Diffraction-Aware Sound Localization for a Non-Line-of-Sight Source

Inkyu An[1], Doheon Lee[2], Jung-woo Choi[3], Dinesh Manocha[4], and Sung-eui Yoon[5]
http://sgvr.kaist.ac.kr/~ikan/papers/DA-SSL

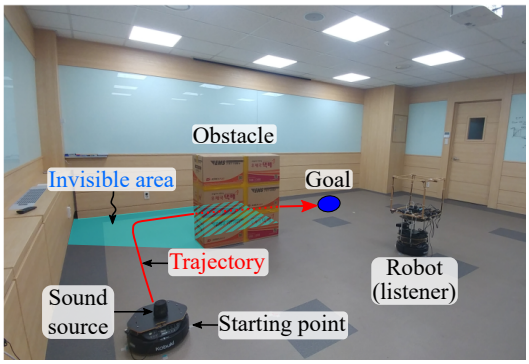
***Abstract*— We present a novel sound localization algorithm for a non-line-of-sight (NLOS) sound source in indoor environments. Our approach exploits the diffraction properties of sound waves as they bend around a barrier or an obstacle in the scene. We combine a ray tracing-based sound propagation algorithm with a Uniform Theory of Diffraction (UTD) model, which simulate bending effects by placing a virtual sound source on a wedge in the environment. We precompute the wedges of a reconstructed mesh of an indoor scene and use them to generate diffraction acoustic rays to localize the 3D position of the source. Our method identifies the convergence region of those generated acoustic rays as the estimated source position based on a particle filter. We have evaluated our algorithm in multiple scenarios consisting of static and dynamic NLOS sound sources. In our tested cases, our approach can localize a source position with an average accuracy error of 0.7m, measured by the L2 distance between estimated and actual source locations in a 7m×7m×3m room. Furthermore, we observe 37% to 130% improvement in accuracy over a state-of-the-art localization method that does not model diffraction effects, especially when a sound source is not visible to the robot.**

I. INTRODUCTION

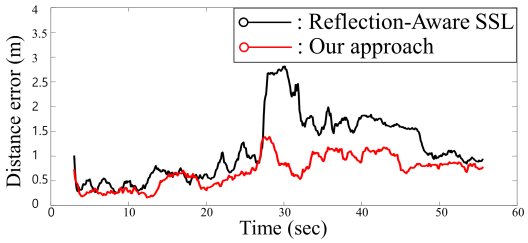
As mobile robots are increasingly used for different applications, there is considerable interest in developing new and improved methods for localization. The main goal is to compute the current location of the robot with respect to its environment. Localization is a fundamental capability required by autonomous robots because the current location is used to guide future movement or actions. We assume that a map of the environment is given and different sensors on the robot are used to estimate its position and orientation in the environment. Some of the commonly used sensors include GPS, CCD or depth cameras, acoustics, etc. In particular, there is considerable work on using acoustic sensors for localization, including sonar signal processing for underwater localization and microphone arrays for indoor and outdoor scenes. In particular, the recent use of smart microphones in commodity or IoT devices (e.g., Amazon Alexa) has triggered interest in better acoustic localization methods [2], [3].

Acoustic sensors use the properties of sound waves to compute the source location. Sound waves are emitted from a source and then transmitted through the media to reach either the listener or microphone locations as direct paths, or after undergoing different wave effects including reflections, interference, diffraction, scattering, etc. Some of the earliest work on sound source localization (SSL) makes use of the time difference of arrival (TDOA) at the receiver [4], [5], [6]. These methods only exploit the direct sound and its direction at the receiver and do not take into account reflections or other wave effects. As a result, these methods do not provide sufficient accuracy for many applications. Other techniques have been proposed to localize the position under different constraints or sensors [1], [7], [8], [9]. This includes modeling higher order specular reflections [1] based on ray tracing and modeling indirect sound effects.

In many scenarios, the sound source is not directly in the line of sight of the listener (i.e. NLOS) and is occluded by obstacles. In such cases, there may not be much contribution in terms of direct sound, and simple methods based on TDOA



(a) A Non-Line-of-Sight (NLOS) moving source scene around an obstacle. Our method can localize its position using acoustic sensors and our diffraction-aware ray tracing.



(b) Accuracy errors, measured as the L2 distance between the estimated and actual 3D locations of a sound source, for the dynamic source. Our method models diffraction effects and improves the localization accuracy as compared to only modeling indirect reflections [1]

Fig. 1. These figures show the testing environment (7m by 7m with 3m height) (a) and the accuracy error of our method with the dynamically moving sound source (b). The source moves along the red trajectory, and the obstacle causes the invisible area for the dynamic source. Invisibility of the source occurs from 27s to 48s, where our method maintains a high accuracy, while the prior method deteriorates due to the diffraction: the average distance errors of our and the prior method are 0.95m and 1.83m.

[1]I. An, [2]D. Lee, and [5]S. Yoon (Corresponding author) are with the School of Computing, KAIST, Daejeon, South Korea; [3]J. Choi is with the School of Electrical Engineering, KAIST; [4]D. Manocha is with the Dept. of CS & ECE, Univ. of Maryland at College Park, USA; {inkyu.an, doheonlee, jwoo}@kaist.ac.kr, dm@cs.umd.edu, sungeui@kaist.edu

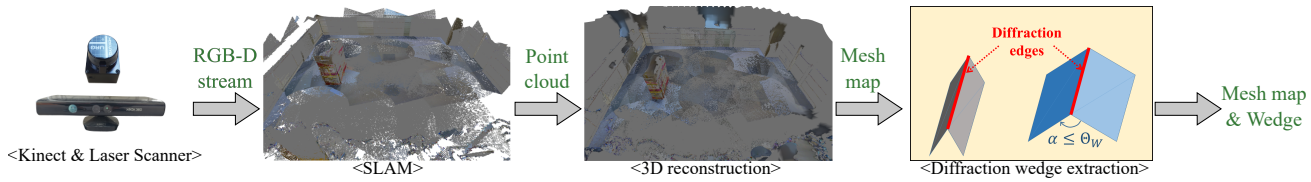


Fig. 2. This figure shows our precomputation phase. We use SLAM to generate a point cloud of an indoor environment from the laser scanner and Kinect. The point cloud is used to construct the mesh map via 3D reconstruction techniques. Wedges whose two neighboring triangles have angles larger than θ_W; their edges are extracted from the mesh map to consider diffraction effects at runtime for sound localization.

may not work well. We need to model indirect sound effects and the most common methods of this type of modeling are based on using ray-based geometric propagation paths. They assume the rectilinear propagation of sound waves and use ray tracing to compute higher order reflections. While these methods work well for high frequency sounds, they do not accurately model many low-frequency phenomena such as diffraction, a type of scattering that occurs from obstacles with sizes of the same order of magnitude as the wavelength. In practice, diffraction is a fundamental mode of sound wave propagation and occurs frequently in building interiors (e.g., the source is behind an obstacle or hidden by walls). These effects are more prominent for low-frequency sources such as vowel sounds in human speech, industrial machinery, ventilation, air-conditioned units.

Main Results. We present a novel sound localization algorithm that takes into account diffraction effects, especially from non-line-of-sight or occluded sources. Our approach is built on a ray tracing framework and models diffraction using the Uniform Theory of Diffraction (UTD) [10] along the wedges. During the precomputation phase, we use SLAM and reconstruct a 3D triangular mesh for an indoor environment. At runtime, we generate direct acoustic rays towards incoming sound directions as computed by TDOA. Once the acoustic ray hits the reconstructed mesh, we generate reflection rays. Furthermore, when acoustic rays pass close enough to the edges of mesh wedges according to our diffraction-criterion, we also generate diffraction acoustic rays to model non-visible paths to include an incident sound direction that can be actually traveled (Sec. III). Finally, we estimate the source position by performing generated acoustic rays using ray convergence.

We have evaluated our method in an indoor environment with three different scenarios including a stationary source and a dynamically moving source along an obstacle that blocks the direct line-of-sight from the listener. In these cases, the diffracted acoustic waves are used to localize the position. We combine our diffraction method with a reflection-aware SSL algorithm [1] and observe improvements from 1.22m to 0.7m, on average, and from 1.45m to 0.79m for the NLOS source. Our algorithm can localize a source generating a clapping sound within 1.38m as the worse error bound in a room of dimensions 7m $\times$ 7m and 3m height.

II. Related Work

In this section, we give a brief overview of prior work on sound source localization and sound propagation.

Sound source localization (SSL). Over the past two decades, many approaches have used time difference of arrival (TDOA) to localize sound sources. Knapp *et al.* presented a good estimation of the time difference using a generalized correlation between a pair of microphone signals [4]. He *et al.* [5] suggested a deep neural network-based source localization algorithm in the azimuth direction for multiple sources. This approach focused on estimating an incoming direction of a sound and did not localize the actual position of the source.

Recently, many techniques have been proposed for estimating the location of a sound source [7], [8], [9]. Sasaki *et al.* [7] and Su *et al.* [8] presented 3D sound source localization algorithms using a disk-shaped sound detector and a linear microphone array such as Kinect and PS3 Eye. Misra *et al.* [9] suggested a robust localization method in noisy environments using a drone. This approach requires the accumulation of steady acoustic signals at different positions, and thus cannot be applied to a transient sound event or to stationary sound detectors.

An *et al.* [1] presented a reflection-aware sound source localization algorithm that used direct and reflected acoustic rays to estimate a 3D source position in indoor environments. Our approach is based on this work and takes into account diffraction effects to considerably improve the accuracy.

Interactive sound propagation. There is considerable work in acoustics and physically-based modeling to develop fast and accurate sound simulators that can generate realistic sounds for computer-aided design and virtual environments. Geometry acoustic (GA) techniques have been widely utilized to simulate sound propagations efficiently using ray tracing techniques. Because ray tracing algorithms are based on the sound propagation model at high frequencies, low-frequency wave effects like diffraction are modeled separately.

In addition, an estimation of the acoustic impulse response between the source and the listener was performed using Monte Carlo path tracing [11], an adaptive frustum tracing [12] or a hybrid combination of geometric and numeric methods techniques [13].

Exact methods to model diffraction are based on solving the acoustic wave equation directly using numeric methods like boundary or finite element methods [14], [15], the wave-geometric approximation method [16], the Kresnel-Kirchoff approximation method [17], or the BTM model [18] and its extension to higher order diffraction models [19]. Commonly used techniques to model diffraction with geometric acoustic methods are based on two models: the Uniform Theory of

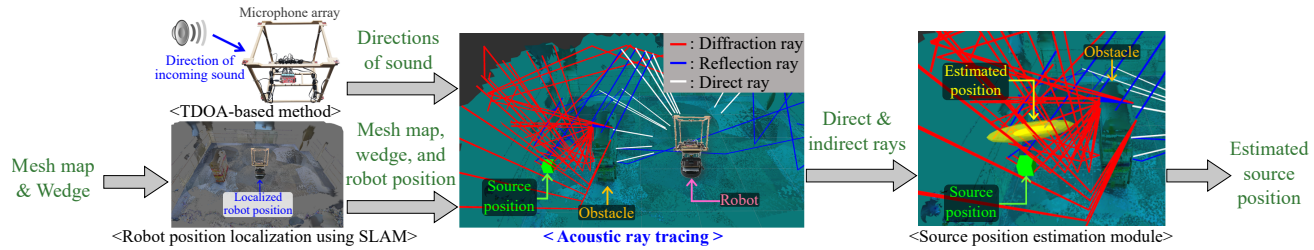


Fig. 3. We show run-time computations using acoustic ray tracing with diffraction rays for sound source localization. The diffraction-aware acoustic ray tracing is highlighted in blue and our main contribution in this paper. The source position estimation is performed by identifying ray convergence.

Diffraction (UTD) [20] and the Biot-Tolstoy-Medwin (BTM) model [18]. The BTM model is an accurate diffraction formulation that computes an integral of the diffracted sound along the finite edges in the time domain [19], [15], [21]. In practice, the BTM model is more accurate, but is limited to non-interactive applications. The UTD model approximates an infinite wedge as a secondary source of diffracted sounds, which can be reflected and diffracted again before reaching the listener. UTD-based approaches have been effective for many real-time sound generation applications, especially in complex environments with occluding objects [11], [22], [23], [24]. Our approach is motivated by these real-time simulations and proposes a real-time source localization algorithm using UTD.

III. Diffraction-Aware SSL

We present our diffraction-aware SSL based on acoustic ray tracing.

A. Overview

Precomputation. Given an indoor scene, we reconstruct a 3D model as part of the precomputation. We use a Kinect and a laser scanner to capture a 3D point cloud representation of the indoor scene. As shown in Fig. 2, the point cloud capturing the indoor geometry information is generated by the SLAM module from raw depth data and an RGB-D stream collected by the laser scanner and Kinect. Next, we reconstruct a 3D mesh map via the generated point cloud. We also extract wedges from the mesh that have an angle between two neighboring triangles smaller than the threshold, Θ_W. The reconstructed 3D mesh map and the wedges on it are used for our diffraction method at runtime.

Runtime Algorithm. We provide an overview of our runtime algorithm as it performs acoustic ray tracing and sound source localization in Fig. 3. Inputs to our runtime algorithm are the audio stream collected by the microphone array, the mesh map reconstructed in the precomputation, and the robot position localized by the SLAM algorithm. Our goal is to find the 3D position of the sound source in the environment. Based on those inputs, we perform acoustic ray tracing supporting direct, reflection, and diffraction effects by generating various acoustic rays (III-B). The source position is computed by estimating the convergence region of the acoustic rays (III-D). Our novel component, acoustic ray tracing with diffraction rays, is highlighted in the blue font in Fig. 3.

B. Acoustic Ray Tracing

In this section, we explain how our acoustic ray tracing technique generates direct, reflection, and diffraction rays.

At runtime, we first collect the directions of the incoming sound signals from the TDOA algorithm [25]. For each incoming direction, we generate a primary acoustic ray in the backward direction; as a result, we perform acoustic ray tracing in a backward manner. At this stage, we cannot determine whether the incoming signal is generated by one of the states: direct propagation, reflection, or diffraction. We can determine the actual states of these primary acoustic rays while performing backward acoustic ray tracing. Nonetheless, we denote this primary ray as the direct acoustic ray since the primary ray is a direct ray from the listener's perspective.

We represent a primary acoustic ray as r_n^0 for the n-th incoming sound direction. Its superscript denotes the order of the acoustic path, where the 0-th order denotes the direct path from the listener. We also generate a (backward) reflection ray once an acoustic ray intersects with the scene information under the assumption that the intersected material mainly consists of specular materials [1]. The main difference from the prior method [1] is that we use a mesh-based representation, while the prior method used a voxel-based octree representation for intersection tests. This mesh is computed during precomputation and we use the triangle normals to perform the reflections. As a result, for the n-th incoming sound direction, we recursively generate reflection rays with increasing orders, encoded by a ray path that is defined by $R_n = [r_n^0, r_n^1, ...]$. The order of rays increases as we perform more reflection and diffraction.

C. Handling Diffraction with Ray Tracing

We now explain our algorithm for efficiently modeling the diffraction effects within acoustic ray tracing to localize the sound source. Since our goal is to achieve fast performance in localizing the sound source, we use the formulation based on the Uniform Theory of Diffraction (UTD) [20]. The incoming sounds collected by the microphone array consist of contributions from different effects in the environment, including reflections and diffractions.

Edge diffraction occurs when an acoustic wave hits the edge of a wedge. In the context of acoustic ray tracing, when an acoustic ray hits an edge of a wedge between two neighboring triangles, the diffracted signal propagates into all possible directions from that edge. The UTD model assumes

that the point on the edge causing the diffraction effect is an imaginary source generating the spherical wave [20].

To solve the problem of localizing the sound source, we simulate the process of backward ray tracing. Suppose that an n-th incoming sound direction denoted by the ray r_n^{j-1} is generated by the diffraction effect at an edge. In an ideal case, the incoming ray will hit the edge of a wedge and generate the diffraction acoustic ray r_n^j, as shown in Fig. 4; in (a), $r_n^{(j,\cdot)}$ is shown. It is important to note that there can be an infinite number of incident rays generating diffractions at the edge. Unfortunately, it is not easy to link the incident direction to the edge generating the diffraction exactly. Therefore, we generate a set of N_d different diffraction rays in a backward manner that covers the possible incident directions to the edge based on the UTD model. This set is generated based on an assumption that one of those generated rays might have the actual incident direction causing the diffraction. When there are sufficient acoustic rays, including the primary, reflection, and diffraction rays, it is highly likely that those rays will pass through or near the sound source location; we choose a proper value of N_d by analyzing diffraction rays (Sec. IV).

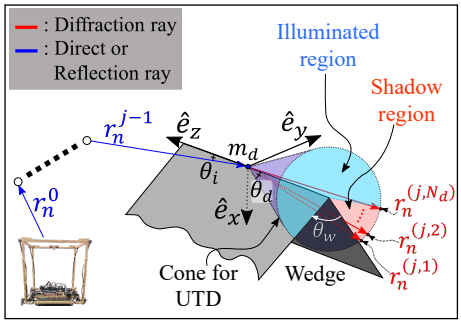
This explanation begins with the ideal case, where the acoustic ray r_n^{j-1} hits the edge of a wedge. Because our algorithm works on a real environment that contains various types of errors from sensor noises and resolution errors from the TDOA method, it is rare that an acoustic ray intersects an edge exactly.

To support various cases that arise in real environments, we propose using the notion of *diffraction-condition* between a ray and a wedge. The diffraction-condition simply measures how close the ray r_n^{j-1} passes to an edge of the wedge. Specifically, we define the *diffractability* v_d according to the angle θ_D between the acoustic ray and its ideally generated ray for the diffraction with the wedge: i.e. $v_d = \cos(\theta_D)$, where the cos function is used to normalize the angle θ_D (Fig. 5).

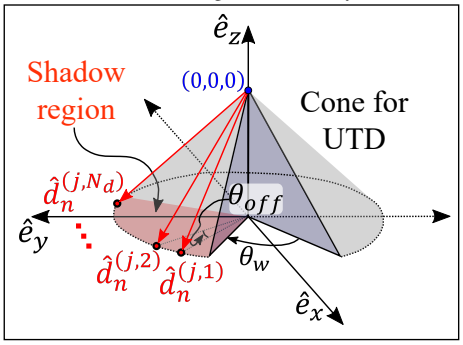
Given an acoustic ray r_n^{j-1}, we define its ideally generated ray r'^{j-1}_n as the projected ray of r_n^{j-1} on the edge of the wedge where the end point m_d of r'^{j-1}_n is on that edge (refer to the geometric illustration on Fig. 5). The point m_d is located at the position closest to the point m_n^{j-1} of the input ray r_n^{j-1}; due to the page limit, we do not show its detailed derivation, but it can be defined based on our high-level description.

If the diffractabilty v_d is larger than a threshold value, e.g., 0.95 in our tests, our algorithm determines that the acoustic ray is generated from the diffraction at the wedge, and we thus generate the secondary, diffraction ray at the wedge in a backward manner.

We now present how we generate the diffraction rays when the acoustic ray satisfies the diffraction-condition. The diffraction rays are generated along the surface of the cone (Fig. 4a) because the UTD model is based on the principle of Fermat [10]: the ray follows the shortest path from the source to the listener. The surface of the cone for the UTD model contains every set of the shortest paths. When



(a) Generating diffraction rays



(b) Computing outgoing directions of diffraction rays.

Fig. 4. This figure illustrates our acoustic ray tracing method for handling the diffraction effect. (a) Suppose that we have an acoustic ray r_n^{j-1} that satisfies the diffraction condition, hitting or passing near the edge of a wedge. We then generate N_d diffraction rays covering the possible incoming directions (especially in the shadow region) of rays that cause diffraction. (b) An outgoing unit vector, $\hat{d}_n^{(j,p)}$, of a p-th diffraction ray is computed on local coordinates $(\hat{e}_x, \hat{e}_y, \hat{e}_z)$ and used after the transformation to the environment in runtime, where $\hat{e}_z$ fits on the edge of the wedge and $\hat{e}_x$ is set half-way between two triangles of the wedge.

the acoustic ray r_n^{j-1} satisfies the diffraction-condition, we compute outgoing directions for those diffraction rays. Those directions are the unit vectors generated on that cone and can be computed on a local domain as shown in Fig. 4b:

$$\hat{d}_n^{(j,p)} = \begin{bmatrix} \cos(\theta_w/2 + p \cdot \theta_{off}) \sin\theta_d \\ \sin(\theta_w/2 + p \cdot \theta_{off}) \sin\theta_d \\ -\cos\theta_d \end{bmatrix}, \quad (1)$$

where $\hat{d}_n^{(j,p)}$ denotes the outgoing unit vector of a p-th diffraction ray among N_d different diffraction rays, θ_w is the angle between two triangles of the wedge, θ_d is the angle of the cone that is the same as the angle between the outgoing diffraction rays and the edge on the wedge, and θ_{off} is the offset angle between two sequential diffraction rays, i.e. $\hat{d}_n^{(j,p)}$and $\hat{d}_n^{(j,p+1)}$, on the bottom circle of the cone.

Given a hit point m_d by an acoustic ray r_n^{j-1} on the wedge, we transform the outgoing directions in the local space to the world space by aligning their coordinates $(\hat{e}_x, \hat{e}_y, \hat{e}_z)$. Based on those transformed outgoing directions, we then compute the outgoing diffraction rays, $\bar{r}_n^{(j)} = \{r_n^{(j,1)}, ..., r_n^{(j,N_d)}\}$, starting from the hit point m_d.

To accelerate the process, we only generate the diffraction rays in the shadow region, which is defined by the wedge; the rest of the shadow region is called the illuminated region.

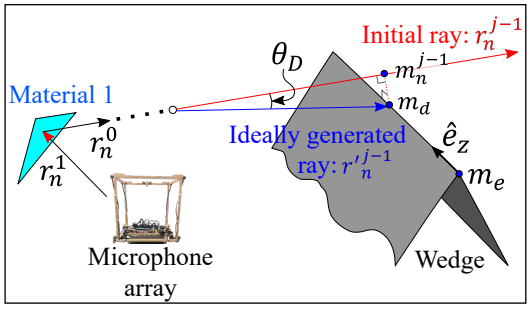


Fig. 5. This figure shows the diffraction condition. When a ray r_n^{j-1} passes close to an edge of a wedge, we consider the ray to be generated by the edge diffraction. We measure the angle θ_D between the ray and its ideal generated ray, which hits the edge exactly, to check our diffraction condition.

We use this process because covering only the shadow region but not the illuminated region generates minor errors in the simulation of sound propagation [22].

Given the new diffraction rays, we apply our algorithm recursively and generate another order of reflection and diffraction rays. Given the n-th incoming direction signal, we generate acoustic rays, including direct, reflection, and diffraction rays and maintain the ray paths R_n in a tree data structure. The root of this tree represents the direct acoustic ray, starting from the microphones. The depth of the tree denotes the order of its associated ray. Note that we generate one child and N_d children for handling reflection and diffraction effects, respectively.

D. Estimating the Source Position

We explain our method used for localizing the sound source position using acoustic rays. Our estimation is based on Monte-Carlo localization (MCL), also known as the particle filter [1]. Our estimation process assumes that there is a single sound source in the environment that causes a high probability that all those acoustic ray paths pass near that source; handling multiple targets using a particle filter has been also studied [26]. In other words, the acoustic rays converge in a region located close to the source, and our estimation aims to identify such a convergence region out of all the generated rays.

The MCL approach generates initial particles in the space as an approximation to the source locations. It allocates higher weights to particles that are closer to acoustic rays and re-samples the particles to get more particles in regions with higher weights [1]. Specifically, we adopt the generalized variance, which is a one-dimensional measure for multi-dimensional scatter data, to see whether particles have converged. When the generalized variance is less than a threshold (e.g., $\sigma_c = 0.5$), we treat the sound that occurs and the mean position of those particles as the estimated sound source position.

IV. Results and Discussion

In this section, we describe our setup, which consists of a robot with microphones and testing environments, and highlight the performance of our approach. The hardware platform is based on Turtlebot2 with a 2D laser scanner, Kinect, a computer with an Intel i7 process, and a microphone array, which is an embedded system for streaming multi-channel audio [27], consisting of eight microphones. For all the computations, we use a single core, and perform our estimation every 200ms, supporting five different estimations in one second.

Benchmarks. We have evaluated our method in indoor environments containing a box-shaped object that blocks direct paths from the sound to the listener. We use two scenarios: a stationary sound source and a moving source. As shown in Fig. 6, we place an obstacle between the robot and the stationary sound source, such that the source is not in the direct line-of-sight of the robot (i.e. NLOS source). We use another testing environment with a source moving along the red trajectory, as shown in Fig. 1a. These two scenarios are tested on the same room size: 7m $\times$ 7m and 3m height.

During the precomputation phase, we perform SLAM and reconstruct a mesh of the testing environment. We ensure that the resulting mesh has no holes using the MeshLab package.

Stationary sound source with an obstacle. We evaluate the accuracy by computing the L2 distance errors between the positions estimated by our method and the ground-truth positions. We use two types of sound signals: clapping and male speech, where male speech has more low-frequency components than the clapping sound (dominant frequency range of the clapping sound is 2k$\sim$2.5kHz and the range is 0.1k$\sim$0.7kHz for male speech).

We compare the accuracy of our approach with that of Reflection-Aware SSL (RA-SSL) [1], which models direct sound and indirect reflections, but does not handle diffraction. For the stationary source producing the clapping sound (Fig. 7a), the average distance errors of the RA-SSL and our method are 1.4m and 0.6m, respectively. There are configurations of the sound source that are not visible to the microphone (NLOS). In this case, we observe 130% better accuracy by modeling these diffraction rays.

Fig. 7b shows the localization accuracy for the male speech signal, which has more low-frequency components. The measured distance errors are, on average, 1.12m for RA-SSL and 0.82m for our approach. While we also observe meaningful improvement, it is less than we see with the

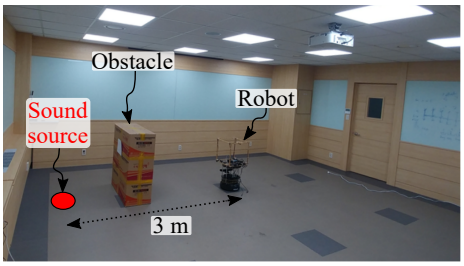
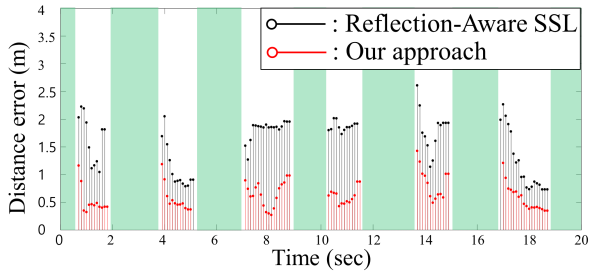
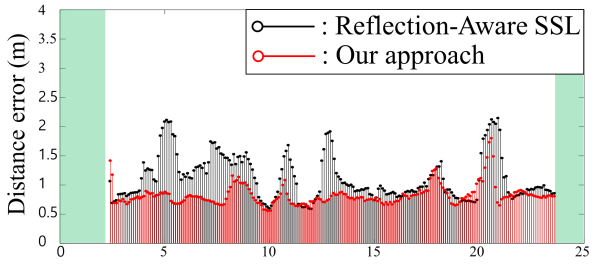


Fig. 6. The evaluation environment for the static sound source. Direct paths from the sound source to the listener are blocked by the obstacle. We use our diffraction-based algorithm for localization.



(a) Stationary source (clapping).



(b) Stationary source (male speech).

Fig. 7. These graphs compare the localization distance errors of our method with the prior, reflection-aware SSL method [1] using the clapping sound source (a) and male speech signal source (b); green regions indicate no sound in that period. The average distance errors of RA-SSL and our method are 1.4m and 0.6m in (a), and 1.12m and 0.82m in (b), respectively. The use of diffraction considerably reduces the localization errors.

clapping sound. Our method supports diffraction, but diffuse reflection is not yet supported. Given the many low-frequency components of male speech, we observe that it is important to support diffuse reflection in addition to diffraction. Nonetheless, by modeling diffraction for male speech, we observe meaningful improvement (37% on average) in localization accuracy.

Moving sound source around an obstacle. We also evaluate our algorithm on a more challenging environment that contains a sound source (clapping sound) moving along the red trajectory shown in Fig 1a. Its accuracy graphs are presented in Fig. 1b; the average distance errors of the RA-SSL and our method are 1.15m and 0.7m, respectively, indicating a 64% improvement in accuracy using our localization algorithm. It is interesting that, when the dynamic source is in the area corresponding to these time values (27s $\sim$ 48s), which are NLOS with respect to the robot, the average distance errors of the RA-SSL and our method are 1.83m and 0.95m, respectively, indicating a 92% improvement. This clearly demonstrates the benefits of our method in terms of localization.

Overall, we achieved 130%, 37%, and 64% improvement, resulting in 77% average improvement, on the stationary source with a clapping sound, the stationary source with male speech, and the dynamic source, respectively, compared with the prior method RA-SSL [1]. The summary of the accuracy of our method compared with RA-SSL is in Table I.

Analysis of diffraction rays. By modeling the diffraction effects, we increase the number of generated rays, resulting in a computational overhead. As a result, we measure the average accuracy error and computation time as a function of N_d, the number of diffraction rays for simulating each

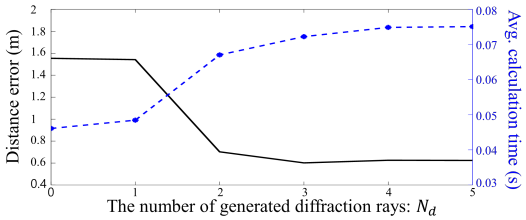


Fig. 8. This figure shows the average accuracy error and computation time for our method on an Intel i7 processor 6700 as a function of N_d that is the number of diffraction rays generated for simulating the edge diffraction.

TABLE I

SUMMARY OF THE ACCURACIES OF THE DIFFERENT METHODS (∗: ONLY NLOS SOURCE)

Scenario Sound	**Stationary**∗ Clapping	**Stationary**∗ Male voice	**Dynamic** Clapping	**Dynamic**∗ Clapping
RA-SSL	1.4m	1.12m	1.15m	1.83m
Ours	0.6m (130%)	0.82m(37%)	0.7m(64%)	0.95m(92%)

edge diffraction. As shown in Fig. 8, the average accuracy error gradually decreases, but we found that when N_d is in a range of 2 to 5, the accuracy is rather saturated. Since we can accommodate up to $N_d = 5$ given our runtime computation budget (0.2 s), we use $N_d = 5$ across all the experiments. In this case, the average numbers of direct, reflection, and diffraction rays are 18, 26, and 184, respectively, in the case of the static source with the clapping sound. In addition, the average running times for acoustic ray tracing and particle filters are 0.09*ms* and 72*ms*; our un-optimized particle filter uses 100 particles and computes weights of them against all the other rays. When we are done estimating the location within the time budget, we allow our process to rest in an idle state.

V. CONCLUSIONS & FUTURE WORK

We have presented a novel, diffraction-aware source localization algorithm. Our approach can be used for localizing an NLOS source and it models the diffraction effects using the uniform theory of diffraction. We have combined our method with indirect reflections and have tested our method in various scenarios with static and moving sound sources with different sound signals.

While we have demonstrated various benefits of our approach, it has some limitations. The UTD model is an approximate model and is mainly designed for infinite wedges. As a result, its accuracy may vary in different environments. We observed lower accuracy for low-frequency sounds (male voices), mainly due to the diffusion effect. Our implemented approach is limited to a single sound source in the environment and does not model all the scattering effects. As part of future work, we would like to address these problems.

ACKNOWLEDGMENT

This research was supported by the SW StartLab program (IITP-2015-0-00199), NRF (NRF-2017M3C4A7066317), ARO grant W911NF-18-1-0313, and Intel.

REFERENCES

[1] Inkyu An, Myungbae Son, Dinesh Manocha, and Sung-eui Yoon, "Reflection-aware sound source localization", in *ICRA*, 2018.

[2] Craig C Douglas and Robert A Lodder, "Human identification and localization by robots in collaborative environments", *Procedia Computer Science*, vol. 108, pp. 1602–1611, 2017.

[3] Muhammad Imran, Akhtar Hussain, Nasir M Qazi, and Muhammad Sadiq, "A methodology for sound source localization and tracking: Development of 3d microphone array for near-field and far-field applications", in *Applied Sciences and Technology (IBCAST), 2016 13th International Bhurban Conference on*. IEEE, 2016, pp. 586–591.

[4] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay", *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327.

[5] Petr Motlicek Weipeng He and Jean-Marc Odobez, "Deep neural networks for multiple speaker detection and localization", in *ICRA*, 2018.

[6] João Filipe Ferreira, Cátia Pinho, and Jorge Dias, "Implementation and calibration of a bayesian binaural system for 3d localisation", in *Robotics and Biomimetics, 2008. ROBIO 2008. IEEE International Conference on*. IEEE, 2009, pp. 1722–1727.

[7] Y. Sasaki, R. Tanabe, and H. Takemura, "Probabilistic 3d sound source mapping using moving microphone array", in *IROS*, 2016.

[8] D. Su, T. Vidal-Calleja, and J. V. Miro, "Towards real-time 3d sound sources mapping with linear microphone arrays", in *ICRA*, 2017.

[9] Pragyan Mohapatra Prasant Misra, A. Anil Kumar and Balamuralidhar P., "Droneears: Robust acoustic sound localization with aerial drones", in *ICRA*, 2018.

[10] Joseph B Keller, "Geometrical theory of diffraction", *JOSA*, vol. 52, no. 2, pp. 116–130, 1962.

[11] Carl Schissler, Ravish Mehra, and Dinesh Manocha, "High-order diffraction and diffuse reflections for interactive sound propagation in large environments", *ACM Transactions on Graphics (TOG)*, vol. 33, no. 4, pp. 39, 2014.

[12] Anish Chandak, Christian Lauterbach, Micah Taylor, Zhimin Ren, and Dinesh Manocha, "Ad-frustum: Adaptive frustum tracing for interactive sound propagation", *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1707–1722, 2008.

[13] Hengchin Yeh, Ravish Mehra, Zhimin Ren, Lakulish Antani, Dinesh Manocha, and Ming Lin, "Wave-ray coupling for interactive sound propagation in large complex scenes", *ACM Transactions on Graphics (TOG)*, vol. 32, no. 6, pp. 165, 2013.

[14] B Teng and R Eatock Taylor, "New higher-order boundary element methods for wave diffraction/radiation", *Applied Ocean Research*, vol. 17, no. 2, pp. 71–77, 1995.

[15] Sara R Martin, U Peter Svensson, Jan Slechta, and Julius O Smith, "A hybrid method combining the edge source integral equation and the boundary element method for scattering problems", in *Proceedings of Meetings on Acoustics 171ASA*. ASA, 2016, vol. 26, p. 015001.

[16] Atul Rungta, Carl Schissler, Nicholas Rewkowski, Ravish Mehra, and Dinesh Manocha, "Diffraction kernels for interactive sound propagation in dynamic environments", *IEEE transactions on visualization and computer graphics*, vol. 24, no. 4, pp. 1613–1622, 2018.

[17] Nicolas Tsingos and Jean-Dominique Gascuel, "Fast rendering of sound occlusion and diffraction effects for virtual acoustic environments", in *Audio Engineering Society Convention 104*. Audio Engineering Society, 1998.

[18] U Peter Svensson, Roger I Fred, and John Vanderkooy, "An analytic secondary source model of edge diffraction impulse responses", *The Journal of the Acoustical Society of America*, vol. 106, no. 5, pp. 2331–2344, 1999.

[19] Andreas Asheim and U Peter Svensson, "An integral equation formulation for the diffraction from convex plates and polyhedra", *The Journal of the Acoustical Society of America*, vol. 133, no. 6, pp. 3681–3691, 2013.

[20] Robert G Kouyoumjian and Prabhakar H Pathak, "A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface", *November*, vol. 88, pp. 1448–1461, 1974.

[21] Lakulish Antani, Anish Chandak, Micah Taylor, and Dinesh Manocha, "Efficient finite-edge diffraction using conservative from-region visibility", *Applied Acoustics*, vol. 73, no. 3, pp. 218–233, 2012.

[22] Nicolas Tsingos, Thomas Funkhouser, Addy Ngan, and Ingrid Carlbom, "Modeling acoustics in virtual environments using the uniform theory of diffraction", in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM, 2001, pp. 545–552.

[23] Micah Taylor, Anish Chandak, Zhimin Ren, Christian Lauterbach, and Dinesh Manocha, "Fast edge-diffraction for sound propagation in complex virtual environments", in *EAA auralization symposium*, 2009, pp. 15–17.

[24] Micah Taylor, Anish Chandak, Qi Mo, Christian Lauterbach, Carl Schissler, and Dinesh Manocha, "Guided multiview ray tracing for fast auralization", *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 11, pp. 1797–1810, 2012.

[25] J.-M. Valin, F. Michaud, and J. Rouat, "Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering", *Robot. Auton. Syst.*, vol. 55, no. 3.

[26] K. Okuma, A. Taleghani, N. d. Freitas, J. J. Little, and D. G. Lowe, "A boosted particle filter: Multitarget detection and tracking", in *ECCV*, 2004.

[27] S. Briere, J.-M. Valin, F. Michaud, and D. Létourneau, "Embedded auditory system for small mobile robots", in *ICRA*, 2008.