Non-Field-Of-View Indoor Sound Source Localization based on Reflection and Diffraction

Kuya Takami*[1], Tomonari Furukawa[1,2] Makoto Kumon[3] and Lin Chi Mak[4]

***Abstract*— This paper presents a new acoustic approach to locate a mobile target outside the field-of-view (FOV), or the non-field-of-view (NFOV) of an optical sensor, based on the reflection and diffraction signals. In this approach, a sensor platform determines a reflection time-difference-of-arrival (TDOA) and frequency dependent diffraction to two distinct observation likelihoods from a single target's sound. The fusion of these likelihoods, a joint acoustic observation likelihood, estimate the NFOV target probabilistically within the recursive Bayesian estimation (RBE) framework. The approach was formulated and derived mathematically. Through parametric studies in simulation, the potential of the proposed approach for practical implementation has been demonstrated by the successful localization of the sound source. Finally, a preliminary validation of sound separation was performed in a controlled experimental environment showing the difference between diffraction alone and combination of diffraction and reflection signals.**

I. INTRODUCTION

Sound localization and speech recognition are increased demand for variety of robotic applications where autonomous mobile robots are operating in human and natural settings. Complex environment, such as an indoor setting, produces difficulties to the target localization and tracking, or mobile target estimation. However, it has a variety of applications such as home security, home health care, and urban search-and-rescue with the limitation constrained by the complexity of the indoor structure [1]–[3]. Indoor structures make estimation problems challenging as they can introduce largely unobservable regions when an optical sensor like a camera is deployed. This is because optical sensors' field-of-view (FOV) is determined by the line-of-sight (LOS) and range of the optical sensor, which is small in highly constrained environments. However, in human settings, humans have the capability to perform searching and finding the target person who is not in the FOV by communicating with the person and estimating the location. In such a case, target estimation is mostly performed not through vision but audition.

Sound source localization traditionally focused strictly on LOS conditions based on both time and frequency domain approach. The most common approach utilizes the time-of-arrival (TOA)/time-difference-of-arrival (TDOA) information of acoustic signals [4]–[6]. Recent work has tackled non-field-of-view (NFOV) target estimation with use of the reflection signal within a known environment. Narang et al. [7] detected reflected sound effectively using a combination of an image model and dynamic environment map to assists in robot navigation. This approach identifies and estimate the direction of the sound source. However, it does not estimate the target position in NFOV and does not integrate the vision information for localization. Even et al. [8] localized the sound source by a ray tracing method based on the reflected signal arrival directions. However, their approach resulted in NFOV estimation with meter order accuracy and some inconsistent variations between trails.

For a numerical technique, existing approaches enhance NFOV target estimation by including a sensor with a limited FOV, such as an optical sensor. Mauler [9] stated the NFOV estimation problem mathematically, and Furukawa, *et al.* [10], [11] developed a generalized numerical solution. In this technique, the event of "no detection" is converted into an observation likelihood and utilized to positively update probabilistic belief on the target. This belief is dynamically maintained by the RBE. The technique, however, has been found to fail in target estimation unless the target is re-discovered within a short period after being lost. Takami et al. [12] incorporated an acoustic sensor to maintain belief with no optical detection with more reliability. Nevertheless, the technique performed poorly unless the target re-entered the optical FOV since the acoustic sensing is only conducted in an assistive capacity. Extending Kumon's approach, Takami, *et al.* [13] focused more towards complex indoor environment using interaural level difference (ILD) *a priori* fixed microphone array knowledge. However, this approach requires prior data collection and can not be applied to the dynamic sensor platforms.

This paper presents a new acoustic approach to estimate a NFOV mobile target using sound wave physical properties, reflection, and diffraction. In the approach, sound source reflection and diffraction signal TDOA and frequency dependent property construct two distinct observation likelihoods from target sound. The fusion of these likelihoods, a joint acoustic observation likelihood, is derived to perform the target estimation. The location of the NFOV target is finally estimated within the recursive Bayesian estimation framework. This process of target estimation was derived mathematically. Following the formulation, the proposed approach was tested through simulation and parametrically studied under multiple conditions. Finally, an experimental environment was constructed to identify the distinct differ-

*Corresponding author

[1]Department of Mechanical Engineering, Virginia Polytechnic Institute and State University, USA {kuya, tomonari}@vt.edu

[2]Center for Autonomous Systems, University of Technology, Sydney ,NSW, Australia

[3]Department of Mechanical System Engineering, Kumamoto University, Kumamoto, Japan makoto@gpo.kumamoto-u.ac.jp

[4]AP Photonics Limited,Science Park, Hong Kong maklinchi@gmail.com

ences between a diffraction signal alone and combination of diffraction and reflection signals.

II. NON-FIELD-OF-VIEW (NFOV)

A. Auditory Recursive Bayesian Estimation

The proposed approach is mathematically described as follows. Let the state of the robot s at time step k be $\bar{\mathbf{x}}_k^s \in \mathcal{X}^s$. Consider a target t, the state is given by $\mathbf{x}_k^t \in \mathcal{X}^t$, and a sequence of observations of the target t by the robot s from time step 1 to time step k given by ${}^s\tilde{\mathbf{z}}_{1:k}^t \equiv \{{}^s\tilde{\mathbf{z}}_\kappa^t | \forall \kappa \in \{1, ..., k\}\}$. The RBE represents belief on the target in the form of a probability density function and iteratively updates the belief in time and observation. Let the belief given a sequence of observations and the robot state at time step $k-1$ be $p\left(\mathbf{x}_{k-1}^t | {}^s\tilde{\mathbf{z}}_{1:k-1}^t, \bar{\mathbf{x}}_{k-1}^s\right)$. Chapman-Kolmogrov equation updates the prior belief in time, or predicts the belief at time step k, by the probabilistic motion model $p\left(\mathbf{x}_k^t | \mathbf{x}_{k-1}^t, \bar{\mathbf{x}}_{k-1}^s\right)$:

$$p\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_{1:k-1}^t, \bar{\mathbf{x}}_{k-1}^s\right) = \int_{\mathcal{X}^t} p\left(\mathbf{x}_k^t | \mathbf{x}_{k-1}^t, \bar{\mathbf{x}}_{k-1}^s\right) p\left(\mathbf{x}_{k-1}^t | {}^s\tilde{\mathbf{z}}_{1:k-1}^t, \bar{\mathbf{x}}_{k-1}^s\right) d\mathbf{x}_{k-1}^t. \quad (1)$$

Note that the motion model is $p\left(\mathbf{x}_k^t | \mathbf{x}_{k-1}^t\right)$ if the target is not reactive to the robot. The observation update, or the correction process, is performed using the Bayes theorem. The target belief is corrected using the new observation ${}^s\tilde{\mathbf{z}}_k^t$ as

$$p\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_{1:k}^t, \bar{\mathbf{x}}_k^s\right) = \frac{q\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_{1:k}^t, \bar{\mathbf{x}}_{k-1:k}^s\right)}{\int_{\mathcal{X}^t} q\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_{1:k}^t, \bar{\mathbf{x}}_{k-1:k}^s\right) d\mathbf{x}_k^t}, \quad (2)$$

where $q(\cdot) = l\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_k^t\right) p\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_{1:k-1}^t, \bar{\mathbf{x}}_{k-1}^s\right)$ and $l\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_k^t, \bar{\mathbf{x}}_k^s\right)$ represents the observation likelihood of $\mathbf{x}_k^t$ given ${}^s\tilde{\mathbf{z}}_k^t$, $\bar{\mathbf{x}}_k^s$.

$$l\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_k^t, \bar{\mathbf{x}}_k^s\right) = \prod_j l_j^a\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_k^t, \bar{\mathbf{x}}_k^s\right) \quad (3)$$

where $l_j^a(\cdot)$ are the likelihoods of jth acoustic sensor.

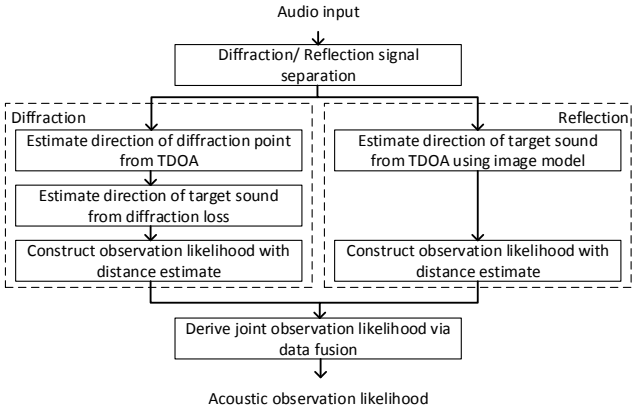


Fig. 1: Construction of auditory NFOV target likelihood

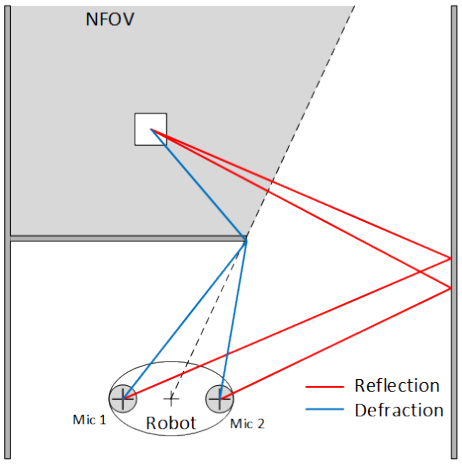
B. Construction of Auditory NFOV Target Observation Likelihood

Figure 1 shows the overview of the approach proposed for constructing a NFOV target observation likelihood using auditory sensors. The proposed approach extracts the first-arrival diffraction and reflection signals by taking the wave propagation physical properties into account. The approach begins with obtaining a time-domain signal of a relatively impulsive sound at each microphone. In each curve, notable peaks are then extracted as a candidate for first-arrival diffraction and reflection signals. When each candidate signal is described in the frequency domain, the first-arrival diffraction and reflection signals can be identified since they are the first signals that are correlated in the low-frequency range. The diffraction signal is then used to identify the so-called diffraction point by deriving the TDOA for each pair of microphones and further estimate the direction of target sound beyond the diffraction point from the loss of sound energy through diffraction, or the diffraction loss. An observation likelihood is eventually constructed by additionally estimating the distance from the sound magnitude and features. The reflection signal estimates the target direction directly from the TDOA by mirroring and creating a virtual target. It also creates an observation likelihood with distance estimate by considering the sound magnitude and characteristics and environmental properties. A joint observation likelihood is finally created by the fusion of the diffraction and reflection observation likelihoods.

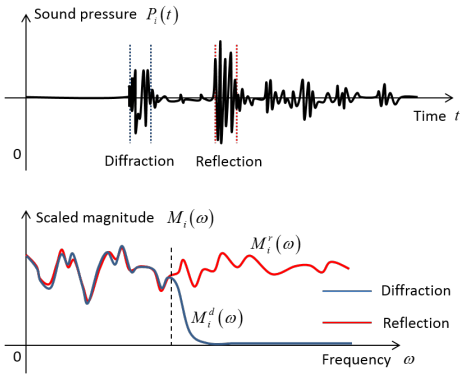
The proposed approach infers the location of the sound target using both the first-arrival diffracted and reflected sound signals. The next subsection describes the extraction of the first-arrival diffraction and reflection signals, followed by the target estimation using the diffracted and reflected signals in the subsequent two subsections. The final goal of this project is to develop a probabilistic RBE based framework, but the preliminary study has succeeded in the proof-of-concept in deterministic formulations. The two subsections will present the deterministic NFOV target estimation using diffraction and reflection sound waves. The final subsection derives the joint observation likelihood as a result of data fusion.

III. EXTRACTION OF FIRST-ARRIVAL DIFFRACTION AND REFLECTION SIGNALS

Figure 2 illustrates the extraction process of the diffraction and reflection signals in a simple scenario, where a robot carrying two microphones receives sound emitted by a NFOV target in a two-dimensional indoor environment with three walls (Figure 2(a)). As shown in the figure, sound waves emitted from the target reach the robot first through diffraction and second through reflection. If the sound is relatively impulsive, the first-arrival diffraction and reflection signals can be extracted clearly. The assumption of the sound specular reflection holds for wall texture which are smooth compared to the wavelength. Figure 2(b) shows not only the sound pressure in the time domain, $P_i(t)$, but also the magnitude of the resulting first-arrival diffraction

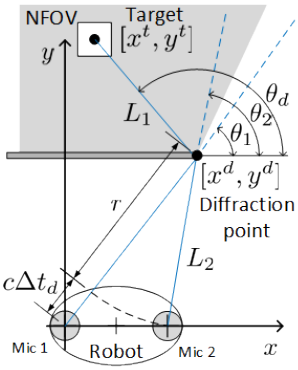


(a) Acoustic signals from NFOV target

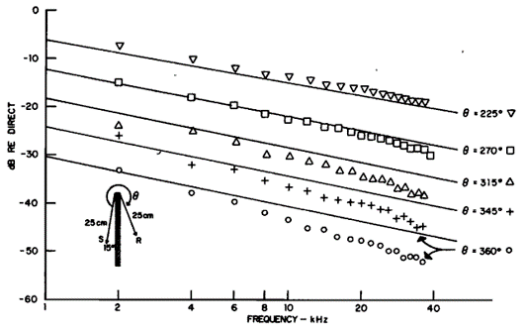


(b) Diffracted and reflected signals

Fig. 2: Auditory NFOV target observation



(a) Proposed approach



(b) Magnitude with different orientation angles [14]

Fig. 3: Estimation of sound direction from diffraction signals

and reflection signals in the frequency domain, $M_i^d(\omega)$ and $M_i^r(\omega)$, where $i \in \{1, 2\}$ is the index of microphone. Note that the magnitude is scaled to examine the correlation. Signals are considered to be from the same sound source if they share the same low frequency characteristics because low-frequency signals reflect and diffract. The proposed approach thus selects the first set of signals that have the same low-frequency characteristics but are dissimilar in high frequency as the first-arrival diffraction and reflection signals of all candidate signals. Diffraction signals have little high-frequency components whilst reflection signals see components in all frequencies. Each microphones, 1 and 2, constructs a different data set.

A. Estimation of Sound Direction from Diffraction Signals

Figure 3(a) shows the notations used for target estimation from diffraction signals in the scenario introduced in the last subsection. Since the diffraction signal microphones 1 and 2 receive is originated from the LOS location at which the sound diffracts, the proposed approach starts target estimation from diffraction signals with the selection of diffraction point from all candidates, which are corners of all structures. The measured quantity used for the selection is the TDOA, $\Delta t_d = t_{d2} - t_{d1}$, where t_{d1} and t_{d2} are the time-of-arrivals (TOAs) at Microphones 1 and 2 respectively. The diffraction point can be easily found from candidates as it satisfies the following equation:

$$\left(x^d\right)^2 + \left(y^d\right)^2 = \left(c\Delta t_d + r\right)^2. \tag{4}$$

where $\left[x^d, y^d\right]$ is the location of a candidate diffraction point, c is the speed of sound and r is a shorter distance between a microphone and the candidate diffraction point. With the diffraction point identified, the proposed approach further identifies the direction of the sound target from the diffraction point by analyzing the magnitudes of diffraction and reflection signals $M_i^d(\omega)$ and $M_i^r(\omega)$. The loss of high-frequency signal components is assumed to be less with a microphone closer to LOS, microphone 2 in this case, as there is no loss with a microphone in LOS of the sound target. Medwin [14] proved the validity of this assumption over a quarter of a century ago, as shown in Fig. 3(b). The magnitude of diffraction sound drops when the "degree of non-line-of-sight (NLOS)" represented by the orientation angle is increased. This makes the proposed approach define the diffraction loss as

$$L_i = \int \left[M_i^r(\omega) - M_i^d(\omega)\right] \mathrm{d}\omega \geq 0, \forall i \in \{1, 2\} \tag{5}$$

and associate it with the degree of NLOS. The work of Medwin also proved that the diffraction loss is approximately proportional to the degree of NLOS. The sound direction

from the diffraction point is given by

$$\theta_d = \theta_1 + \frac{\theta_2 - \theta_1}{L_1 - L_2} L_1. \tag{6}$$

B. Estimation of Sound Direction from Reflection Signals

Figure 4 shows the proposed approach for estimation of sound direction from reflection signals. Reflection makes the sound propagation and the subsequent target estimation complicated, but if the wall is smooth and yields specular reflection, the sound direction can be estimated easily by introducing a virtual target [15], which is located symmetrically to the real target relative to the wall of reflection. Let the position of the virtual target be $\left[\hat{x}^t, \hat{y}^2\right]$. The measured TDOA can be associated with the position of the virtual target as

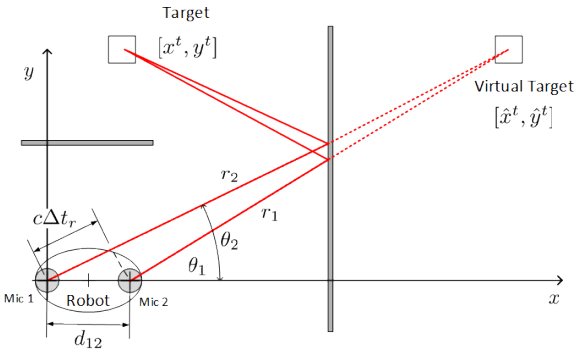


Fig. 4: Estimation of sound direction from reflection signals by proposed approach

$$\begin{cases} c\Delta t_r = \Delta r_{12} = \|r_1 - r_2\| \\ \|r_1\| \sin\theta_1 = \|r_2\| \sin\theta_2 \\ \|r_2\| \cos\theta_2 - \|r_1\| \cos\theta_1 = d_{12} \end{cases}, \tag{7}$$

where d_{12}, $\{r_1, r_2\}$ are the distance between Microphones 1 and 2 and vectors from sensor to the virtual sound source, respectively. Derivation attempted as a preliminary study for this project yields the relationship as a TDOA curve expressing r_1 as a function of θ_1

$$\|r_1(\theta_1)\| = \frac{-d_{12}^2 + \Delta r_{12}^2}{2(\Delta r_{12} + d_{12}\cos(\theta_1))} \tag{8}$$

Further mathematical manipulation shows that this equation asymptotically yields the sound direction as

$$\frac{\theta_2 + \theta_1}{2} = \lim_{\|r_1\| \to \infty} \tan^{-1} \frac{\hat{y}^t}{\hat{x}^t} = \cos^{-1} \frac{c\Delta t_d}{d_{12}}. \tag{9}$$

C. Construction of Joint Observation Likelihood through Data Fusion

While the sound can be better identified in direction rather than distance, it is also possible to make an estimate as to how far away the sound target is. The proposed approach estimates the positions by utilizing any available information including the magnitude, sound patterns, or sound features stored in a knowledge base and constructs an observation

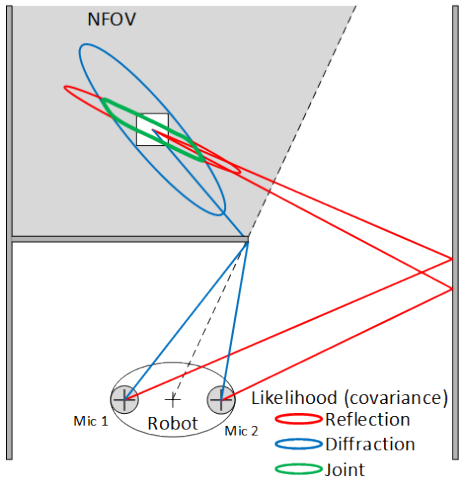


Fig. 5: Construction of joint observation likelihood through data fusion

likelihood for each of the diffraction and reflection signals by modeling uncertainties. For the j-th pair of microphones, the diffraction and reflection likelihoods are then combined to create an auditory joint observation likelihood via the canonical data fusion formula:

$$l_j^a\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_k^t, \bar{\mathbf{x}}_k^s, \bar{\mathbf{m}}_k\right) = l_j^d\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_k^t, \bar{\mathbf{x}}_k^s, \bar{\mathbf{m}}_k\right) l_j^r\left(\mathbf{x}_k^t | {}^s\tilde{\mathbf{z}}_k^t, \bar{\mathbf{x}}_k^s, \bar{\mathbf{m}}_k\right) \tag{10}$$

where $l_j^d(\cdot)$ and $l_j^r(\cdot)$ are the diffraction and reflection observation likelihood. Figure 5 illustrates the diffraction and reflection observation likelihoods as well as the joint observation likelihood where the observation likelihood is represented by an ellipsoid indicating a probability distribution with a covariance. The diffraction and reflection likelihoods are shown to have high eccentricity due to more accuracy in direction than in distance. Since the difference of the diffraction and reflection likelihoods in orientation may not be significant, the resulting auditory joint likelihood is also given by an ellipsoid with high eccentricity. However, the proposed approach, utilizing the diffraction and reflection physics of sound, could estimate the location of the sound target.

IV. Numerical Analysis

The preliminary simulation experiments demonstrate the validity of the proposed approach. The experiments were setup with the parameters in Table I. Figure 6 shows an example of localization. The dashed line, square, circles and red cross are the reflective wall, target, microphone sensors and target estimation, respectably. As shown in the figure, diffraction and reflection observation likelihood construct the green lined joint likelihood to estimate the target location under NFOV condition. Based on the initial simulation, the parametric study was performed to measure the sensitivity of estimation based on the sensor distance d_{12} in a range of 1 cm to 50 cm. The sensor position was also varied at three

different locations. Figure 7 shows the certainty of the target estimation as KL-Divergence at different sensor location with increasing d_{12}. As shown in the figure, increasing the distance between the microphone leads to a better estimate of the target. This is expected because increasing d_{12} will increase the time of arrival difference, resulting in better angle estimation. However, increasing the sensor position in y-direction lead to the reduction in KL-divergence.

TABLE I: Dimensions and parameters for the simulation

Parameter	Value
x^t	$[0.5m, 1.5m]$
$[x^d, y^d]$	$[1m, 1m]$
x_{wall}	$2m$
x_s	$[0.6m, 0.3m]$

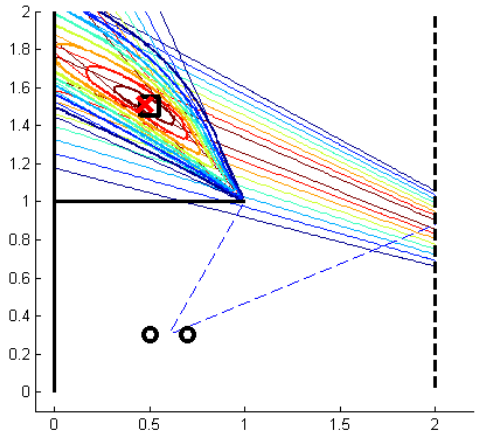


Fig. 6: Target estimation

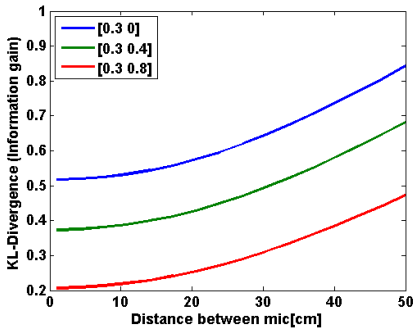
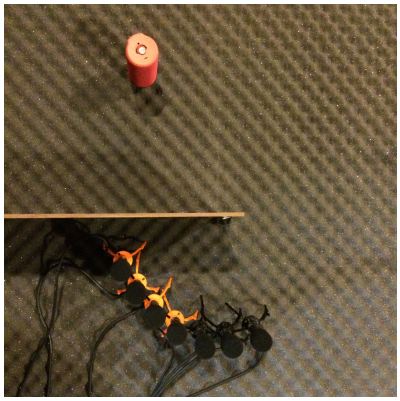
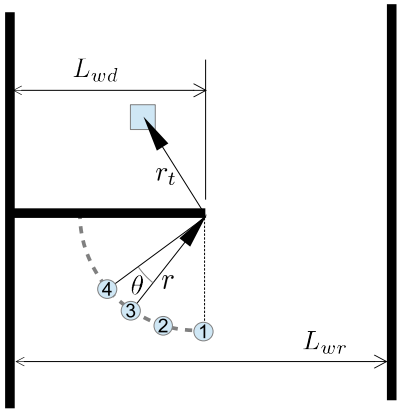


Fig. 7: d_{12} variation effect in certainty

Figure 8(a) show the experimental testing environment. As shown in the figure, the pink cylinder on top is the omni-directional target sound source, and separator wall and microphone array shown below. The sound insulator was installed at the top and the bottom of the environment to mitigate additional reflection and minimize reverberation. A removable smooth reflective wall was used for introducing the reflection. The experiments included microphones in a circular pattern as shown in Fig. 8(b) with dimensions in Table II.



(a) Setup



(b) Environemnt parameters

Fig. 8: Experimental environment

TABLE II: Dimensions and parameters for the experiments

Parameter	Value
L_{wd}	62cm
r_t	$[-20\text{cm}, 25\text{cm}]$
$\|r\|$	35cm
θ	15°
L_{wr}	112cm

Figure 9 depicts incoming sound to the microphone array without the reflection wall. The figure consist of a reference microphone at the target and the microphone array signal. As shown in the figure, the diffraction loss increased when angle of diffraction is enlarged from mic-1 to mic-4. The region A in particular displays first arrival diffraction signal. The diffraction loss is more clearly identified in the frequency domain depicted in Fig. 10.

Figure 11 is a comparison of mic-2 signals with and without the reflection wall. The figure clearly shows the effect of the wall between those two cases where region A only contains diffraction, and region B contains mixture of diffraction and reflection. The duration of region A was

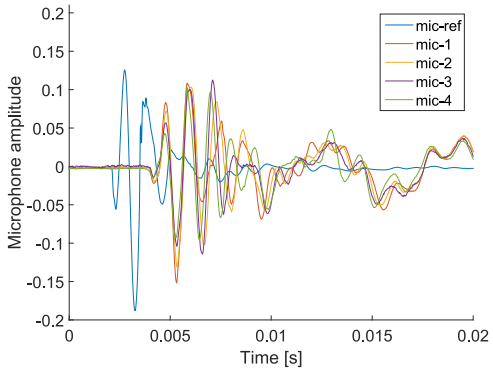


Fig. 9: Reference mic signal at the target and signal of mic array

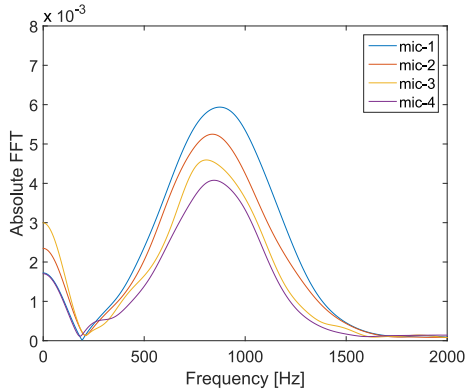


Fig. 10: Magnitude with different orientation angles

measured to be less than 5% error from the actual distance measurements.

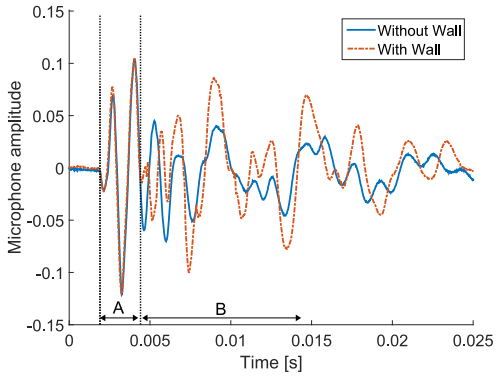


Fig. 11: Signal with and without the reflective wall

V. CONCLUSIONS

This paper demonstrated the target estimation capability of a new acoustic approach in NFOV using sound wave physical properties, reflection, and diffraction. This approach using sound source reflection and diffraction signal to construct two distinct observation likelihoods from the target sound. The joint acoustic observation likelihood can then be used to estimate the target location. This process was formulated mathematically with two-dimensional assumptions. The proposed approach was tested and parametrically studied under different conditions in the simulation. Finally, an experimental environment was constructed for further validation.

Future work consists of extending the application to an indoor environment. This includes the incorporation of an actual mobile target estimation using a mobile robot equipped1 with a microphone array.

REFERENCES

[1] N. B. Priyantha, H. Balakrishnan, E. D. Demaine, and S. Teller, "Mobile-assisted localization in wireless sensor networks," in *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, vol. 1. IEEE, 2005, pp. 172–183.

[2] H. M. Khoury and V. R. Kamat, "Evaluation of position tracking technologies for user localization in indoor construction environments," *Automation in Construction*, vol. 18, no. 4, pp. 444–457, 2009.

[3] S. Argentieri, P. Danes, P. Soueres *et al.*, "A survey on sound source localization in robotics: from binaural to array processing methods," 2014.

[4] J.-S. Hu, C.-Y. Chan, C.-K. Wang, M.-T. Lee, and C.-Y. Kuo, "Simultaneous localization of a mobile robot and multiple sound sources using a microphone array," *Advanced Robotics*, vol. 25, no. 1-2, pp. 135–152, 2011.

[5] D. B. Ward, E. A. Lehmann, and R. C. Williamson, "Particle filtering algorithms for tracking an acoustic source in a reverberant environment," *Speech and Audio Processing, IEEE Transactions on*, vol. 11, no. 6, pp. 826–836, 2003.

[6] L. C. Mak and T. Furukawa, "Non-line-of-sight localization of a controlled sound source," in *Advanced Intelligent Mechatronics, 2009. AIM 2009. IEEE/ASME International Conference on*. IEEE, 2009, pp. 475–480.

[7] G. Narang, K. Nakamura, and K. Nakadai, "Auditory-aware navigation for mobile robots based on reflection-robust sound source localization and visual slam," in *Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on*. IEEE, 2014, pp. 4021–4026.

[8] J. Even, Y. Morales, N. Kallakuri, C. Ishi, and N. Hagita, "Audio ray tracing for position estimation of entities in blind regions," in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*. IEEE, 2014, pp. 1920–1925.

[9] R. Mauler, *Recent Developments in Cooperative Control and Optimizatio*. Kluwer Academic Publishers, Norwell, MA, 2003, ch. Objective Functions for Bayesian Control-Theoretic Sensor Management, II: MHC-Like Approximation, pp. 273–316.

[10] T. Furukawa, F. Bourgault, B. Lavis, and H. F. Durrant-Whyte, "Recursive bayesian search-and-tracking using coordinated uavs for lost targets," in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*. IEEE, 2006, pp. 2521–2526.

[11] T. Furukawa, L. C. Mak, H. Durrant-Whyte, and R. Madhavan, "Autonomous bayesian search and tracking, and its experimental validation," *Advanced Robotics*, vol. 26, no. 5-6, pp. 461–485, 2012.

[12] K. Takami, T. Furukawa, M. Kumon, D. Kimoto, and G. Dissanayake, "Estimation of a nonvisible field-of-view mobile target incorporating optical and acoustic sensors," *Autonomous Robots*, 2015.

[13] K. Takami, T. Furukawa, M. Kumon, and G. Dissanayake, "Non-field-of-view acoustic target estimation in complex indoor environment," in *Proc. of the Int. Conf. on Field and Service Robotics*, 2015.

[14] H. Medwin, "Shadowing by finite noise barriers," *The Journal of the Acoustical Society of America*, vol. 69, no. 4, pp. 1060–1064, 1981.

[15] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, June 1997.

ACKNOWLEDGMENT

The authors wish to acknowledge the support of the research team member, Hanxing Liu, for his efforts and very productive work and development of the experimental system and data acquisition.