Sound Source Localization and Its Applications for Robots

Sung-eui Yoon
Associate Professor
KAIST

http://sgvr.kaist.ac.kr

KAIST

Sound Source Localization and Its Applications for Robots

- **Acoustic signals are everywhere**
- **Relative little work on utilizing acoustic cues for robots**
- **Recent developments thanks to AI speakers**
  - **Limited to direct sound signals or near-field speech recognition**

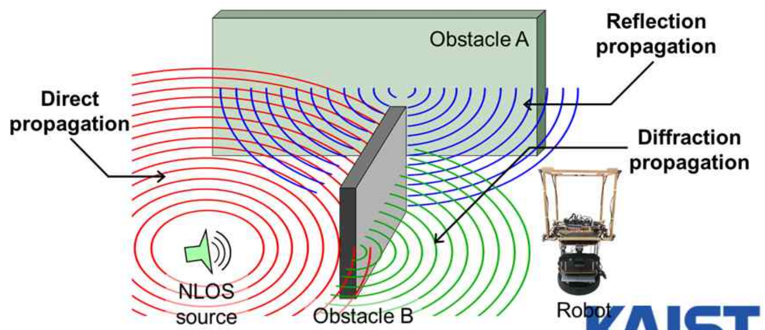
Image: robot 'Curi' of Mayfield robotics

- **Proposed half-day workshop to build a community on the topic and facilitate further research**
  - **Organized by me and Dinesh Manocha**

KAIST

Topics

- Interactive sound propagation technologies
- Localization applications for drone and underwater robots
- Deep learning techniques for multi-party interactions
- Real-time localization techniques considering non line of sight wave effects (diffraction)
- Joint optimizations between SLAM and acoustic approaches.
- Audio-motor localization



Schedule

- **8:30, Overview of the workshop**
- **8:50, Factor graphs for SLAM with visual and acoustic sensors**
  - **Frank Dellaert, Professor, Georgia Tech., USA**
- **9:10, Direct and inverse sound propagation methods**
  - **Dinesh Manocha, Professor, Univ. of Maryland at College Park, USA**

- **9:30, Poster Highlight Presentation**
- **10:00, Afternoon Tea and Poster Presentation**

KAIST

Schedule

- **10:30, Ray tracing based Sound Source Localization, Sung-eui Yoon**
- **10:50, Deep learning for robust audio perception in human-robot interactions**
  - **Jean-Marc Odobez, Idiap research institute and EPFL, Switzerland**
- **11:10, How can audio-motor binaural localization be made "active"?**
  - **Patrick Danes, Université de Toulouse and LAAS-CNRS, France**
- **11:30, Robot-Human Interaction based on Sound Source Localization**
  - **Hiroshi G. Okuno, Waseda Univ., Japan**

KAIST

Schedule

- **11:50, From Robot Audition to Drone Audition**
  - **Kazuhiro Nakadai, Honda Research Institute, and Tokyo Institute of Tech., Japan**
- **12:10, How can we quickly find an underwater acoustic source with a single mobile hydrophone?**
  - **Mandar Chitre, National University of Singapore,**
- **12:30, Closing**

Notice to Speakers

- **20 min including Q&A for each talk**

- **4 min for each poster talk**

1, A Bayesian System for Noise Robust Binaural Speaker Counting for Humanoid Robots

2, Informative Path Planning for Source Localization

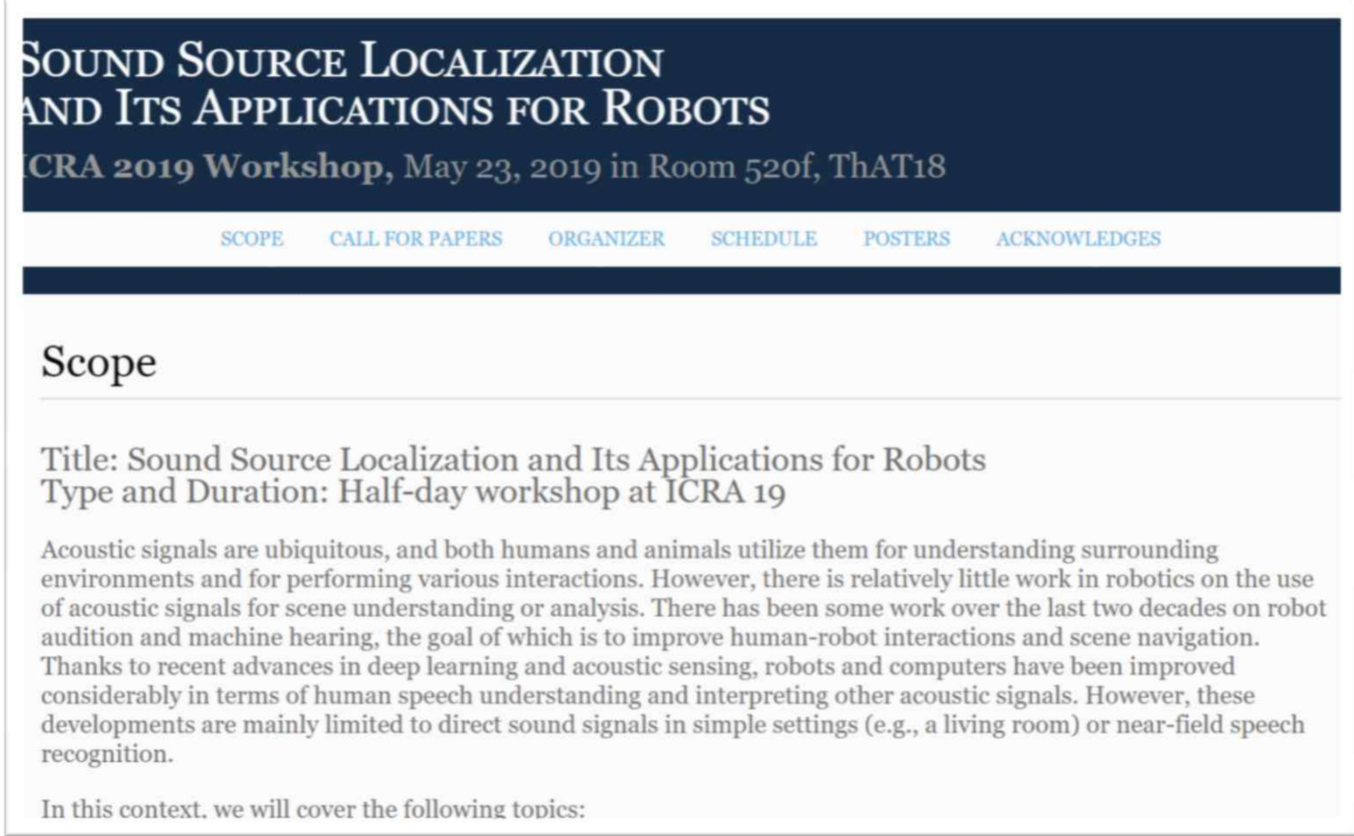
3, Sound Source Tracking Using Multiple Microphone Arrays Mounted to an Unmanned Aerial Vehicle

4, Design of a Microphone Array for Rollin' Justin

5, Robust Sound Source Localization considering Similarity of Back-Propagation Signals

Resource

- **Workshop homepage**
  - **https://sglab.kaist.ac.kr/SSL_ICRA19/**
    - **Talk slides and posters will be available**



SOUND SOURCE LOCALIZATION
AND ITS APPLICATIONS FOR ROBOTS

ICRA 2019 Workshop, May 23, 2019 in Room 520f, ThAT18

SCOPE | CALL FOR PAPERS | ORGANIZER | SCHEDULE | POSTERS | ACKNOWLEDGES

Scope

Title: Sound Source Localization and Its Applications for Robots
Type and Duration: Half-day workshop at ICRA 19

Acoustic signals are ubiquitous, and both humans and animals utilize them for understanding surrounding environments and for performing various interactions. However, there is relatively little work in robotics on the use of acoustic signals for scene understanding or analysis. There has been some work over the last two decades on robot audition and machine hearing, the goal of which is to improve human-robot interactions and scene navigation. Thanks to recent advances in deep learning and acoustic sensing, robots and computers have been improved considerably in terms of human speech understanding and interpreting other acoustic signals. However, these developments are mainly limited to direct sound signals in simple settings (e.g., a living room) or near-field speech recognition.

In this context, we will cover the following topics:

Reflection and Diffraction Aware Sound Source Localization using Ray Tracing

Sung-eui Yoon

Associate Professor

KAIST

http://sgvr.kaist.ac.kr

Motivation

- **Robots to interact with humans**
  - The importance of communication between users and robots using "sound"

An example of communicating with home robots and children

Image: robot 'Curi' of Mayfield robotics

Motivation

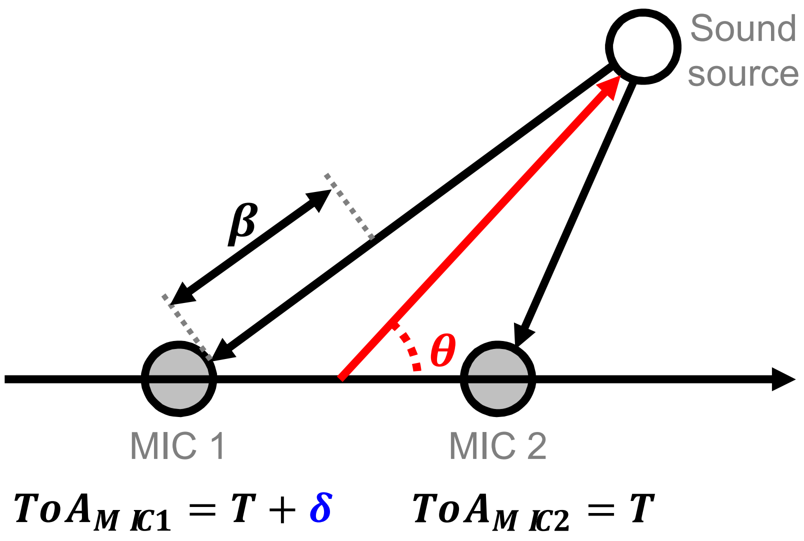
- **Robots to interact with humans**
  - Needs for techniques of localizing people using "Sound"

An example of needs for localizing users using sound

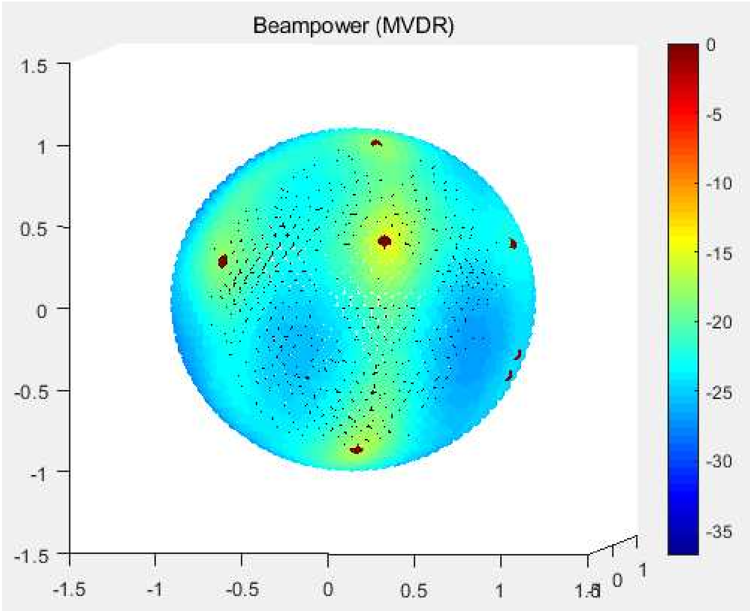
Image: Xiaofei et al, Sound Source …, IROS 2011

Related works

- **Estimating the incoming direction of sound based on a microphone array**
  - In many prior works, they have tried to estimate the <u>incoming direction of sound</u>



Time Difference Of Arrival (TDOA) method



Beamformer algorithm

Motivation

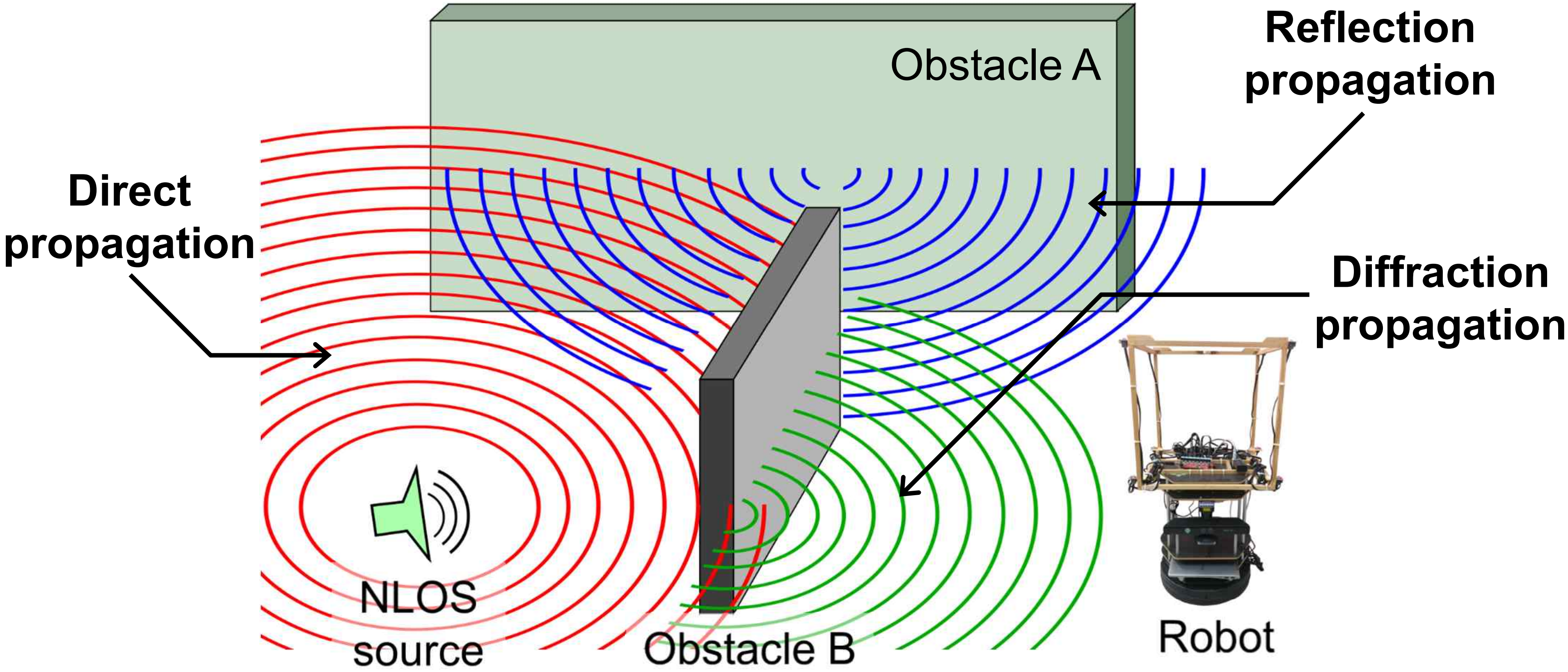
- **Needs for novel sound localization techniques**
  - In real environments, there are many obstacles causing invisible areas for a robot
  - Should be able to localize the sound source behind the obstacle (the non-line-of-sight source)

Image: Room service robot, Savioke

An example of needs to localize the NLOS source

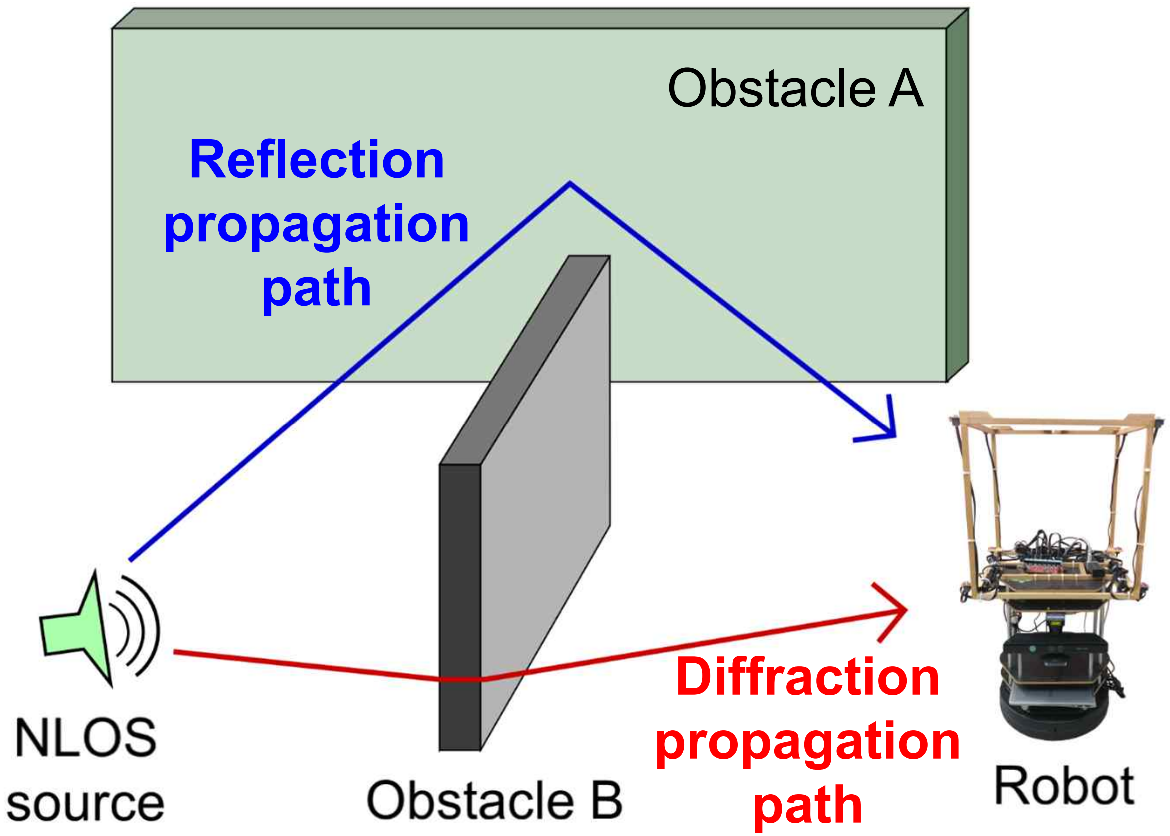
Sound Propagation

- Sound propagates through various paths (e.g., direct, diffraction, reflection)



Simplification into Acoustic Ray Paths

- The prominent propagation paths can be approximated by ray paths



My General Research Topics: Scalable Ray Tracing, Image Search, Motion Planning

- **Designing *scalable graphics and geometric algorithms* to efficiently handle massive models on commodity hardware**

Photo-realistic rendering

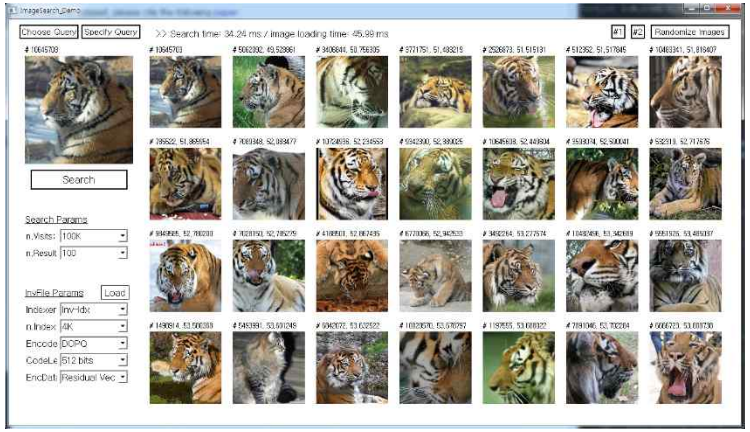


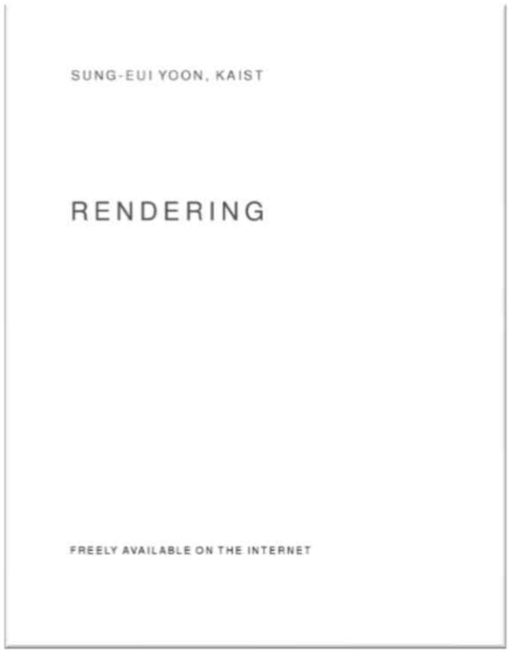
Image search

Motion planning

KAIST

Ray Tracing for Physically-based Rendering

- **Has been studied for many decades**
- **Adopted in movie industry, and used a lot in recent CG movies**

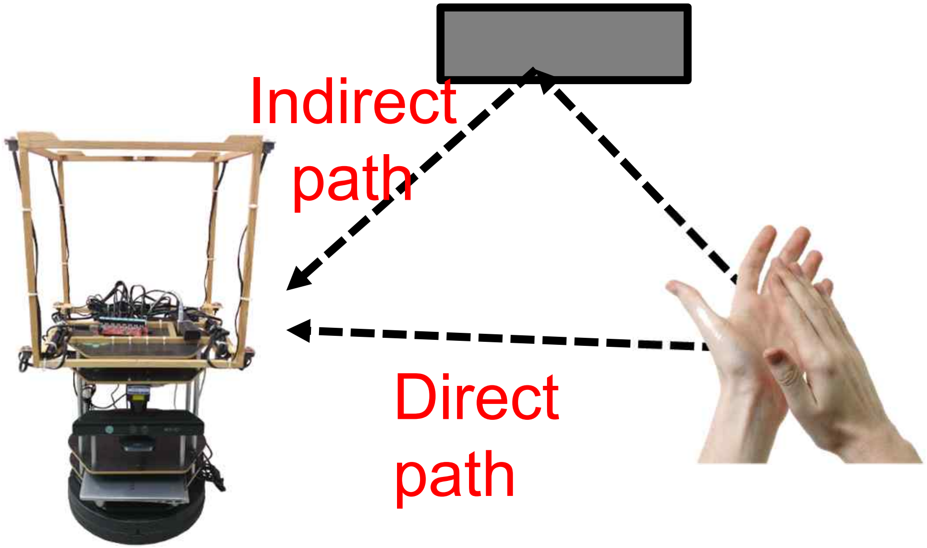


Rendering
Sung-eui Yoon
1st edition, July 2018, 148 pages
Freely available on the internet
Copyright 2018

PovRay

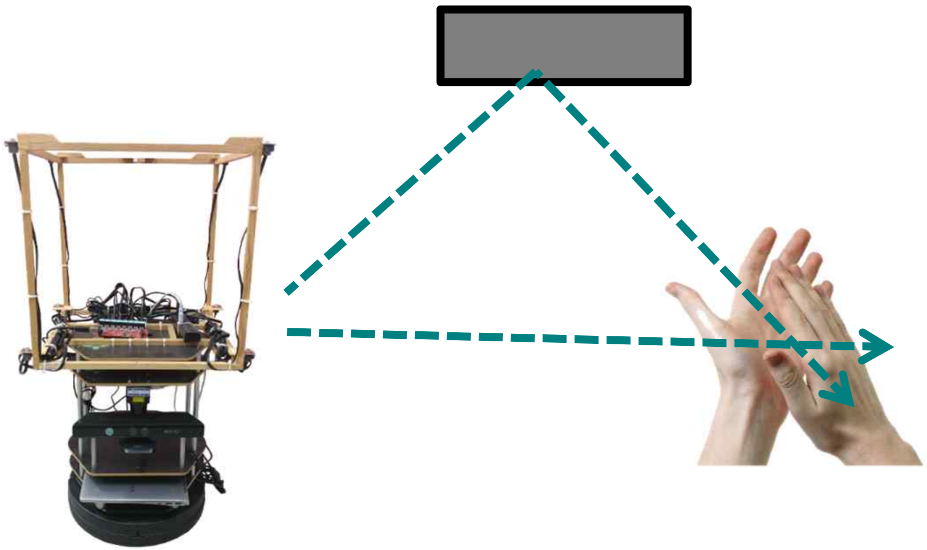
Reflection-Aware Sound Source Localization [ICRA 18]

- Collect the direct and indirect directions of the sound from a TDOA-based method.



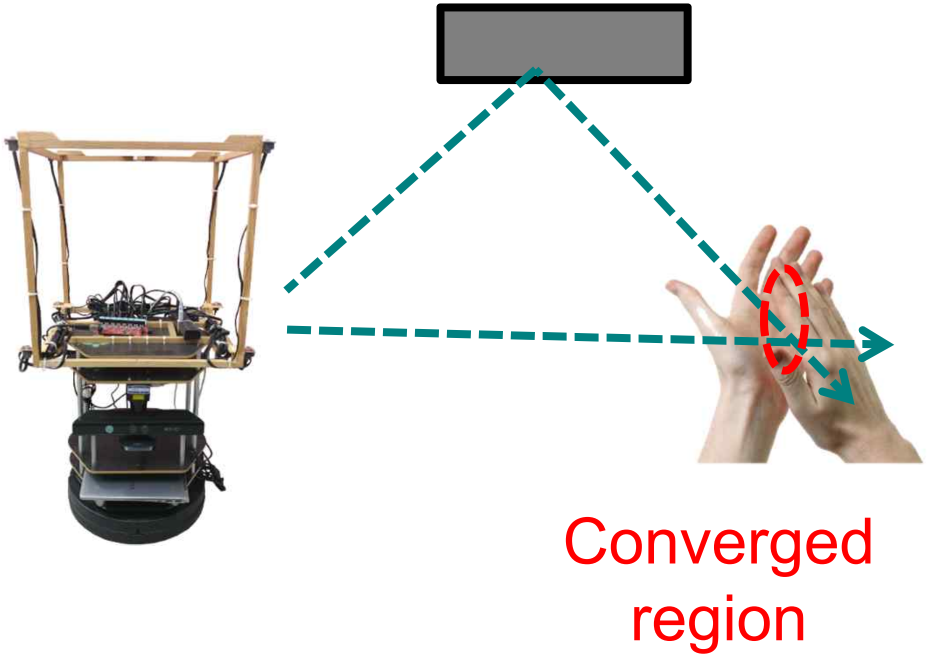
Key Idea of Our Approach

- Propagate acoustic rays to the free space considering specular reflection.



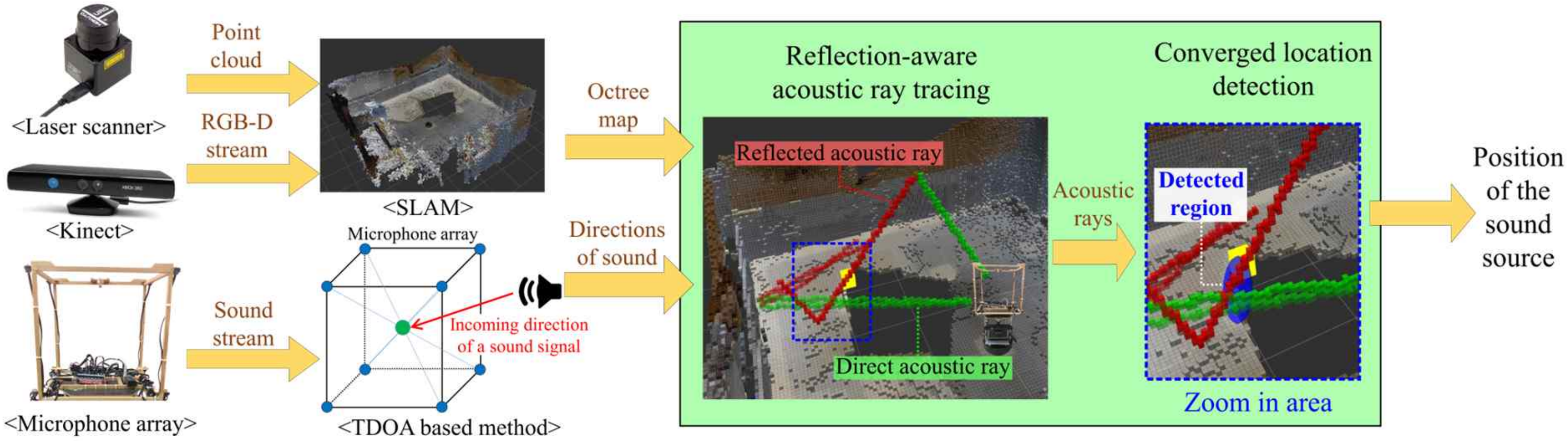
Key Idea of Our Approach

- Find the converged region of rays, and estimate the region as a source position.



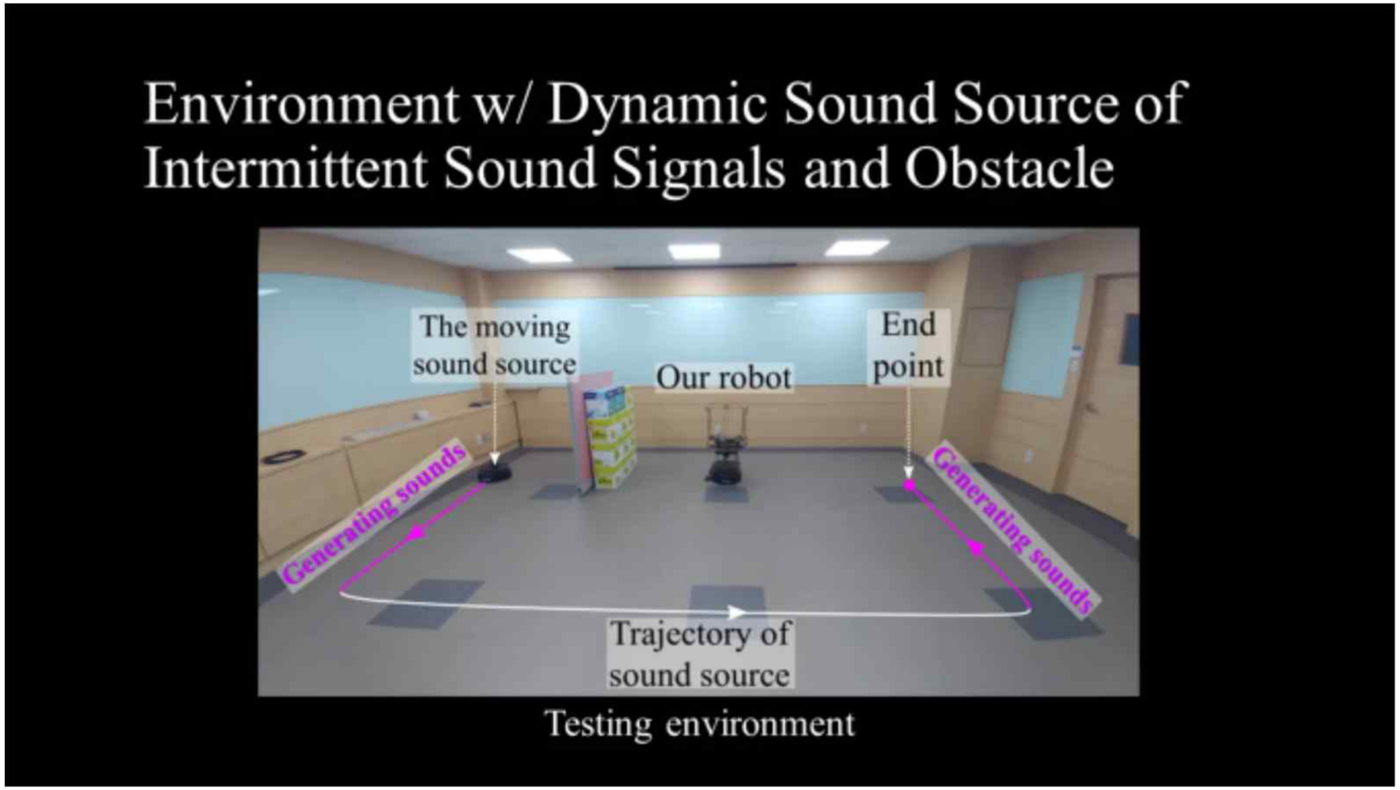
Overview of Our Approach

- Input: Octree map(SLAM) and Directions of sounds(TDOA based SSL)
- Acoustic rays are generated by Reflection-aware acoustic ray tracing
- The position of the sound source is estimated on Converged location detection



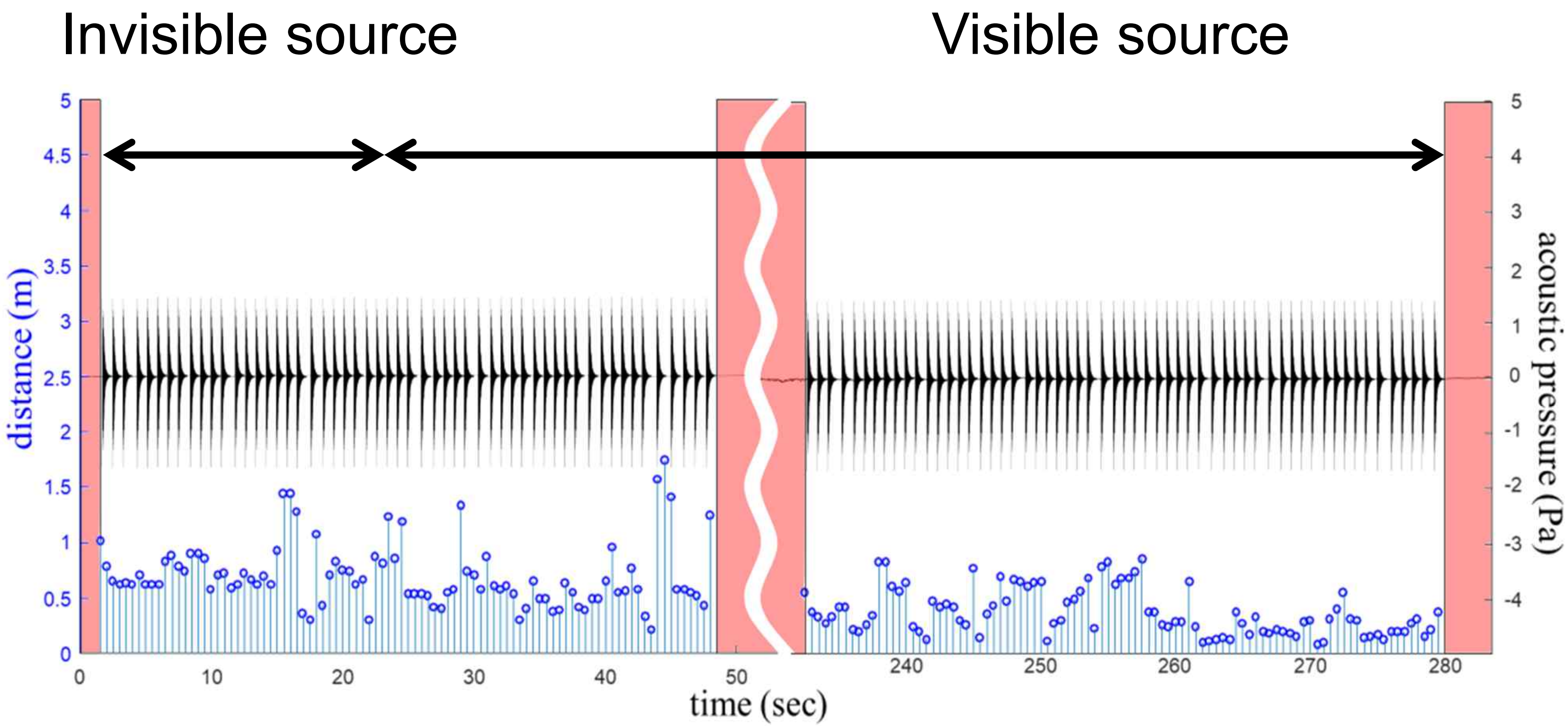
Result: Dynamic Sound Source and Obstacle

- Test on the indoor environments with a dynamic sound source of the intermittent signals and an obstacle



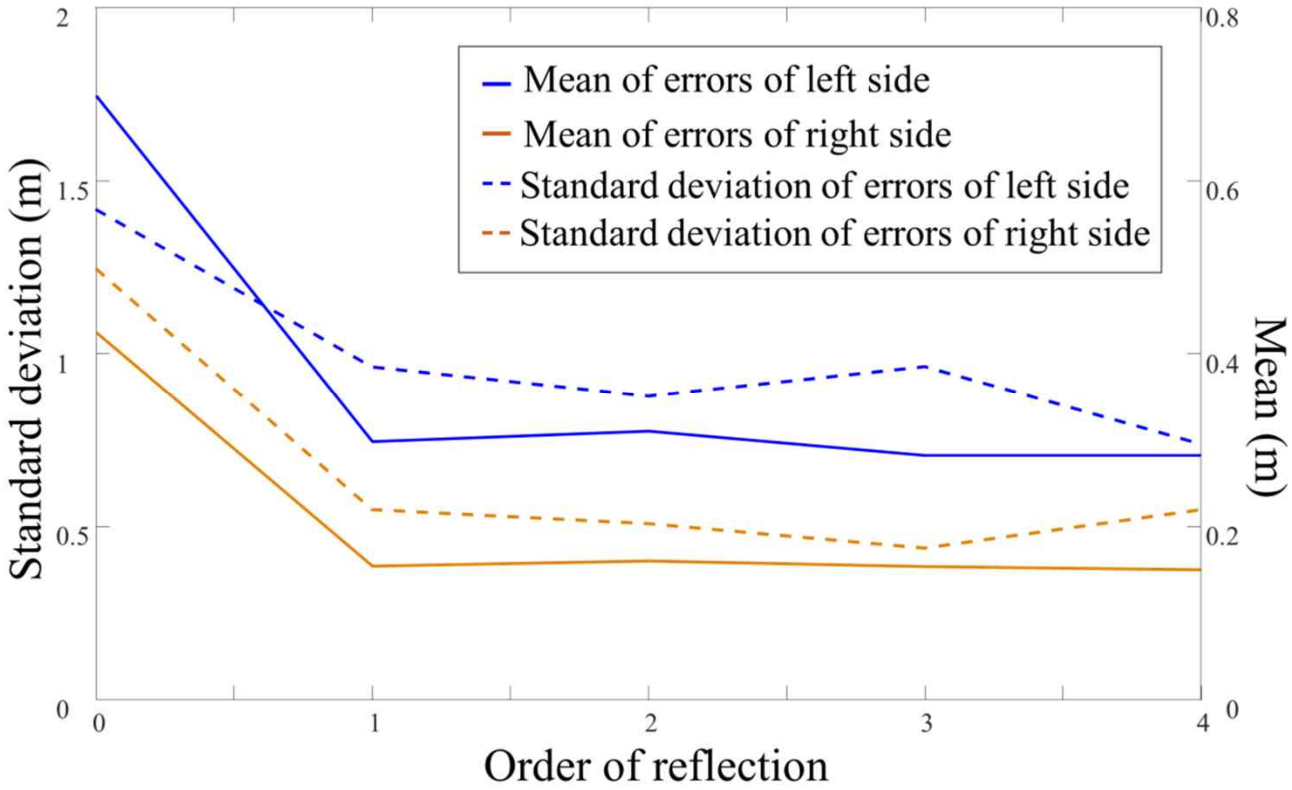
Result: Dynamic Sound Source and Obstacle

- The average error of the left side is 0.7m, and right side is 0.3m.



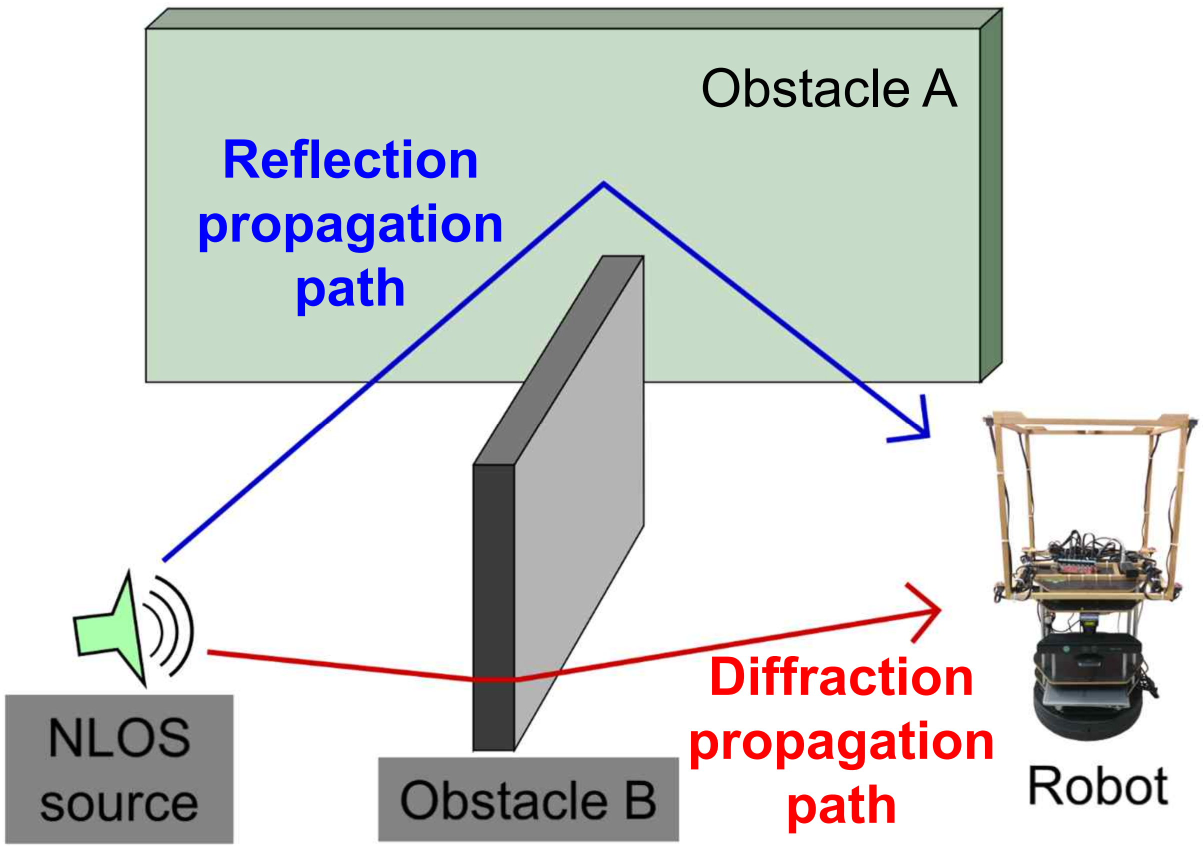
Result: Dynamic Sound Source and Obstacle

- Reflected rays improves the localization accuracy by 40% over only using direct rays.



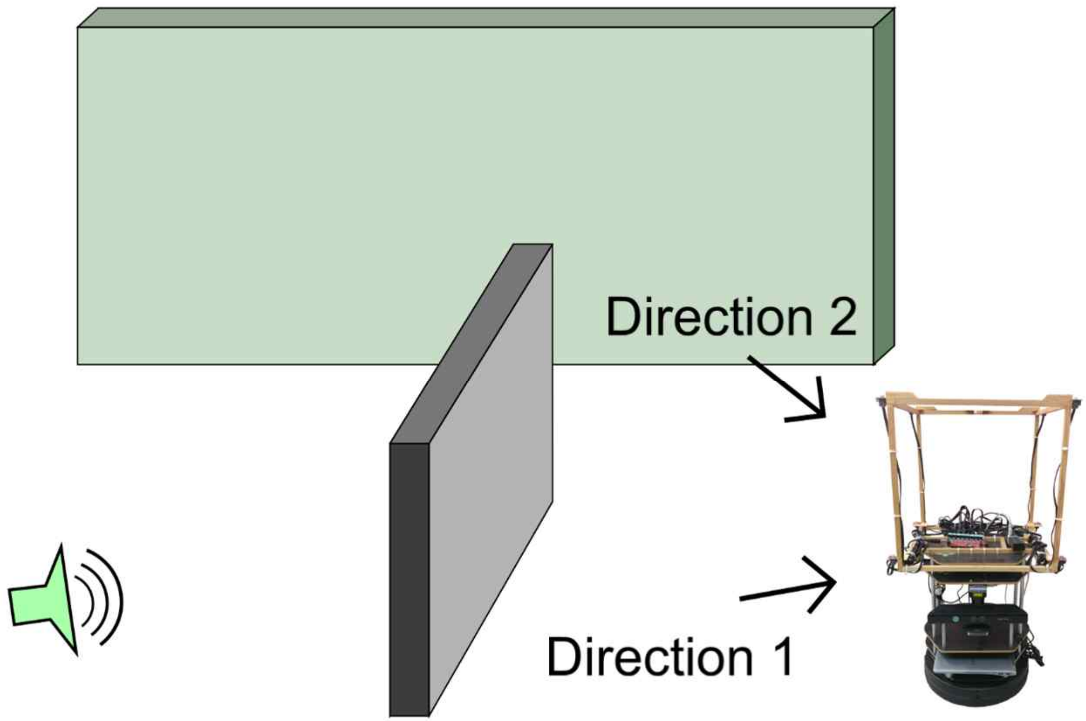
Extended for Supporting Diffraction [ICRA 19]

- The prominent propagation paths can be approximated by ray paths



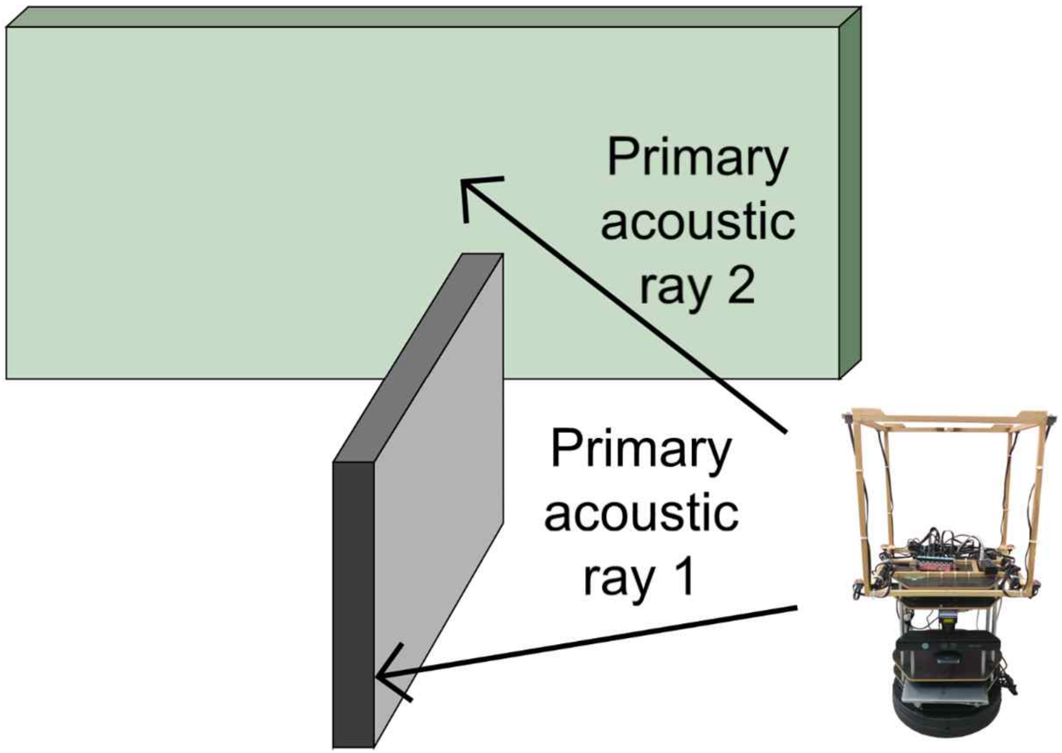
Key Idea of Our Approach

- **Diffraction-aware SSL for a NLOS source**
  - Estimate the incoming directions of sound from a signal measured by the microphone array; using a TDOA (Time Difference Of Arrival) method

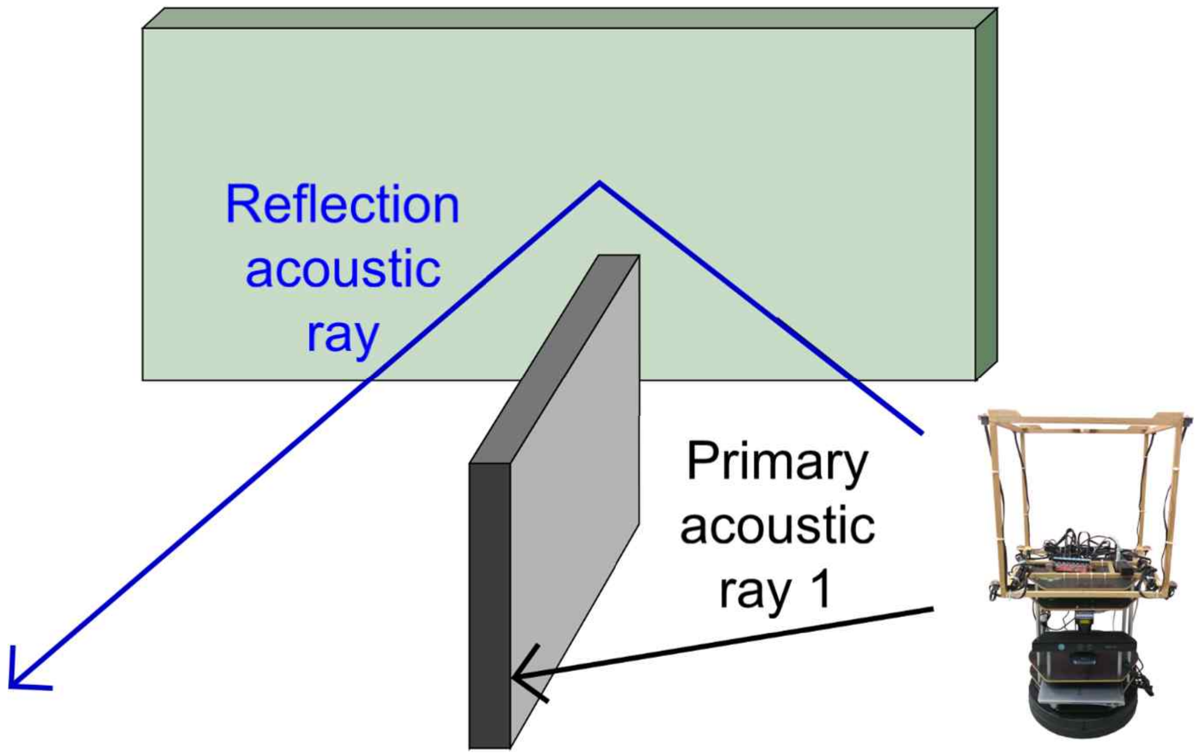


Diffraction-aware SSL for a NLOS source

- Generate the primary acoustic rays into reverse directions of the incoming directions of sound

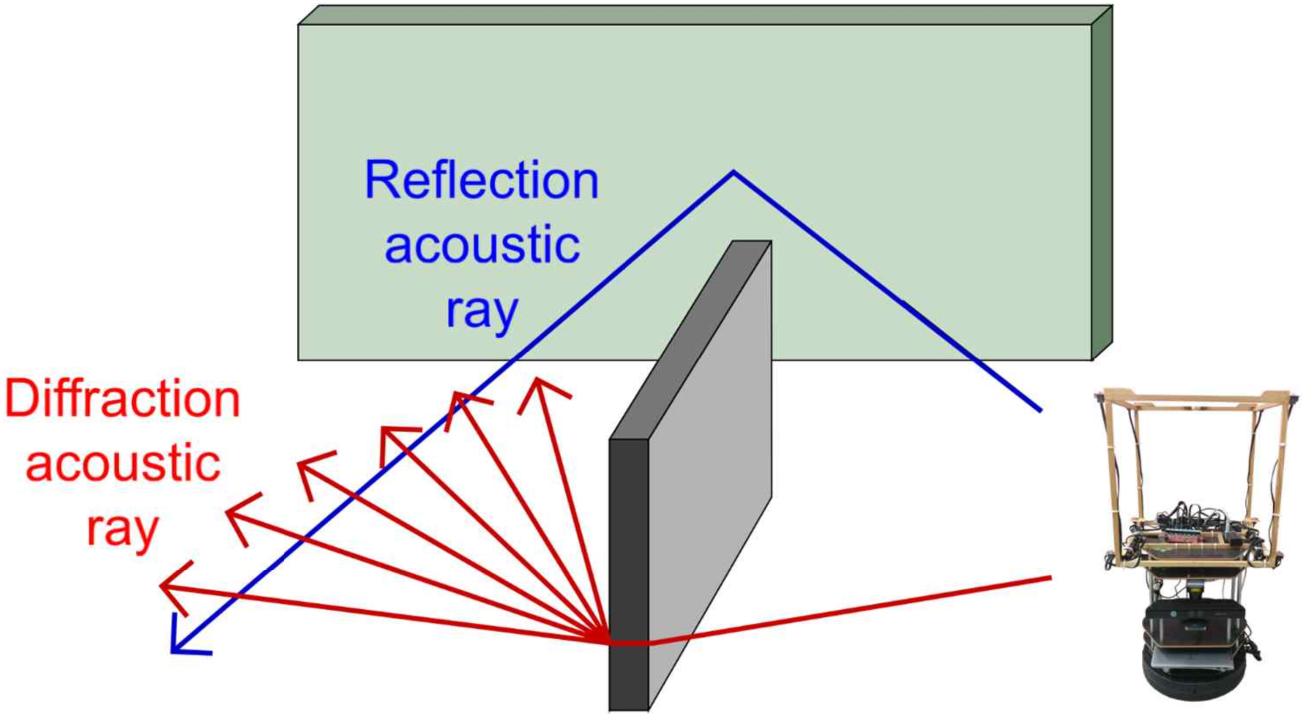


Reflection Acoustic Rays



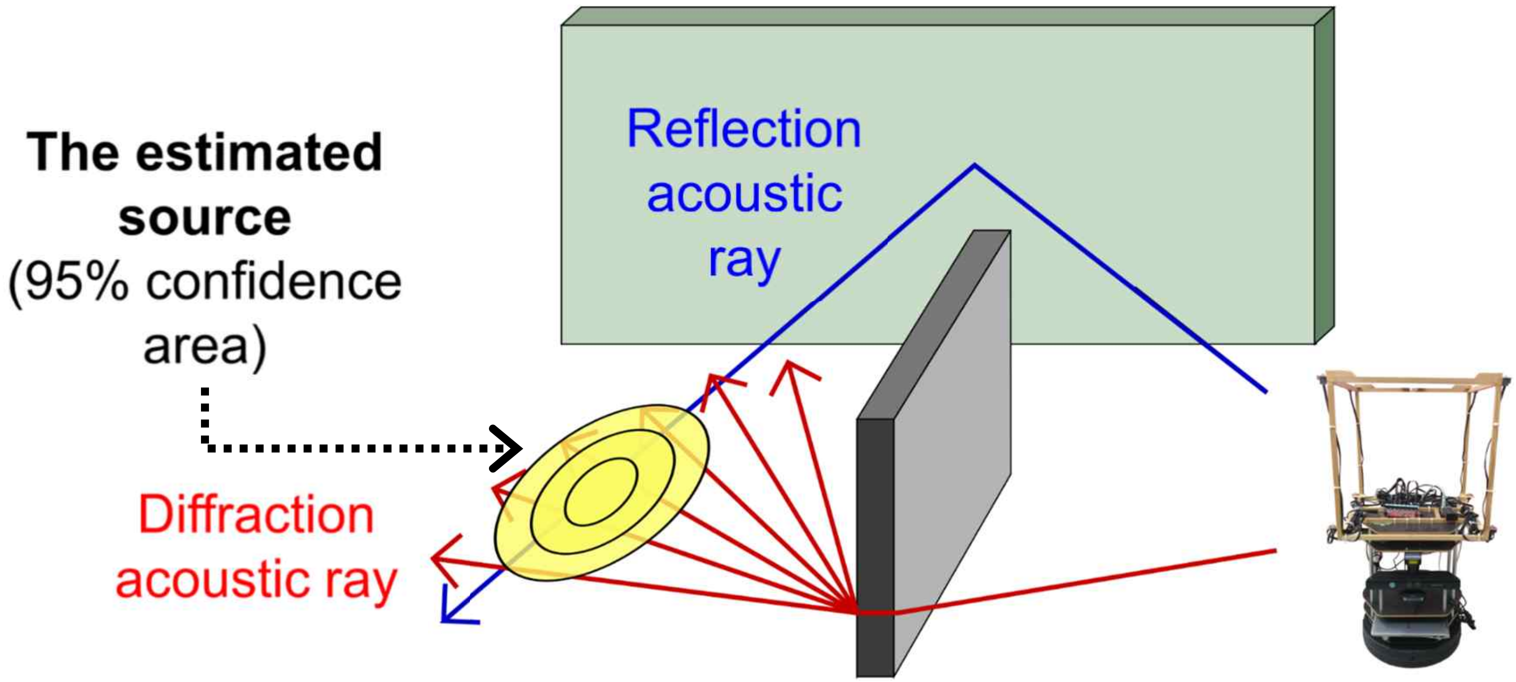
Diffraction Acoustic Rays

- Check whether the primary acoustic ray becomes close enough to the wedge of the obstacle
- If so, generate the diffraction ray based on the Uniform Theory of Diffraction model



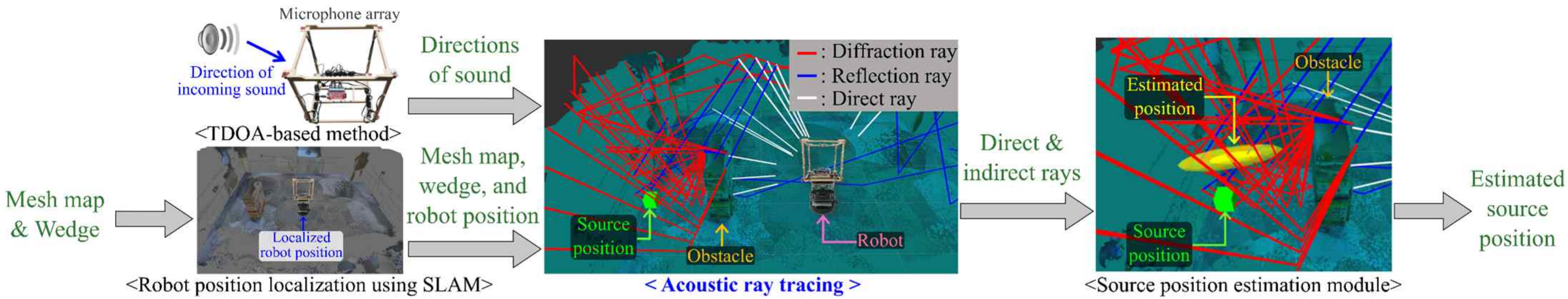
Estimating a Converged Region as Source Location

- Acoustic rays are candidates of the sound propagation paths
- The convergence region of acoustic rays are determined by estimated source position
  - Particle filter is used for identifying the region



Overview of Our Approach

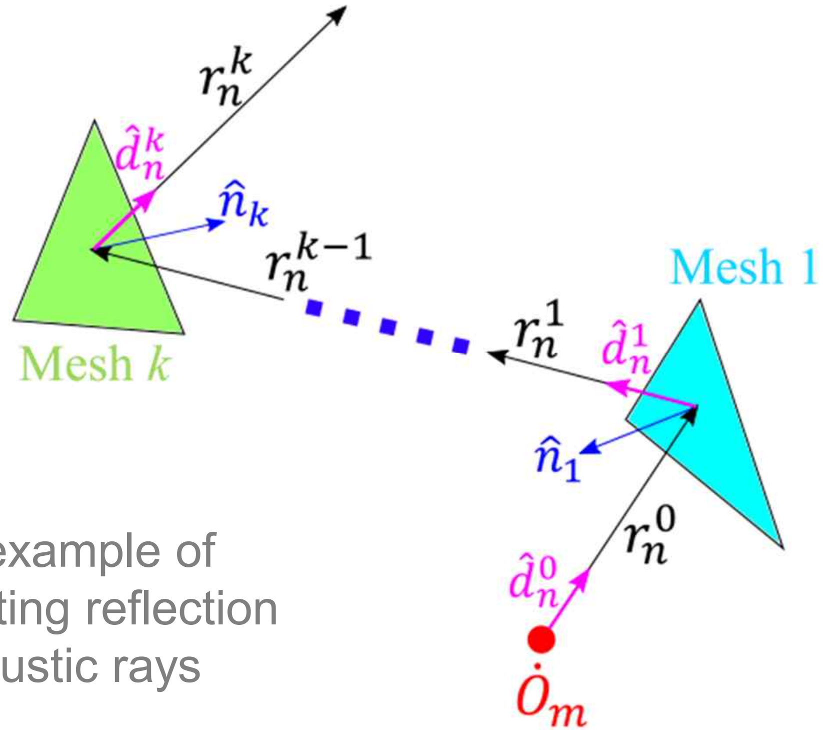
- Inputs are point cloud and audio stream
  - The indoor environment is reconstructed by using SLAM
- The incoming direction of sound is estimated by the TDOA method
- Based on the mesh map (reconstructed environment) and incoming directions, generate the acoustic rays



Overview of our approach

Reflection acoustic rays

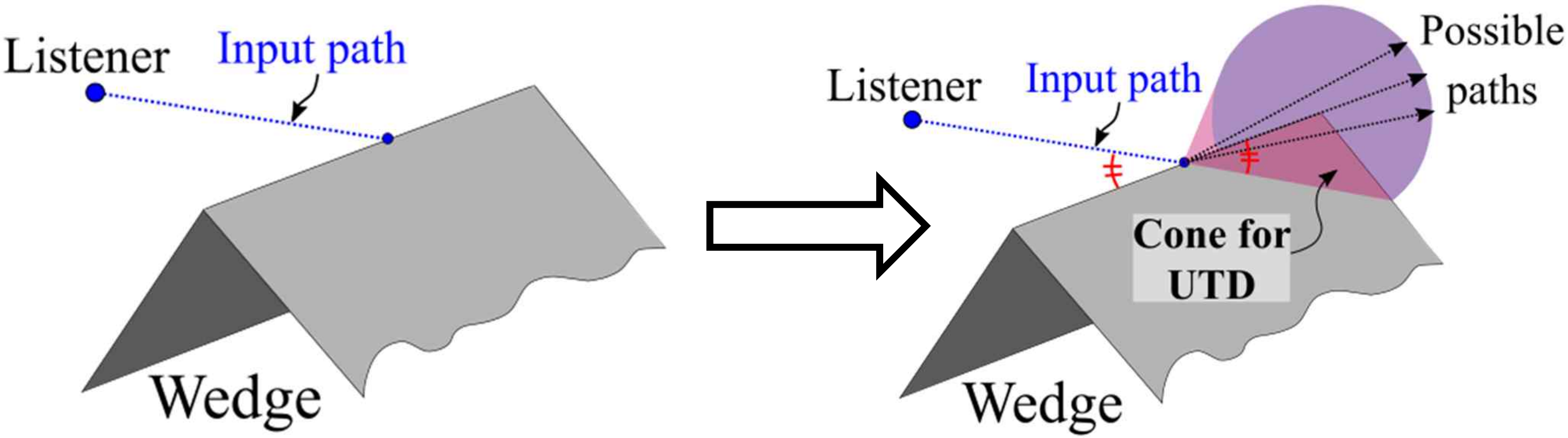
- Generate primary acoustic rays assuming the specular reflection



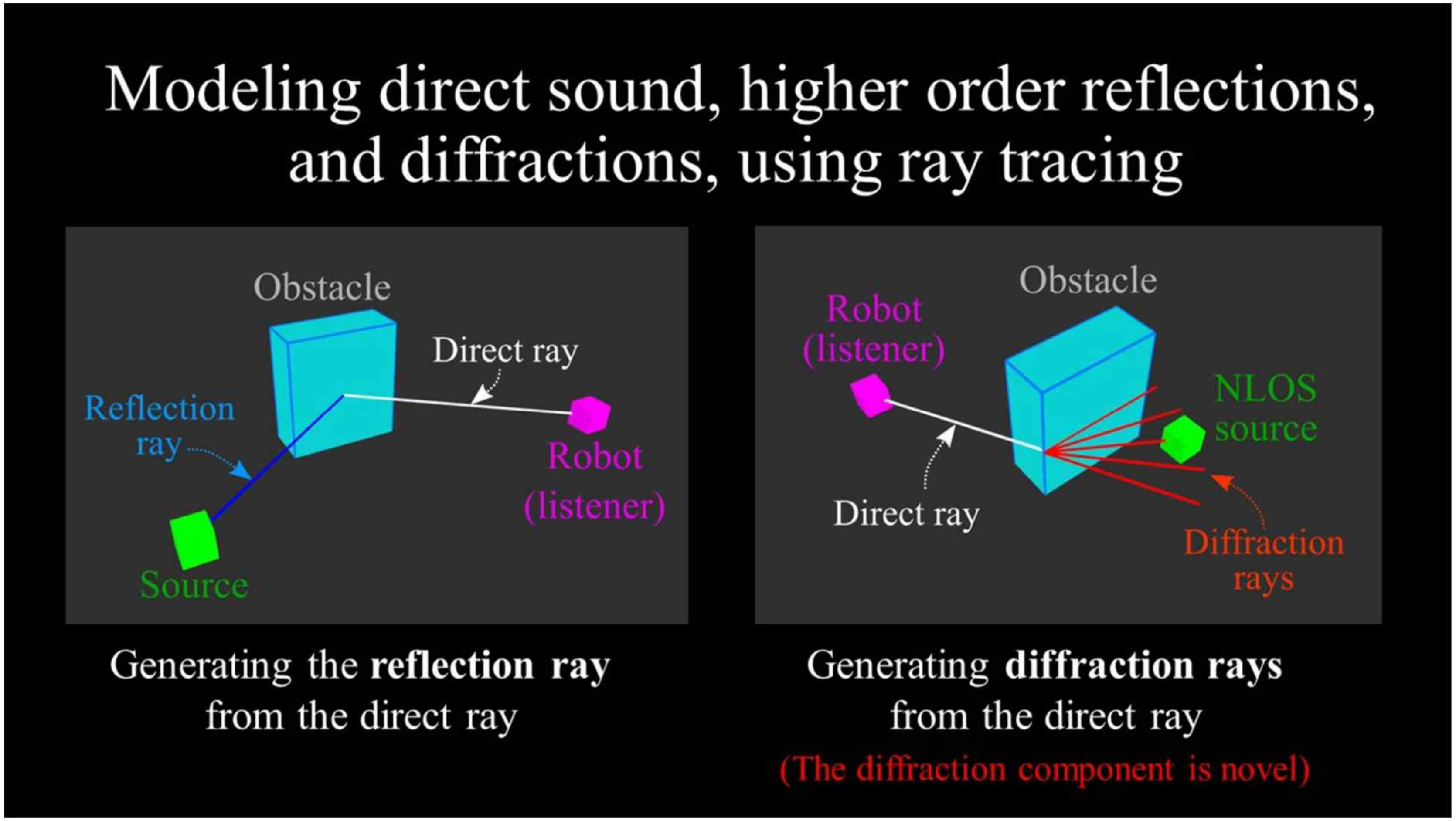
An example of generating reflection acoustic rays

Diffraction acoustic rays

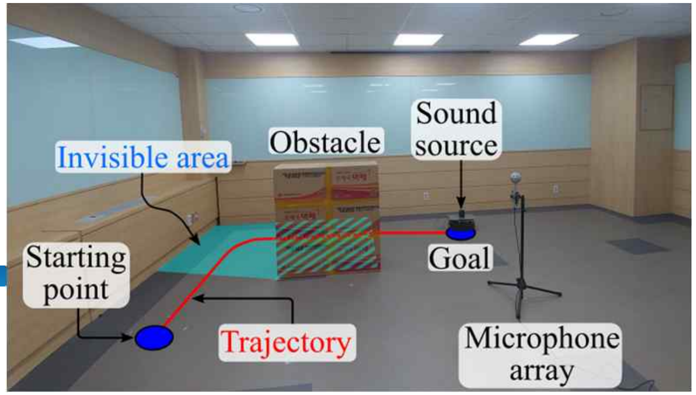
- The UTD model is based on **the principle of Fermat**: the ray follows the shortest path from the source to the listener.
- If the input is the path from the listener to the point on the edge and the source position is unknown, the set of possible shortest paths should be a surface of a cone (cone for UTD)
  → **Possible paths ≈ Diffraction rays**



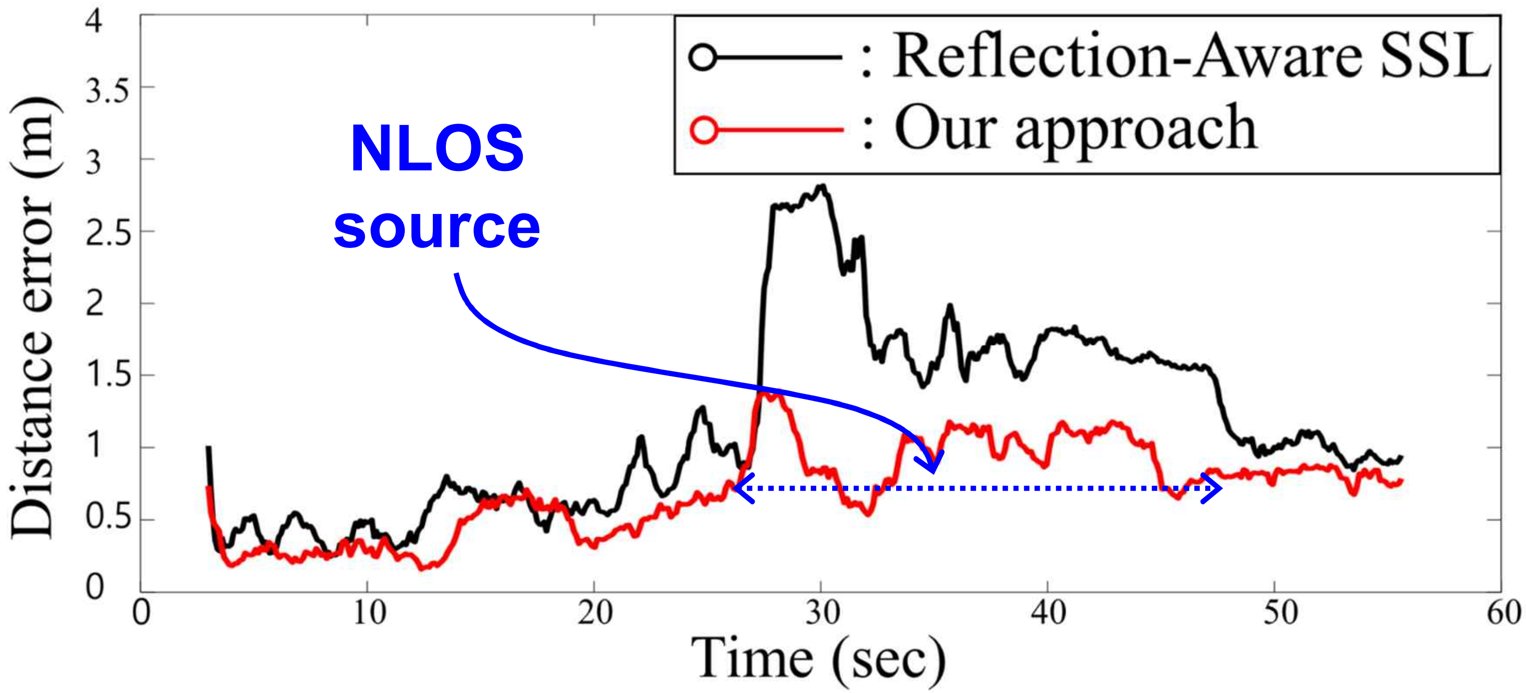
Working Video



Results | *A NLOS moving source scene around an obstacle*

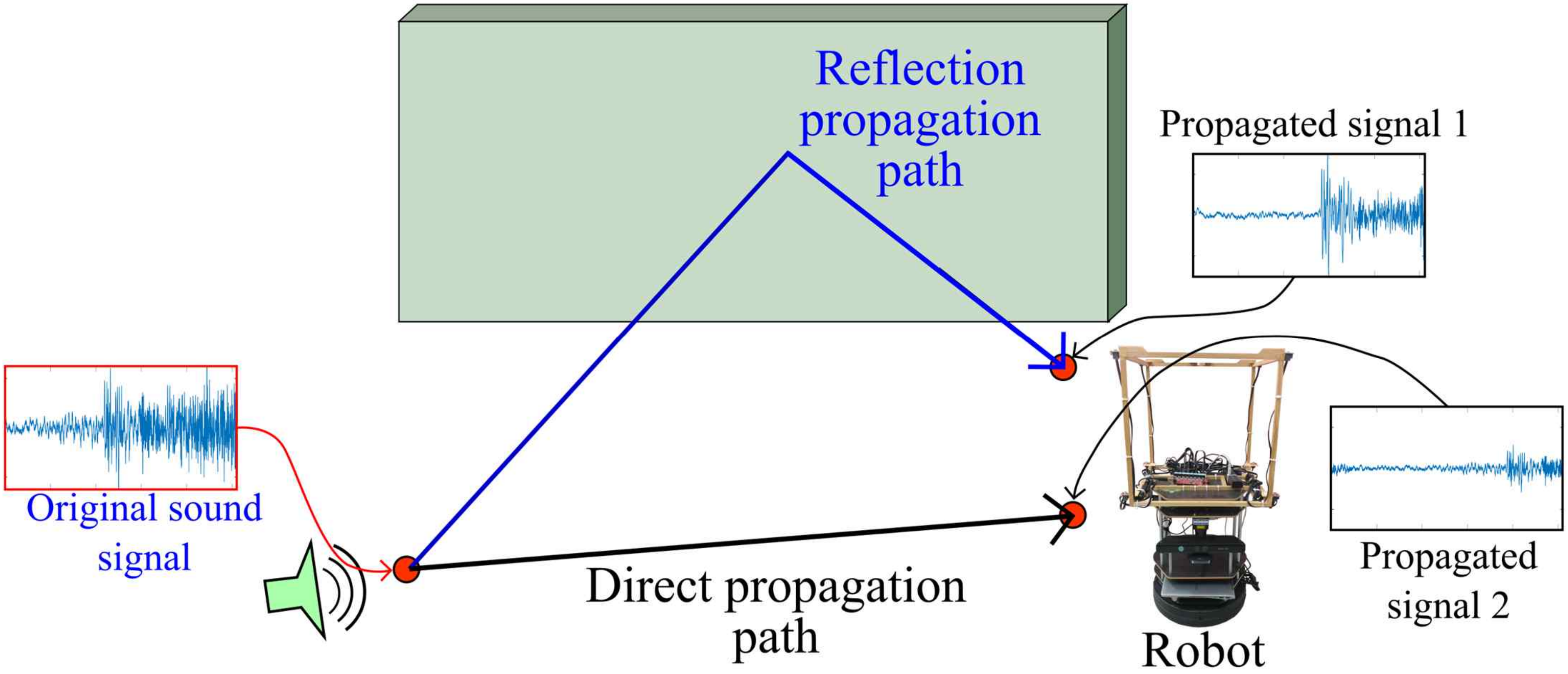


- **The distance errors between a ground truth and estimated positions**
  - The average distance errors of the RA-SSL and our method are 1.15m and 0.7m (during a NLOS source, 1.83m and 0.95m)



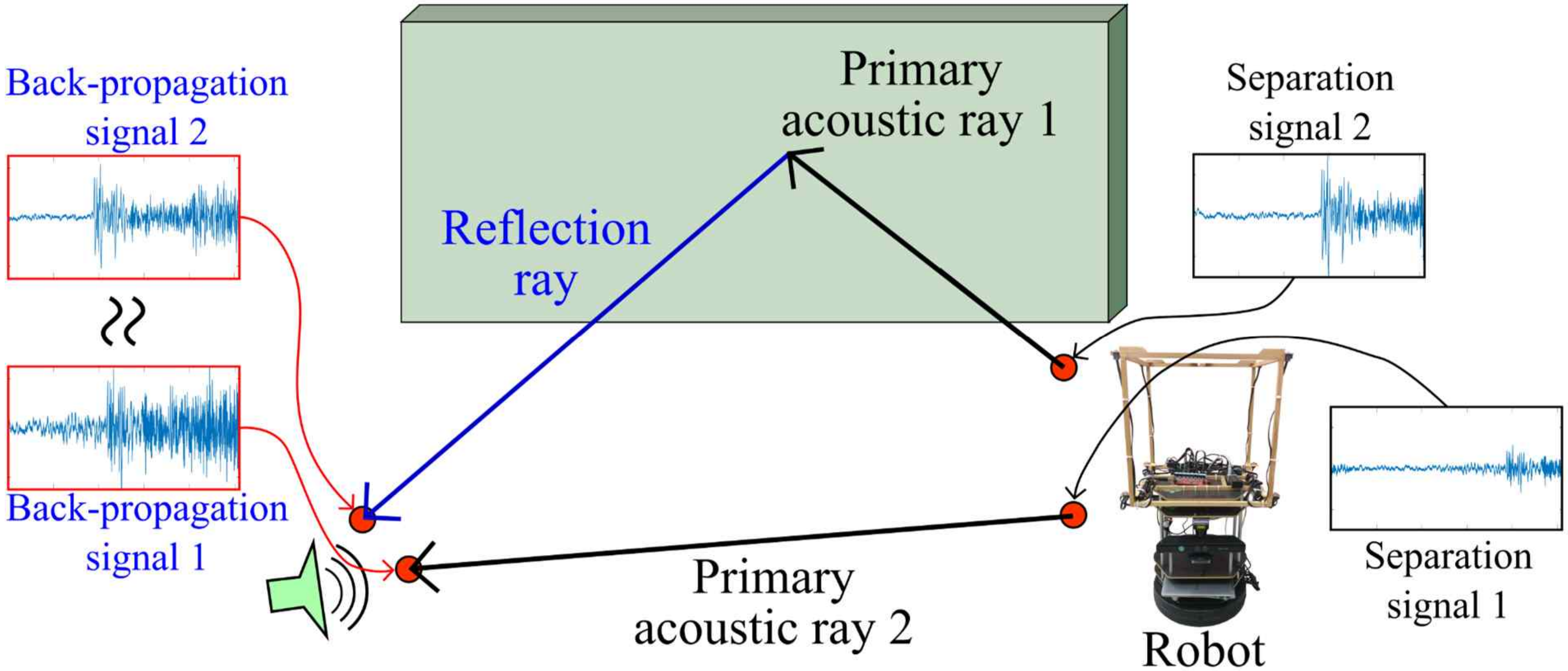
Considering Signals of Acoustic Paths

- The sound signal propagates along the propagation paths and the characteristics of signals are changed



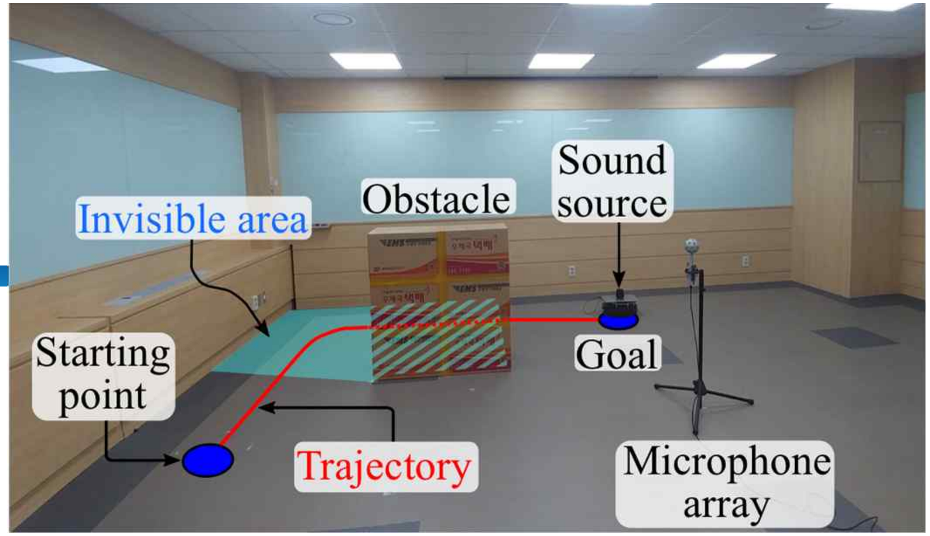
Utilizing Back-Propagation Signals

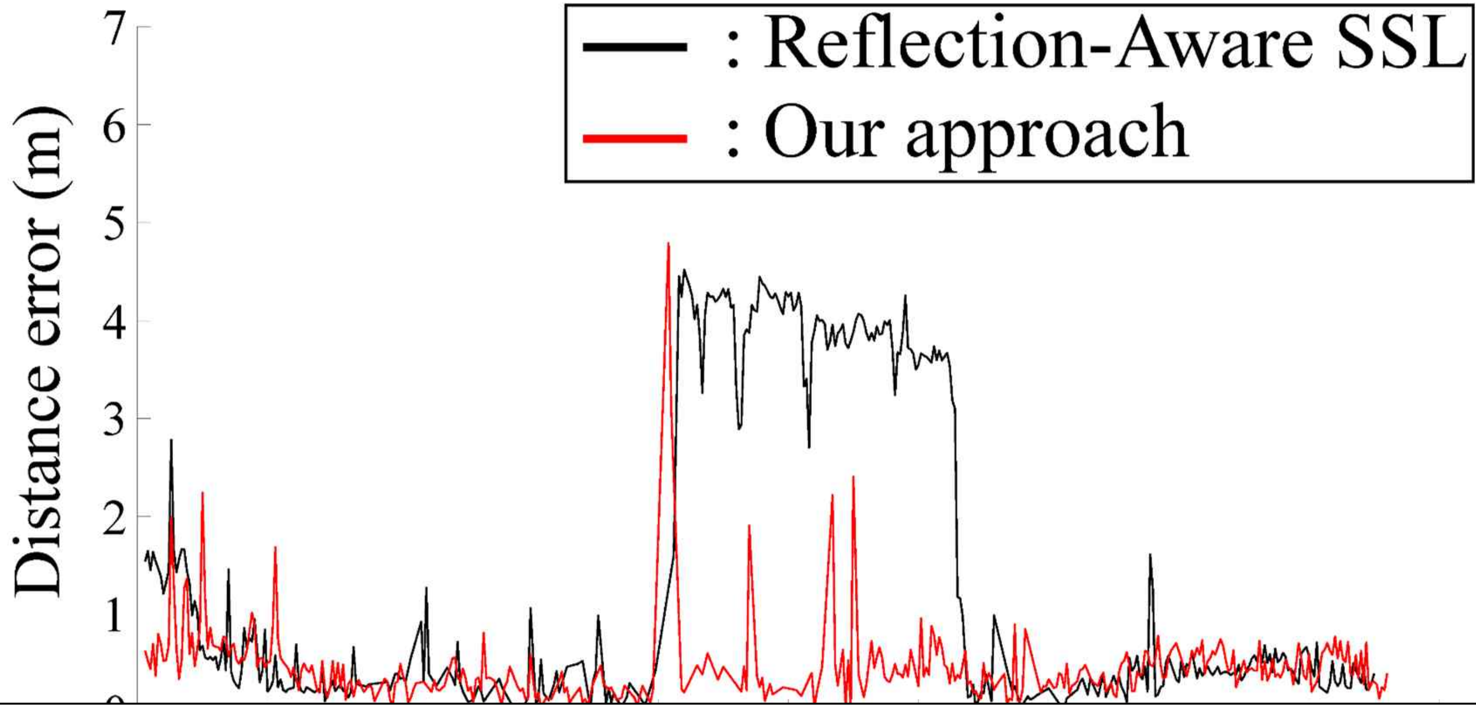
- Separate the sound signal came from a specific direction through the sound propagation path → Separation signals
- Propagate the separation signals based on acoustic ray paths → Back-propagation signals



Early Results

- The source moves along the red trajectory with a clapping sound
- The environment contains an obstacle





The avg. distance errors are 1.49m (RA-SSL) and 0.46m (Ours), 220% improvement

Conclusions

- **Discussed ray tracing based SSL for supporting reflection and diffraction**

- **Source codes are available**

- **Many future directions**
  - **Estimating geometry and materials**
  - **Jointly considering visual cues**
  - **Handling noise environments and multiple speakers**

KAIST