account for the acoustic system's response that affects the performance of the TOA estimator, while Fig. 9 shows that the proposed method is 10% better than the EMI method overall SNR values which are on par with the ScLAM techniques. According to the results obtained in Fig. 8, the proposed method can estimate an acoustic reflector up to a distance of 1.5 m with 60% accuracy under low SNR of 0 dB. Similarly, the proposed method is robust against different SNR levels as seen in Fig. 9 compared to EMI. The results obtained from Table 1 shows that the proposed method offers a limited range as it estimates the acoustic reflector's range up to a distance of 1.5 m with an RMSE of 0.2671 m at a high SNR value of 30 dB. Under low SNR value of 0 dB, the μ, σ, and RMSE remain similar which indicates that the proposed method is robust under changing environmental conditions.

In the last experiment, we combined the proposed method with a multilateration technique so that the direction, as well as the location of the acoustic reflector, is determined by a robotic system as it navigates an indoor environment. Here, we test EMI, EMR, and

ScLAM under an SNR of 30 dB and place the multichannel setup at varying distances. According to the results obtained in Fig. 11, all methods can estimate an acoustic reflector up to a distance of 0.7 m with 80% accuracy. The results obtained in Table 2 also indicates that the μ, σ and the RMSE are similar for all 3 methods (EMI, EMR and ScLAM). The μ value is around 0.6154 m while the RMSE value is 0.16176 m when the setup is placed at a distance of 0.7 m. The μ and RMSE values increase as the distance between the wall and the setup increases to 1.1 m and 1.5 m. This reduction in accuracy could be due to the loudspeaker blocking the acoustic echoes from reaching one of the microphones placed behind the loudspeaker which could affect the TOA estimation. This could result in spurious estimates that can reduce the performance of the multilateration technique when locating an acoustic source. Similar performance is seen in the remaining methods. However, for multilateration technique to work within robotics, the robotic platform requires the knowledge of its Cartesian position in the environment, i.e., the position of the loudspeaker and microphones should be known. One way to acquire this information is by utilizing sensors used for tracking the odometry and orientation of a robot, e.g., the inertial measurement unit. However, in this paper, we assume that the location of the loudspeaker and microphones will be known.

6 Conclusions and future work

The contribution of this paper is to propose a robust expectation-maximization technique for acoustic reflector localization, intended for the robotic platform using echolocation. The proposed method builds on existing work proposed by [19], i.e., their work assumed that the gain or filter parameters are assumed to be the same which in practice is not a valid assumption as this can hinder the acoustic reflector estimation process. Hence, in this paper, we introduced this uncertainty within the signal formulation. Three experiments were performed in a simulated and practical environment. To test the performance of the proposed method, two proof-of-concept platforms are used: one consists of a collocated microphone-loudspeaker arrangement while the other consists of a uniform circular microphone array with a loudspeaker placed at the center of an array. From our experimental results, we deduce that our proposed method can estimate an acoustic reflector up to a distance of 1.5 m with 60% accuracy and can be combined with a multilateration technique to locate the direction of an acoustic reflector. Our proposed method can be beneficial to the robotic platforms as it can complement existing laser- and camera-based technologies for generating a spatial map of an indoor environment as done in our previous works. Our proposed echolocation method can aid a robotic platform in detecting and estimating transparent surfaces and can also estimate multiple acoustic echoes when a robot moves to a corner of a room.

In the future iteration of this work, we aim to implement the proposed method on an existing robotic platform, e.g., Softbank's NAO robot, and also improve the algorithm and combine it with eco-labeling techniques as proposed in [21] so that multiple acoustic echoes are estimated and categorized to represent an indoor environment. We also intend to test the proposed method using the robotic platform outlined in [28]. This way, we can test the performance of the proposed method against the ScLAM and McLAM algorithms and also evaluate the performance in generating a spatial map of a typical office environment. The current proof-of-concept is a fixed loudspeaker-microphone setup, while in [28], the setup is placed on top of a robotic platform that moves within an indoor environment. Moreover, this method could also be used in a wireless acoustic sensor network (WASN) to detect acoustic sources [28, 34].

Abbreviations

TOA	Time-of-arrival
EM	Expectation-maximization
UCA	Uniform circular array
SNR	Signal-to-noise ratio
DOA	Direction-of-arrival
aSLAM	Acoustic simultaneous localization and mapping
RIR	Room impulse response
TDOA	Time difference-of-arrival
ML	Maximum likelihood
T_{60}	Reverberation time (60 dB)
RPM	revolutions per minute
DREGON	Database of drone audio recordings
NLS	nonlinear least squares

Acknowledgements
Not applicable.

Authors' contributions
JRJ, MGC, and US designed the idea for the manuscript. JRJ and US conducted the experiments. All the authors contributed to the writing of this work. Moreover, all author(s) read and approved the final manuscript.

Funding
This work was funded by Aalborg Unversity, Denmark.

Availability of data and materials
Not applicable.

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 17 November 2022 Accepted: 22 March 2024
Published online: 18 April 2024

References

1. J. Steckel, H. Peremans, BatSLAM: Simultaneous localization and mapping using biomimetic sonar. PLoS ONE **8**(1), 1–11 (2013)
2. R. Kuc, Echolocation with bat buzz emissions: Model and biomimetic sonar for elevation estimation. J. Acoust. Soc. Am. **131**(1), 561–568 (2012)
3. M. Kreković, I. Dokmanić, M. Vetterli, EchoSLAM: Simultaneous localization and mapping with acoustic echoes, Proc. IEEE Int. Conf. Acoust., Speech, Signal Process, IEEE, pp. 11–15 (2016)
4. S. Tervo, J. Pätynen, T. Lokki, Acoustic reflection localization from room impulse responses. ACTA Acustica U. Acustica **98**(3), 418–440 (2012)
5. G. Defrance, L. Daudet, J.D. Polack, Detecting arrivals within room impulse responses using matching pursuit, Proc. of the 11th Int. Conference on Digital Audio Effects (DAFx-08), Espoo, Finland. vol. 10, pp. 307–316 (2008)
6. G. Defrance, L. Daudet, J.D. Polack, Using matching pursuit for estimating mixing time within room impulse responses. Acta Acustica U. Acustica **95**(6), 1071–1081 (2009)
7. G. Moschioni, A new method for measurement of early sound reflections in theaters and halls, Proceedings of the 19th IEEE Instrumentation and Measurement Technology Conference (IEEE Cat. No.00CH37276), IEEE, vol. 1, pp. 425–430 (2002)
8. U. Saqib, S. Gannot, J. Jensen, Estimation of acoustic echoes using expectation-maximization methods. EURASIP J. Audio Speech Music. Process. **2020**(1), 1–15 (2020)
9. Y. Geng, J. Jung, Sound-source localization system for robotics and industrial automatic control systems based on neural network, 2008 International Conference on Smart Manufacturing Application, IEEE, pp. 311–315 (2008)
10. S. Dey, S. Boppu, M.S. Manikandan, Design of a real-time automatic source monitoring framework based on sound source localization, 2019 Seventh International Conference on Digital Information Processing and Communications (ICDIPC), IEEE, pp. 35–40 (2019)
11. H. Zhu, H. Wan, Single sound source localization using convolutional neural networks trained with spiral source, 5th International Conference on Automation, Control and Robotics Engineering (CACRE), IEEE, pp. 720–724 (2020)
12. N. Riopelle, P. Caspers, D. Sofge, Terrain classification for autonomous vehicles using bat-inspired echolocation, 2018 International Joint Conference on Neural Networks (IJCNN), IEEE, pp. 1–6 (2018)
13. J.H. Christensen, S. Hornauer, S.X. Yu, BatVision: Learning to see 3D spatial layout with two ears, IEEE International Conference on Robotics and Automation (ICRA), IEEE, pp. 1581–1587 (2020)
14. E. Tracy, N. Kottege, Catchatter: Acoustic perception for mobile robots. IEEE Robot. Autom. Lett. **6**(4), 7209–7216 (2021)
15. D.W. Gunness, Loudspeaker transfer function averaging and interpolation. J. Audio Eng. Soc. (2001)
16. U. Saqib, J.R. Jensen, Sound-based distance estimation for indoor navigation in the presence of ego noise, Proc. 27th European Signal Processing Conf. (EUSIPCO), IEEE, pp. 1-5 (2019)
17. P. Ahgren, P. Stoica, A simple method for estimating the impulse responses of loudspeakers. IEEE Trans. Consum. Electron. **49**(4), 889–893 (2003)
18. Z. Sü, M. Çalışkan, Acoustical design and noise control in metro stations: Case studies of the ankara metro system. Build. Acoust. **14**(3), 203–221 (2007)
19. M. Feder, E. Weinstein, Parameter estimation of superimposed signals using the em algorithm. IEEE Trans. Acoust. Speech Signal Process. **36**(4), 477–489 (1988)
20. A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the em algorithm. J. R. Stat. Soc. Ser. B (Methodol.) **39**(1), 1–22 (1977)
21. I. Dokmanic, R. Parhizkar, A. Walther, Y.M. Lu, M. Vetterli, Acoustic echoes reveal room shape. Proc. Natl. Acad. Sci. **110**(30), 12186–12191 (2013)
22. L. Nguyen, J.V. Miro, X. Qiu, Can a robot hear the shape and dimensions of a room?, International Conference on Intelligent Robots and Systems (IROS), IEEE, pp. 5346–5351 (2019)
23. M. Boutin, G. Kemper, Can a ground-based vehicle hear the shape of a room?. Studies in Applied Mathematics. **151**(1), 352-368 (2023)
24. E.A.P. Habets, I. Cohen, S. Gannot, Generating nonstationary multisensor signals under a spatial coherence constraint. J. Acoust. Soc. Am. **124**(5), 2911–2917 (2008)
25. U. Saqib, J. Jensen, A model-based approach to acoustic reflector localization using robotic platform,in Proc. IEEE Int. Conf. Intell., Robot, Automation (IROS), IEEE, pp. 1–8 (2018)
26. R. Humphrey, Playrec: Multi-channel MATLAB audio. (2007). http://www.playrec.co.uk. Accessed Mar 2001
27. H. Herlufsen, Dual channel FFT analysis (part I), Brüel & Kjær Technical Review. (1984)
28. U. Saqib, J.R. Jensen, A framework for spatial map generation using acoustic echoes for robotic platforms. Robot. Auton. Syst. **150**, 104009 (2022)
29. J. Yang, H. Lee, K. Moessner, Multilateration localization based on singular value decomposition for 3D indoor positioning, Int. Conf. Indoor Positioning and Indoor Navigation, IEEE, pp. 1–8 (2016)
30. J. Wan, N. Yu, R. Feng, Y. Wu, C. Su, Localization refinement for wireless sensor networks. Comput. Commun. **32**(13), 1515–1524 (2009)
31. Y. Zhou, Jun Li, L. Lamont, Multilateration localization in the presence of anchor location uncertainties, IEEE Global Communications Conference (GLOBECOM), IEEE, pp. 309–314 (2012)
32. A. Yazici, U. Yayan, H. Yücel, An ultrasonic based indoor positioning system, Int. Symposium on Innovations in Intell. Sys. and Applications, IEEE, pp. 585–589 (2011)
33. C. Chen, K. Yao, in *Classical and Modern Direction-of-Arrival Estimation*, ed. by T.E. Tuncer, B. Friedlander. Source and node localization in sensor networks (Academic Press, Boston, 2009), pp. 343–383
34. M. Cobos, F. Antonacci, A. Alexandridis, A. Mouchtaris, B. Lee, A survey of sound source localization methods in wireless acoustic sensor networks. Wirel. Commun. Mob. Comput, pp. 1-24 (2017)

Publisher's Note