Robust Acoustic Reflector Localization for Robots

Usama Saqib, Mads Græsbøll Christensen, Jesper Rindom Jensen

Robust Acoustic Reflector Localization for Robots

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

***Abstract*— The transfer function or response of an acoustic system, e.g., loudspeakers, contributes to a phase lag due to propagation delay that can hinder the performance of a time-of-arrival (TOA) estimator intended for acoustic reflector localization using echolocation. In this paper, we propose a Robust Expectation-Maximization (EM) algorithm that takes into account the response of acoustic systems to enhance the TOA estimation accuracy. 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. Moreover, we also show that the proposed method can also be used to estimate multiple acoustic echoes that originate when a robot is placed at a corner of a room.**

I. 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 to construct a spatial map of an indoor environment. This can be accomplish by a collocated microphone-loudspeaker combination. This enables a robotic platform to detect obstacles and acoustic reflectors such as transparent surfaces that are difficult to detect using laser- and camera-based technologies. Methods for localizing an acoustic reflector using echolocation on robotic platforms for spatial map construction have been proposed in our previous work [4]–[6], which is later implemented on a proof-of-concept robotic platform [7].

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 aid a robot to estimate the distance between the acoustic reflector and itself. Traditionally, TOA information is extracted from room impulse response (RIR) estimates which is normally done using a peak-picking approach [2], [3], [8]–[10]. However, in our previous work [4]–[6], we proposed estimating TOAs directly from the observed signal, by exploiting a model of sound reflections. This model is broadly divided into two distinct parts: the direct-path including early reflections and late reflections which are comprised by a stochastic dense tail [11]. 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 early reflections helps in inferring the distance of the closest acoustic reflector [2], [3], [6]. While, TOA estimation enables a robot to determine the distance of an acoustic reflector, 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 [12]–[14].

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 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. Therefore, we propose an algorithm that utilize the existing loudspeaker-microphone setup to estimate the distance of an acoustic reflector [7]. However, the loudspeaker's and microphone's response was not taken into account when deriving a TOA estimator. In practice, loudspeakers/microphones are non-ideal. One reason is the transfer function of the acoustical systems, e.g., loudspeakers/microphones, contributes to a phase lag due to propagation delay [15]. This can lead to model-mismatch and have a detrimental impact on the performance of the TOA estimation method proposed in [4], [6] and may hinder estimation of multiple acoustic echoes. Estimating the transfer function of the loudspeaker is usually done using a loudspeaker enclose microphone (LEM) setup which involves placing the setup within an anechoic environment. However, in [16], 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 off, e.g., concrete, glass, and cardboards. This is because, some materials absorb certain sound frequencies that could lead to non-ideal characteristics of the observed signals [17].

In this paper, we therefore extend the method originally proposed in [18] and later used in our previous work [6] 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

[1]The authors are with the Audio Analysis Lab, CREATE, Aalborg University, Denmark {ussa,mgc,jrj}@create.aau.dk

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 [16]. Moreover, to test the proposed method, a proof-of-concept setup is built to conduct experiments using real data. The proposed method and proof-of-concept setup enable estimating the acoustic echoes.

The remaining part of this paper is organized as follows: Section II introduces the problem formulation, Section III proposed the TOA estimation method based on expectation-maximization (EM) while Section IV introduces a multilateration technique that is used to estimate the DOA of the acoustic reflector. Additionally, the proof-of-concept setup used to test the proposed method is discussed in Section V-A while the experimental results followed by discussion and conclusion can be found in Section V, VI and VII, respectively.

II. 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

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

where h 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 $x(n) = h*s(n)$. The acoustic impulse response can be further modelled by decomposing the reverberation into early and late reverberation components. The early reflections are modelled 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 rth 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 g_r is a dirac function for all $r = 1, \ldots, R$, we get the ideal model used in [6], 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.,

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

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

$$\mathbf{G}_r = \begin{bmatrix} g_{M-1,r} & \cdots & g_{0,r} & 0 & 0 & \cdots \\ 0 & \cdots & g_{1,r} & g_{0,r} & 0 & \cdots \\ & & \vdots & \vdots & & \\ \cdots & 0 & g_{M-1,r} & g_{M-2,r} & \cdots & g_{0,r} \end{bmatrix}, \tag{5}$$

$$\mathbf{S}(n-\tau) = \begin{bmatrix} s(n-\tau+M-1) & \cdots & s(n-\tau) \\ s(n-\tau+M) & \cdots & s(n-\tau+1) \\ \vdots & & \vdots \\ s(n-\tau+N-1) & \cdots & s(n-\tau+N-M) \end{bmatrix}, \tag{6}$$

$$\mathbf{s}(n-\tau) = [s(n-\tau), s(n-\tau+1), \cdots, s(n-\tau+N-1)]^T, \tag{7}$$

$$\mathbf{g}_r = [g_{0,r}, g_{1,r}, \cdots, g_{M-1,r}]^T. \tag{8}$$

Here, the matrix $\mathbf{G}_r$ has a dimension of $(N - M + 1) \times N$ while $\mathbf{S}_r$ 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

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

Compared to [18], we do not assume that the gain or filter $\mathbf{g}_r$ is the same. Hence, the problem at hand is to estimate the delay τ_r and the filter $\mathbf{g}_r$ parameters. Moreover, in this paper, we are interested to estimate these parameter to localize the position of an acoustic reflector which was not addressed in [18]. 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).

III. ROBUST EM-BASED ACOUSTIC REFLECTOR LOCALIZATION

Iterative estimation schemes such as the expectation-maximization (EM) algorithm may be employed to alleviate the complexity. 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

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

(a) Hardware setup for experiments with single channel microphone-loudspeaker

(b) Hardware setup for experiments with multi channel microphones organized in a uniform circular array with a loudspeaker placed at the center of the array

Fig. 1: Proof of concept used to obtained acoustic data

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 $\mathcal{N}(\mathbf{0}, \beta_r \mathbf{C})$. Moreover, the scaling factors, β_r, are non-negative, real-valued scalars, that satisfy

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

With these assumptions, the EM algorithm for the problem at hand is given by *E-step:*

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

M-step:

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

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

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

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

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

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:

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

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

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

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

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

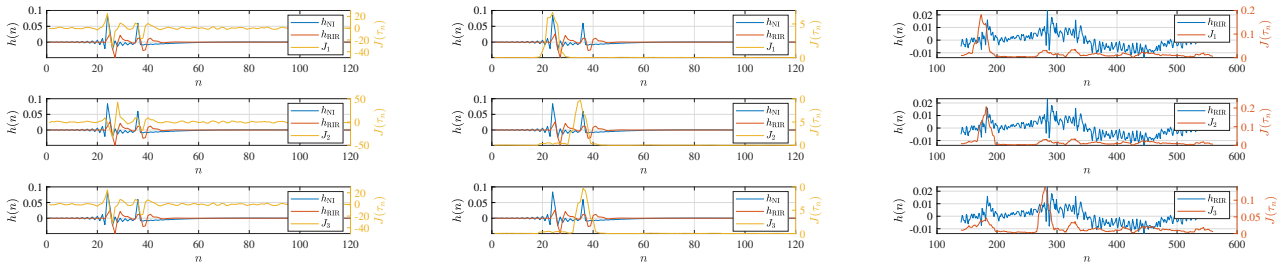
That is, the filter estimate becomes

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

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 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:

$$d = \frac{c \times \tau}{2}, \tag{22}$$

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

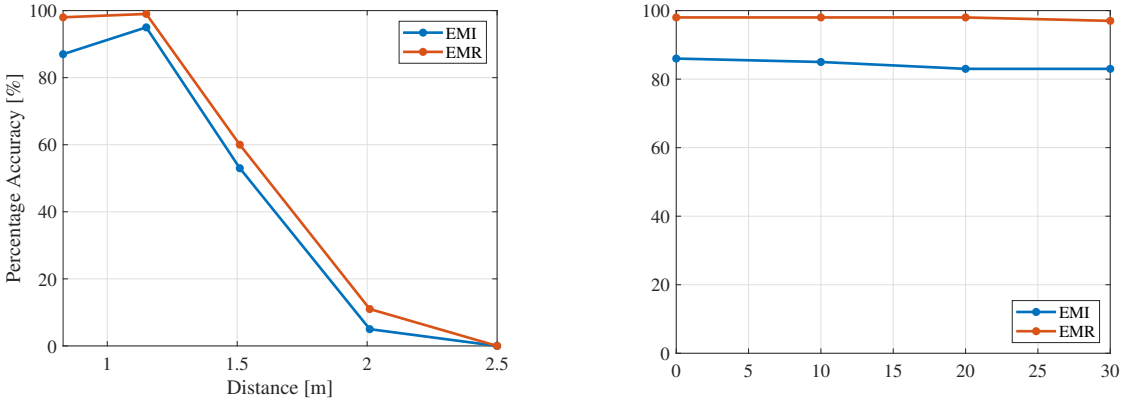


(a) Cost functions of the M-step for $M = 1$ using the EMI method in [18]

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

(c) Estimating multiple acoustic echoes using real data obtained from hardware platform in Fig. 1(a)

Fig. 2: Comparison of the EMI and EMR method to estimate individual peaks. EMI fails to estimate the individual peaks corresponding to individual acoustic reflections while EMR correctly estimates the acoustic reflections



(a) 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

(b) 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.

Fig. 3: Comparison of the EMI and EMR method under different distances and background noise

IV. Multilateration technique for DOA estimation

The proposed method discussed in Section III uses a collocated microphone-loudspeaker setup to estimate the distance of an acoustic reflector. However, the proposed method could also be extended by incorporating multiple microphones. In this way, we can estimate the DOA of the acoustic echoes which can aid robotic platforms to locate the source of 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 techniques popularly used in telecommunication to estimate the direction and distance of a transmitter/source [19]–[21]. Moreover, multilateration was also used to estimate robot's position in 3D space as proposed in [22]. Within the context of this paper, multilateration is used to estimated 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 robot's motor encorder or Inertial Mass Unit (IMU) but this aspect of robot navigation is beyond the scope of this paper. More specifically, let us assume that we have M microphones and the source is placed on the same xy-plane. Using (17), we can estimate the range value vector, $\mathbf{d}$]. If the microphones are located on the xy-plane or 2D plane, at positions, $[\mathbf{x}_m, \mathbf{y}_m] = [(x_1, y_1), (x_2, y_2), \ldots, (x_M, y_M)]$ then based on the range data $\mathbf{d}_m$ a circle can be draw from each microphones. The point of intersection of these individual circles would yield the location of the acoustic reflector as seen in Fig. 4(a). The true acoustic reflector position (x, y) is at the intersection of all the circles and satisfies the following equations:

$$(x - x_m) = d_m^2, \quad m = 1, \cdots, M. \tag{23}$$

In the presence of noisy estimations of $[\mathbf{d}]$, the circles will not intersect at a single point. Therefore, a least-square fit can used to obtain the acoustic reflector location estimate [23], i.e.,

$$\mathbf{r}_s = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}, \tag{24}$$

	EMI SNR = 30dB			EMI SNR = 0dB		
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
2.5	1.238	0.2832	1.2931	1.2361	0.2842	1.2952
	EMR $M = 5$ $\lambda = 100$ **SNR = 30dB**			**EMR** $M = 5$ $\lambda = 100$ **SNR = 0dB**		
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
2.5	1.2063	0.3129	1.3307	1.2483	0.3120	1.2896

TABLE I: Comparison of EMI against the proposed method EMR under different distances and background noise

where

$$\mathbf{A} = \begin{bmatrix} 2(x_1 - x_M) & 2(y_1 - y_M) \\ \vdots & \\ 2(x_{M-1} - x_M) & 2(y_{M-1} - y_M) \end{bmatrix} \tag{25}$$

$$\mathbf{b} = \begin{bmatrix} x_1^2 - x_M^2 + y_1^2 - y_M^2 + d_M^2 - d_1^2 \\ \vdots \\ x_{M-1}^2 - x_M^2 + y_{M-1}^2 - y_M^2 + d_M^2 - d_{M-1}^2 \end{bmatrix} \tag{26}$$

The multilateration approach is only suitable for localizing single reflector position because it relies on finding a single solution that stems from the intersection of all the circles based on the distances of the acoustic reflector from the individual microphones.

However, by taking the acoustic response within the model, we can estimate multiple reflections originating from two acoustic reflector, i.e., first-order and second-order reflection. By combining the proposed method with echolabeling, we can estimate the position of multiple acoustic echoes.

V. 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^3. The analysis window considered were set to τ_{min} and τ_{max} corresponding to a distance of 0.5 m to 3 m similar to the work performed in [7]. This analysis window also helps in estimating the first-order early reflection and prevents direct-path component from being estimated. Moreover, the probe signal $s(n)$ is a broadband signal of length $2,000$ samples drawn from a Gaussian burst with zero padding to form a signal of length $20,000$ samples.

A. Proof-of-concept

The experimental platform used to evaluate the performance of the proposed method. Two design variations are proposed to test the proposed method for 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 on 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. 1(a). The second variation consist of a 6 microphones 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. 1(b). 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^3. Furthermore, both the loudspeaker and microphones is 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 drivers[1] is installed from the internet. Moreover, MATLAB is used as a data acquisition software tool to record and save the observed signals and for statistical analysis on the proposed method. Furthermore, for multichannel data acquisition playRec [25] 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

B. Simulated and real Results

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

$$h_{\text{NI}} = h_{\text{RIR}} * h_{\text{BP}}. \tag{27}$$

The band pass 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 $R = 3$, 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 RIR generator. Using this setup, we ran the Ideal-EM (EMI) method with a filter length $M = 1$

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

as proposed in [18], and the presented Robust-EM method (EMR) with filter length $M = 5$. The resulting cost functions are depicted in Figures 2(a) and Fig.2(b), 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 practical setting using the hardware platform in Fig. 1(a). The platform was placed at a corner of a room with a distance to the walls, 1 m and 0.65m, 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 RIR of the environment using dual-channel method [26]. This is done by computing $\widehat{H}(f) = Y_1(f)/S(f)$ 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. 2(c), the EMR method successfully estimates all the peaks corresponding to individual acoustic reflector.

C. Impact of distances and background noises

In this experiment, we evaluate the performance of the proposed TOA estimator and compared 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. 1(a). The Accuracy is defined as the percentage of TOA that are within $\pm 10\%$ of the ground truth value obtained from lidar. The proposed method (EMR) is compared with previous method (EMI). The results are shown in Fig. 3(a). The data obtained from this experiment is also summarized in Table I. 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 low signal-to-noise ratio (SNR). The separate loudspeaker is playing an audio clip from Youtube called cocktail party[2]. The SNR is defined as the variance of the observed signal, $\mathbf{x}(n)$, against the variance of the background noise, $\mathbf{v}(n)$.

$$\text{SNR} = \frac{\sigma_x^2}{\sigma_v^2},$$

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 is recorded for 1 sec. 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 SNRs to evaluate the proposed method (EMR). The evaluation results are shown in Fig. 3(b).

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

	EMR SNR = 30dB		
Lidar Data [m]	**μ[m]**	**σ[m]**	**RMSE [m]**
0.7	0.6240	0.1442	0.1617
1.1	0.8428	0.0689	0.2660
1.5	1.1686	0.3247	0.4617

TABLE II: Performance of the proposed on multiple microphones against distances

D. Evaluation of Robust EM using multilateration technique

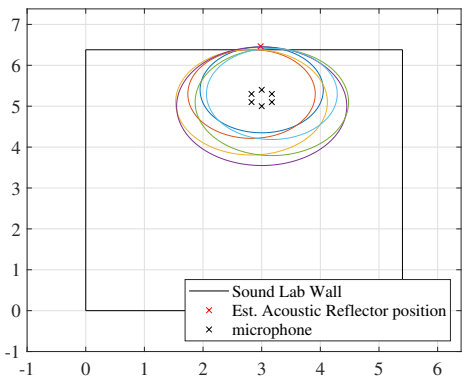
In this experiment, we test the performance of the proposed method using multilateration technique described in Section IV. The setup used for this experiment is shown in Fig.1(b). Here, the setup was fixed at distances $[0.7, 1.1, 1.5]$ m against an acoustic reflector. Furthermore, 50 recordings were made at each distance which was later evaluated. The results are depicted in Fig.4 and listed in Table II.

VI. Discussion and Limitations

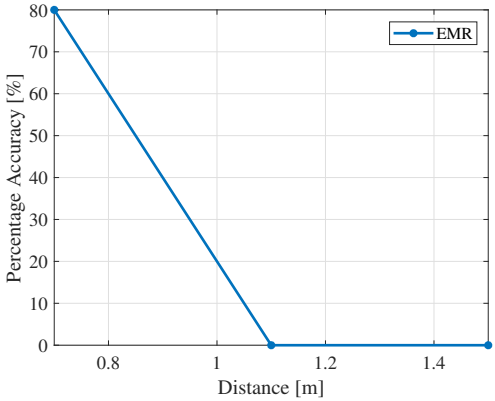
Two platform designs were proposed to test the algorithm: A collocated microphone-loudspeaker as seen in Fig. 1(a) and a uniform circular microphone array with a loudspeaker positioned at the center of the array as seen in Fig.1(b). The results obtained from the first experiment reveals that the proposed method can be used to estimate multiple acoustic reflections as EMR can account for acoustic system's response which can hinder the estimation accuracy of multiple acoustic reflections. As seen in Fig. 2, EMR estimates multiple peaks that corresponds to an acoustic reflectors while EMI estimates a single acoustic reflector. Estimating multiple acoustic reflectors is beneficial for spatial map construction typically in an office environment because often times, a robotic platform is closer to two acoustic reflectors.

In the second experiment, the performance of EMR and EMI is evaluated using the proof-of-concept described in Section V-A. The results in Fig. 2(a) reveals that EMR provides significant improvements in estimating the acoustic reflector as it can account for acoustic system's response that affects the performance of TOA estimator. According to the results EMR can estimate an acoustic reflector up to a distance of 1.5 m with 60% under low SNR of 0 dB. According to Table.I, 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 distance up to 1.5 and for all SNRs.

In the last experiment, we combined EMR with 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 EMR under a SNR of 30 dB and place the multi-channel setup at varying distances. According to the results obtained in Fig. 4(b), EMR can estimate an acoustic reflector up to a



(a) EMR and multilateration technique to localize an acoustic echo situated at a distance of 0.7 m



(b) Performance of EMR against varying distances

Fig. 4: EMR combined with multilateration technique

distance of 0.7 m with 80% accuracy. This reduction in accuracy could be the loudspeaker blocking the acoustic echoes from reaching one of the microphone that could affect the TOA estimation. This could results estimating spurious estimates that can reduce the performance of the multilateration technique when locating an acoustic source. Furthermore, according to Table II, the σ and RMSE values increases as the platform's distance with respect to the wall is also increase while μ value is close to 0.7 m at a SNR of 30. However, for multilateration technique to work, the robotic platform requires the knowledge of its Cartesian position of 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., Inertial Mass Unit. However, in this paper, we assume that the location of the loudspeaker and microphones will be known.

VII. CONCLUSIONS

The contribution of this paper is to propose a Robust Expectation-Maximization technique for acoustic reflector localization, intended for robotic platform using echolocation. The proposed method builds on existing work proposed by [18], 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 introduce this uncertainty within the signal formulation. Three experiments were performed in simulated and practical environment. To test the performance of the proposed method, two proof-of-concept platforms are used: One consist of a collocated microphone-loudspeaker arrangement while the other consist 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 multilateration technique to locate the direction of an acoustic reflector. Our proposed method can be beneficial to robotic platform as it can compliment existing laser- and camera-based technologies for generating a spatial map of an indoor environment. Our proposed echolocation method can aid a robotic platform to detect and estimate 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 paper, we aim to implement the proposed method on existing robotic platform, e.g., Softbank's NAO robot and also improve the algorithm and combining it with echolabeling techniques so that multiple acoustic echoes are estimated and categorized. This will help in constructing accurate map of an indoor environment using echolocation alone.

REFERENCES

[1] J. Steckel and H. Peremans, "BatSLAM: Simultaneous localization and mapping using biomimetic sonar," *PLOS ONE*, vol. 8, no. 1, pp. 1–11, 01 2013.

[2] R. Kuc, "Echolocation with bat buzz emissions: Model and biomimetic sonar for elevation estimation," *J. Acoust. Soc. Am.*, vol. 131, no. 1, pp. 561–568, 2012.

[3] M. Kreković, I. Dokmanić, and M. Vetterli, "EchoSLAM: Simultaneous localization and mapping with acoustic echoes," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2016, pp. 11–15.

[4] U. Saqib and J. R. Jensen, "Sound-based distance estimation for indoor navigation in the presence of ego noise," in *Proc. European Signal Processing Conf.*, 2019.

[5] J. R. Jensen, U. Saqib, and S. Gannot, "An EM method for multichannel TOA and DOA estimation of acoustic echoes," in *Proc. IEEE Workshop Appl. of Signal Process. to Aud. and Acoust.*, New Paltz, NY, US, Oct. 2019, pp. 120–124.

[6] U. Saqib, S. Gannot, and J. Jensen, "Estimation of acoustic echoes using expectation-maximization methods," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2020, no. 1, pp. 1–15, 2020.

[7] U. Saqib and J. Jensen, "A model-based approach to acoustic reflector localization using robotic platform," in *Proc. IEEE Int. Conf. Intell., Robot, Automation.* IEEE, 2018, pp. 1–8.

[8] S. Tervo, J. Pätynen, and T. Lokki, "Acoustic reflection localization from room impulse responses," *ACTA Acustica united with Acustica*, vol. 98, no. 3, pp. 418–440, 2012.

[9] G. Defrance, L. Daudet, and J.-D. Polack, "Detecting arrivals within room impulse responses using matching pursuit," in *Proc. of the 11th Int. Conference on Digital Audio Effects (DAFx-08), Espoo, Finland*, vol. 10. Citeseer, 2008, pp. 307–316.

[10] ——, "Using matching pursuit for estimating mixing time within room impulse responses," *Acta Acustica united with Acustica*, vol. 95, no. 6, pp. 1071–1081, 2009.

[11] G. Moschioni, "A new method for measurement of early sound reflections in theaters and halls," in *IMTC/2002. Proceedings of the 19th IEEE Instrumentation and Measurement Technology Conference (IEEE Cat. No.00CH37276)*, vol. 1, 2002, pp. 425–430 vol.1.

[12] Y. Geng and J. Jung, "Sound-source localization system for robotics and industrial automatic control systems based on neural network," in *2008 International Conference on Smart Manufacturing Application*, 2008, pp. 311–315.

[13] S. Dey, S. Boppu, and M. S. Manikandan, "Design of a real-time automatic source monitoring framework based on sound source localization," in *2019 Seventh International Conference on Digital Information Processing and Communications (ICDIPC)*, 2019, pp. 35–40.

[14] H. Zhu and H. Wan, "Single sound source localization using convolutional neural networks trained with spiral source," in *2020 5th International Conference on Automation, Control and Robotics Engineering (CACRE)*, 2020, pp. 720–724.

[15] d. w. gunness, "loudspeaker transfer function averaging and interpolation," *journal of the audio engineering society*, november 2001.

[16] P. Ahgren and P. Stoica, "A simple method for estimating the impulse responses of loudspeakers," *IEEE Transactions on Consumer Electronics*, vol. 49, no. 4, pp. 889–893, 2003.

[17] Z. Sü and M. Çalışkan, "Acoustical design and noise control in metro stations: Case studies of the ankara metro system," *Building Acoustics*, vol. 14, no. 3, pp. 203–221, 2007.

[18] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the em algorithm," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 36, no. 4, pp. 477–489, 1988.

[19] J. Yang, H. Lee, and K. Moessner, "Multilateration localization based on singular value decomposition for 3d indoor positioning," in *Int. Conf. Indoor Positioning and Indoor Navigation*, 2016, pp. 1–8.

[20] J. Wan, N. Yu, R. Feng, Y. Wu, and C. Su, "Localization refinement for wireless sensor networks," *Computer Communications*, vol. 32, no. 13, pp. 1515–1524, 2009.

[21] Y. Zhou, Jun Li, and L. Lamont, "Multilateration localization in the presence of anchor location uncertainties," in *IEEE Global Communications Conference (GLOBECOM)*, 2012, pp. 309–314.

[22] A. Yazici, U. Yayan, and H. Yücel, "An ultrasonic based indoor positioning system," in *Int. Symposium on Innovations in Intell. Sys. and Applications*, 2011, pp. 585–589.

[23] C. Chen and K. Yao, "Source and node localization in sensor networks," in *Classical and Modern Direction-of-Arrival Estimation*, T. E. Tuncer and B. Friedlander, Eds. Boston: Academic Press, 2009, pp. 343–383.

[24] E. A. P. Habets, I. Cohen, and S. Gannot, "Generating nonstationary multisensor signals under a spatial coherence constraint," *J. Acoust. Soc. Am.*, vol. 124, no. 5, pp. 2911–2917, 2008.

[25] R. Humphrey, "Playrec: Multi-channel matlab audio," *URL http://www.playrec.co.uk*, 2007.

[26] H. Herlufsen, "Dual channel FFT analysis (part I)," in *Brüel & Kjær Technical Review*, no. 1984-1, 1984.