A Model-based Approach to Acoustic Reflector Localization with a Robotic Platform

Usama Saqib¹ and Jesper Rindom Jensen¹

Abstract—Constructing a spatial map of an indoor environment, e.g., a typical office environment with glass surfaces, is a difficult and challenging task. Current state-of-the-art, e.g., camera- and laser-based approaches are unsuitable for detecting transparent surfaces. Hence, the spatial map generated with these approaches are often inaccurate. In this paper, a method that utilizes echolocation with sound in the audible frequency range is proposed to robustly localize the position of an acoustic reflector, e.g., walls, glass surfaces etc., which could be used to construct a spatial map of an indoor environment as the robot moves. The proposed method estimate the acoustic reflector’s position, using only a single microphone and a loudspeaker that are present on many socially assistive robot platforms such as the NAO robot. The experimental results show that the proposed method could robustly detect an acoustic reflector up to a distance of 1.5 m in more than 60% of the trials and works efficiently even under low SNRs. To test the proposed method, a proof-of-concept robotic platform was build to construct a spatial map of an indoor environment.

I. INTRODUCTION

Constructing a spatial map of a dynamic environment using a robotic platform is useful, for example, in navigation, analysis and monitoring environments such as buildings, tunnels, etc., for maintenance purposes [1]. Simultaneous Localization And Mapping (SLAM) is a popular framework among the robotic community to generate a spatial map of an indoor environment as well as to localize and orient the pose of the robot [2]–[4]. SLAM is often used in conjunction with external sensors such as cameras, lasers, etc., to receive environmental data. However, camera-based systems are susceptible to changing light conditions which affects the accuracy of the system. Moreover, camera has limited field of view (FOV) which makes it unsuitable to detect objects around the corner [5]. Similarly, Light Detection and Ranging (LiDAR) is a laser-based range sensing technology that is often used with SLAM to accurately generate a spatial map of an environment [6]. However, both LiDAR and camera systems are unsuitable for detecting transparent surfaces that are typically found in an office environment [7].

These issues could be resolved by employing sound. Sound has been widely used in robotics to detect acoustic sources [8] but in this paper we consider echolocation [9], [10]. The advantage of incorporating echolocation on a robotic platform is that it enable robots to navigate under low light conditions or even under complete darkness [11], [12]. Moreover, microphones can be omnidirectional as opposed to common cameras which make their FOV larger. Additionally, microphones can be used to detect audible sources not within direct line of sight. Therefore, constructing a spatial map of an environment using echolocation can be desirable in such difficult scenarios. In recent years, some works on combining echolocation with camera-based systems have been conducted to generate a spatial map of a room. For instance, in [7], the authors proposed using laser and ultrasonic sensor to detect glass surfaces that aids a robot in navigating a room. However, most robotic platforms especially those used for Human-Robot Interaction (HRI), only includes loudspeakers and microphones in the audible frequency range such as Softbank’s NAO robots. In this paper, we consider echolocation with audible sound for mapping an environment with platforms like these, which would not require additional sensors (e.g., ultrasonic, LiDAR).

Localization of acoustic reflectors is a known problem in acoustic signal processing, which can be achieved by estimating the Time-of-Arrivals (TOAs) of reflected sounds. The echoes recorded by a microphone has a certain structure and is distinctly described in two parts: the direct-path plus early reflections and late reflections which are often described as a stochastic and dense tail. This is described by the room impulse response (RIR), which contains important information about the TOAs of the echoes, eventually enabling estimation of the acoustic reflectors’ position. Based on this knowledge, several methods have been proposed to infer the shape of a room from the RIR. For instance, in [13], [14], a collocated loudspeaker and microphone arrangement was used to detect first-order echoes from RIRs, which are then utilized to construct a spatial map of a room. Another attempt to construct a spatial map of a room using a mobile robot is proposed in [15], where the environment is probed to estimate RIRs as the robot moves in a predefined path. Based on this setup, the authors then proposed two room geometry estimation methods, one using simple trigonometry, while the other uses Bayesian filtering. Moreover, echolocation was proposed in [16] for robotic platforms. The authors proposed a multichannel approach to room geometry and robot position estimation, but their approach is also based on a priori knowledge of RIRs. Relying on RIRs is problematic in at least two ways. Firstly, it is a difficult estimation problem in itself to obtain the RIRs, and, secondly, it is non-trivial to extract TOAs from the RIR estimates. Commonly, TOAs are extracted by employing peak picking from estimated RIRs [17]. This is also seen in [11], which is an attempt to generate a spatial map of an outdoor environment using echolocation. The authors in [11] probes the environment and extract TOAs from the echoes as the platform moves to...
where $g_{q,k}$ is the attenuation of the $q^{th}$ reflections from the loudspeaker at the $k^{th}$ robot position to the microphone, and $\tau_{q,k}$ is the TOA of the reflected sound$^1$.

The acoustic impulse response has a certain structure and is distinctively described in two parts: the direct-path plus early reflections and late reflections often described as a stochastic and dense tail. This means that we could rewrite the equation in (2) as:

$$w_k(n) = \sum_{q=1}^{R} g_{q,k}s(n - \tau_{q,k}) + v'(n), \quad (3)$$

where $R$ is the number of early reflections including the direct-path component and $v'(n)$ is the collective noise term comprised of all the late reverberation (i.e., the $q > R$ components) and the additive background noise. The problem at hand is then to estimate the position of the acoustic reflectors from estimates of the TOAs, $\tau_{q,k}$, at the different robot positions, $r_k$.

### III. TOA Estimation and Mapping

We start this section by showing how the TOAs mentioned in (3) can be estimated based on the method in [19]. Then, based on these estimates, we propose the echolocaton approach to mapping of the acoustic reflectors.

#### A. Nonlinear least squares TOA estimation

If $N$ samples of the reflected signals $w_k(n) = [w_k(n) \ w_k(n + 1) \ \cdots \ w_k(n + N - 1)]^T$ is taken while assuming that $s(n)$ is known and the robot position is assumed fixed within these $N$ samples, then a nonlinear least squares (NLS) estimator can be formulated which is statistically optimal under white Gaussian noise conditions. This is expressed as follows:

$$\{\hat{g}_k, \hat{\tau}_k\} = \arg \min_{g, \tau} \left\| w_k(n) - \sum_{q=1}^{R} g_q s(n - \tau_q) \right\|^2, \quad (4)$$

$^1$In our definition in (2), the direct-path component corresponds to $q = 1$
Fig. 2: A proof-of-concept robotic platform
where
\[
\tau = \begin{bmatrix} \tau_1 & \tau_2 & \cdots & \tau_q \end{bmatrix}^T, \tag{5}
\]
\[
g = \begin{bmatrix} g_1 & g_2 & \cdots & g_q \end{bmatrix}^T, \tag{6}
\]
and \(\hat{\tau}_k\) and \(\hat{g}_k\) are defined similarly.

Using Parseval’s theorem, (4) can be transformed into the frequency domain. This helps in reducing the computational cost of the estimator and could facilitate using only selected frequencies for the estimation. This is expressed below:

\[
\{\hat{g}_k, \hat{\tau}_k\} = \arg\min_{\hat{g},\tau} \left\| W_k - \sum_{q=1}^{R} g_q Z(\tau_q) \odot S \right\|^2, \tag{7}
\]

\[
Z(\tau) = \begin{bmatrix} 1 & e^{-j2\pi \frac{\tau}{\frac{1}{K}}} & \cdots & e^{-j2\pi \frac{\tau}{\frac{K-1}{K}}} \end{bmatrix}^T, \tag{8}
\]

where the upper case vectors, e.g., \(W_k\) denotes the frequency domain version of their time-domain counterparts, e.g., \(w_k(n)\). The matrix, \(Z(\tau)\), delays the source signal \(S\) by \(\tau\) samples while \(\odot\) is the element-wise product operator. In order to estimate \(\hat{g}_k\) and \(\hat{\tau}_k\) parameters of the multiple reflections, \(R\), various cyclic methods could be used like the RELAX method proposed in [21] that iteratively estimates the values of \(\tau_{q,k}\) and \(g_{q,k}\). Solving (7) for \(g_q\) by taking the derivative of the cost function yields:

\[
\hat{g}_q = \frac{W_k^H Z(\tau_q) + Z^H(\tau_q) W_k}{2Z^H(\tau_q) Z(\tau_q)} \tag{9}
\]

where \(Z(\tau_q) = Z(\tau_q) \odot S\) is the frequency domain probe signal delayed by \(\tau_q\) samples. By inserting this back into (7), we get

\[
\hat{\tau}_k = \arg\min_{\tau} \left\| W_k - \sum_{q=1}^{R} \frac{W_k^H Z(\tau_q) + Z^H(\tau_q) W_k}{2Z^H(\tau_q) Z(\tau_q)} \right\|^2 \tag{10}
\]

In the special case, where we assume \(R = 1\), e.g., if we are interested in estimating only the nearest acoustic reflector position, we get that

\[
\hat{\tau}_k = \arg\max_{\tau} \Re\{W_k^H Z(\tau)\} \tag{11}
\]

where the operator \(\Re\) represents taking the real part of the signal. As seen, our derivation leads to cross-correlations.

**B. TOA-based acoustic reflector mapping**

The NLS estimator described earlier estimates \(\tau_k\) for every robot position, \(r_k\). By taking multiple observation at different time instance and position, i.e. taking the robot’s movement into account, the NLS estimator can be used to generate a spatial map of an environment. Consider the platform moving in a predefined trajectory \(R = \{r_1, \ldots, r_K\}\) with \(r_k = (r_{x_k}, r_{y_k})\), such that the platform moves from \(r_k\) to \(r_{k+1}\) etc. Here, we implicitly considered mapping in two dimensions, but if additional microphones or loudspeakers are included the principle could be extended to three dimensions. For every position, \(r_k\), the platform will probe the environment with \(s(n)\) and record the observed signal \(w_k(n)\). The probed signal and the observed signal are then converted into the frequency domain before passing it to the NLS estimator. In practice, the analysis window for the TOA could be restricted to a search interval from \(\tau_{\text{min}}\) up to \(\tau_{\text{max}}\) samples. This leads to

\[
\hat{\tau}_k = \arg\max_{\tau \in [\tau_{\text{min}}, \tau_{\text{max}}]} \Re\{W_k^H Z(\tau)\} \tag{12}
\]

To estimate the position of the acoustic reflector from the estimated TOA, \(\hat{\tau}_k\), we assume the loudspeaker to be directional and place the reflector position at a distance corresponding to the estimated TOA in the direction of the loudspeaker. The direction in which the robot platform is moving, \(\theta_{r,k}\), at position \(r_k\), is related to the direction that the loudspeaker is facing, \(\theta_{l,k}\), by a fixed offset angle, \(\Delta\theta\), i.e.,

\[
\theta_{l,k} = \theta_{r,k} + \Delta\theta. \tag{13}
\]
TABLE I: Evaluation of the proposed method against ground truth and SNRs

<table>
<thead>
<tr>
<th>SNR = 0 dB</th>
<th>SNR = 10 dB</th>
<th>SNR = 20 dB</th>
<th>SNR = 30 dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>LiDAR [m]</td>
<td>μ [m]</td>
<td>σ [m]</td>
<td>RMS error [m]</td>
</tr>
<tr>
<td>0.83</td>
<td>0.8796</td>
<td>0.0214</td>
<td>0.0540</td>
</tr>
<tr>
<td>1.15</td>
<td>1.0874</td>
<td>0.0408</td>
<td>0.0832</td>
</tr>
<tr>
<td>1.51</td>
<td>1.4899</td>
<td>0.2837</td>
<td>0.2833</td>
</tr>
<tr>
<td>2.01</td>
<td>1.1584</td>
<td>0.3521</td>
<td>0.9208</td>
</tr>
<tr>
<td>2.50</td>
<td>1.2208</td>
<td>0.3806</td>
<td>1.3456</td>
</tr>
</tbody>
</table>

Based on the above information, the coordinates of the position of the acoustic reflector is then estimated as follows:

\[ p_{x_k} = r_{x_k} + \frac{c}{2} \tau_{k} \cos \theta_{l,k} \]

\[ p_{y_k} = r_{y_k} + \frac{c}{2} \tau_{k} \sin \theta_{l,k} \]

where \( c \) is the speed of the sound. The procedure is then to estimate the acoustic reflector positions for each of the known robot positions, \( r_{x_k}, \) along its trajectory. The estimated acoustic reflector positions are then concatenated in the set \( P = \{p_1, \ldots, p_K\} \) with \( p_k = (p_{x_k}, p_{y_k}) \) for \( k = 1, \ldots, K \).

The resulting method for mapping the environment in two dimensions based on TOA estimates are outlined in Algorithm 1. However, the fixed offset \( \Delta \theta \) could be avoided when employing multiple microphones but this is left for future iteration of this project.

IV. ROBOTIC PLATFORM OVERVIEW

In this section, the hardware and software of the robotic platform is discussed. The proposed method discussed in Section III was implemented on embedded platform with a microcomputer running Windows 10. The microcomputer was developed by UDOO. The UDOO x86 is a single board development platform. On the platform, MATLAB was used to implement the proposed method in Algorithm 1. Moreover, for multichannel audio data acquisition, Playrec [22] was used to emit and record the sounds. A Kobuki TMR-K01-W1 platform was used as the base unit of the robot. It is a wheeled platform with on-board odometry sensor that allows for precise control and movement. The Kobuki platform has a built-in microcontroller that was programmed with a predefined trajectory. The loudspeaker and microphone arrangement is placed on top of the platform, which was connected to a Presonus 1818VSL audio interface. The Presonus interface was then subsequently connected to the UDOO x86 microcomputer. The sampling frequency of the audio interface was set to 48,000 Hz. Moreover, a pre-calibrated laser range sensor, TFMini micro Lidar, was also attached to the an Arduino Uno on the UDOO platform, which was used as the ground truth in our experiments. This helps in evaluating the performance of the proposed method at varying distances under different Signal-to-Noise Ratios (SNRs). The recorded data was processed by the UDOO x86 microcomputer in real-time as the robot was moving along its trajectory. The system diagram is shown in Fig. 1 and the final assembly is shown in Fig. 2.

V. EXPERIMENTAL SETUP AND RESULTS

Two experiments were performed to evaluate the performance of the proposed method. Both were tested on the proof-of-concept robotic platform discussed in the previous section. The first experiment evaluate the performance of the used NLS estimator under different SNRs and distances while placing it against one reflector as shown in Fig. 3(a). The second experiment tested the system in a real scenario of generating a spatial map as the robot moves in a predefined trajectory. Furthermore, in the second experiment, two indoor environments were conducted in the Sound Lab and an office area with a glass partition, respectively. Both environments are located in the CREATE building at Aalborg University, Denmark.

For our experiments, we assumed the speed of sound to be 343 m/s and considered an analysis window starting from \( \tau_{\text{min}} \) samples to \( \tau_{\text{max}} \) samples corresponding to distances from 0.8 m to 2 m. The interval was selected such that the first-order reflections between distances of 0.8 m and 2 m from the microphone were captured, without capturing the direct-path component. Moreover, \( R \) was set to 1 so that only one reflection from the first-order early reflection was considered for the estimation. For both experiments, the
source signal, $s(n)$, was selected as a broadband signal of length 1,500 samples drawn from a Gaussian burst with zero padding to form a signal with length of $N = 20,000$ samples. Furthermore, a LiDAR was as placed adjacent to the microphone to measure the distance to the acoustic reflector as shown in Fig. 2. This distance serves as the ground truth for our experiments.

**A. Evaluation for different SNRs and distances**

This experiment was performed inside the Sound Lab at Aalborg University, Denmark. The Sound Lab has dimensions of $6.38 \times 5.4 \times 4.05$ m and has sound absorbing materials embedded into the wall. The test was performed to evaluate the performance of the TOA estimation under different SNRs and distances. To simulate low SNRs, a separate loudspeaker was used as the interfering source playing an audio clip called *Cocktail Party*\(^2\). The loudspeaker was placed at the corner of the lab at a distance of 6.4 m away from the robot as shown in Fig. 3(a). The SNR is defined as the ratio between the variance of the recorded probed signal $x(n)$ against the variance of the background noise $v(n)$, i.e.,

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

where $\sigma_x^2 = E[|x(n)|^2]$ and $\sigma_v^2 = E[|v(n)|^2]$. Both the background noise and probe signal were recorded for 1 second. Then, 4 SNR values $[30, 20, 10, 0]$ dB were selected for this experiment. To test the performance of the proposed method in estimating the distance of the wall, the platform was placed at the following distances from one of the walls $[0.8, 1.0, 1.5, 2.0, 2.5]$ m. The results from this experiment are depicted in Fig. 3 and Table I.

**B. Mapping of an indoor environment**

The second experiment was conducted to evaluate the spatial mapping method in Algorithm 1 on the robotic hardware platform. The objective here is to successfully generate a spatial map of the environment, while detecting transparent surfaces, e.g., glass partitions. Two trajectories were predefined for both environment. To construct the spatial map of the office area with the glass surface in Fig. 4(a), the robot was predefined to move in straight line along the glass surface, while a rectangular trajectory was selected for the experiment in the Sound Lab shown in Fig. 5(a). In both environments, the robot follows the predefined trajectories and stops every 1 m before coming to a momentary stop for 2 seconds. During these 2 seconds, the platform then probes the environment using a known signal, $s(n)$ and uses the proposed method to determine the location of the acoustic reflector before moving to a new location and repeating the echolocation process for each of the positions, $r_k$, for $k = 1, \ldots, K$ as outlined in Algorithm 1. The results from the mapping experiment are shown in Fig. 4 and 5.

**VI. DISCUSSION**

The data collected from the first experiment is summarized in Fig. 3 and Table I. These results show that the proposed method can accurately measure the acoustic
reflector positions even when testing with sound absorbing materials and under highly noisy conditions. In general, the material of the environment is expected to be more reflective then in the considered experimental setting, which should only make the reflector position estimation easier, since the reflections will be stronger. To measure the performance of the TOA estimator, we considered its accuracy defined as the percentage of the TOA estimates that are within ±ε of the ground truth (LiDAR) data, where ε was chosen as 10%. The results in Fig. 3(b) shows that the proposed method has above 60% accuracy up to a distance of 1.5 m even under low SNRs. For each SNR and distance configuration, we conducted 100 experiments. In each of these, the probe signal and the background noise were sampled randomly.

Furthermore, as seen in the Table I, as the distance between the platform and robot increases, the standard deviation, σ, and Root Mean Square Error (RMSE), also seen in Fig. 3(c), increases. Additionally, the mean, μ, is close to the ground truth value for distances up to 1.5 m and for all SNRs. In conducting the second experiment, we only considered a single high SNR level of approximately 30 dB. As seen in both Fig. 4(b) and Fig. 5(b), a spatial map of both environment was obtained based on the sound recordings at the different robot position along the trajectory. Moreover, in Fig. 4(b), the algorithm accurately constructed reflector position estimates even in the presence of transparent glass surfaces as opposed to the LiDAR data. Careful examination of Fig. 4(b) and Fig. 5(b) would reveal that the data points are not aligned to the layout of the wall. This is due to the drift in the robotic platform. One way to overcome this drift is to monitor the motor state and compensate the movement of the platform. However, this aspect is beyond the scope of this paper and will be tackled in the future iteration of this research.

VII. CONCLUSION

The contribution of this paper is to propose a new mapping algorithm that could benefit many robotic applications by localizing acoustic reflectors from recorded microphone data. Our proposed method make use of a single loudspeaker-microphone arrangement which are commonly found in many robotic platforms used for Human Robot Interaction (HRI). Two experiments were conducted: one to test the performance of the proposed method, and another to construct a spatial map of two indoor environments. According to the data, the proposed method can robustly detect acoustic reflector at distances up to 1.5 m with above 60 % accuracy. It is also seen from the results that at higher distances, the standard deviation and the RMSE also increases which reduces the overall performance of the algorithm. Furthermore, in the second experiment, spatial maps of two environments were estimated using the proposed method on a robotic platform following a predefined trajectory. Fig. 4 and 5 shows that the method accurately detects even sound absorbing and transparent surfaces when compared with traditional sensing technologies. In the future iteration of this research, the proposed method will be extended to a multichannel approach, which should increase the accuracy further and enable estimation of multiple reflectors. Moreover, movement will be taken into account, in which case the reflector localizing can be conducted while the robot is moving.

REFERENCES


