Map-less Navigation for Blind People

Masaki Kuribayashi

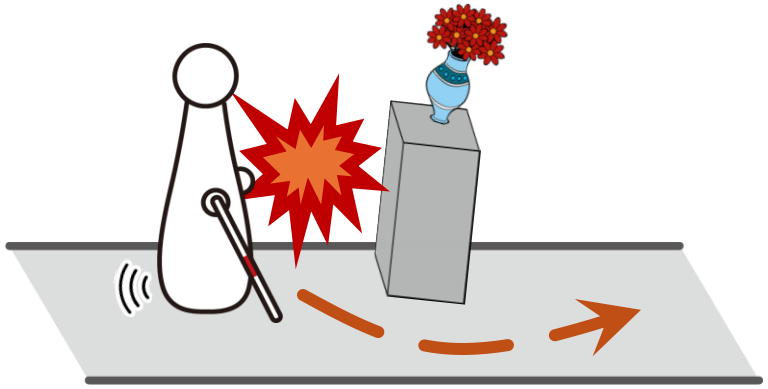
Morishima Lab., Waseda University



Difficulties Blind People Face when Navigating Indoor

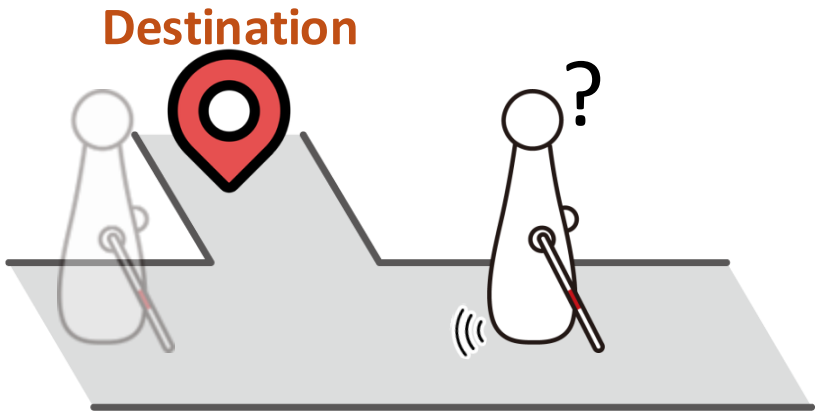
Mobility

- Cane can still miss obstacles
- Risk of injury and breaking objects



Orientation

- Wayfinding is cognitively demanding even in familiar place
- Can miss intersection get lost



Spatial Learning

- Need for constructing mental map and enjoying
- Also important for wayfinding



Navigation Systems for Blind People

Smartphones

Supports **orientation** and **spatial leaning**

e.g., App for in/outdoor turn-by-turn guidance[1] and LLM-powered window-shopping[2]

Robots

Additionally supports **mobility**

e.g., Cane-like, quadruped and wheeled robots

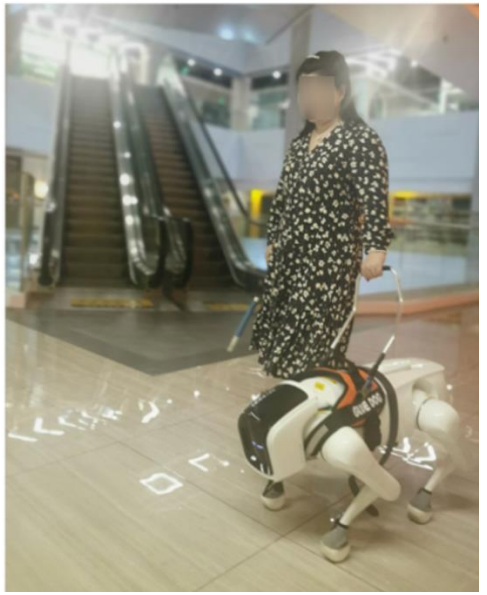
BlindSquare[1]



ChitChatGuide[2]



Quadruped[3]



Wheeled[4]



[1] BlindSquare, Retrived in July 16 https://www.blindsquare.com/

[2] Kaniwa et al., "ChitChatGuide: Conversational Interaction Using Large Language Models for Assisting People with Visual Impairments to Explore a Shopping Mall"

[5] Cai et al, "Navigating real-world challenges: A quadruped robot guiding system for visually impaired people in diverse environments," CHI2024

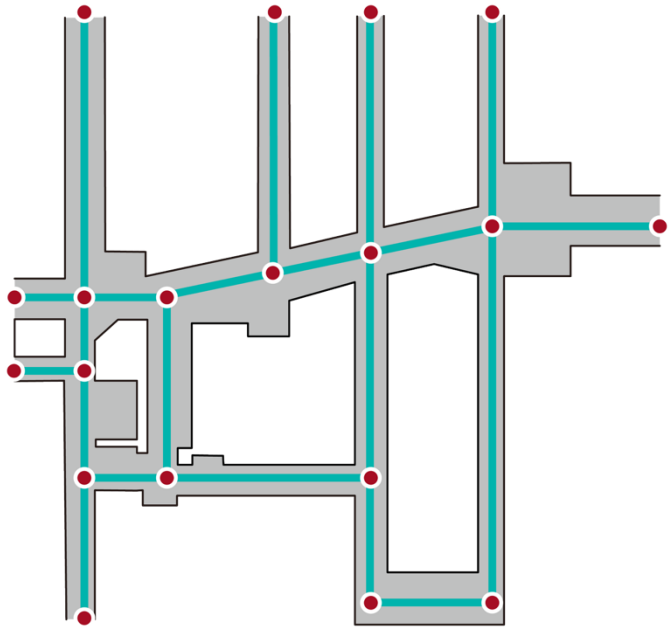
[6] Zhang et al, ""I am the follower, also the boss": Exploring Different Levels of Autonomy and Machine Forms of Guiding Robots for the Visually Impaired", CHI2023

Systems with Prebuilt Maps Cannot be Used Anywhere

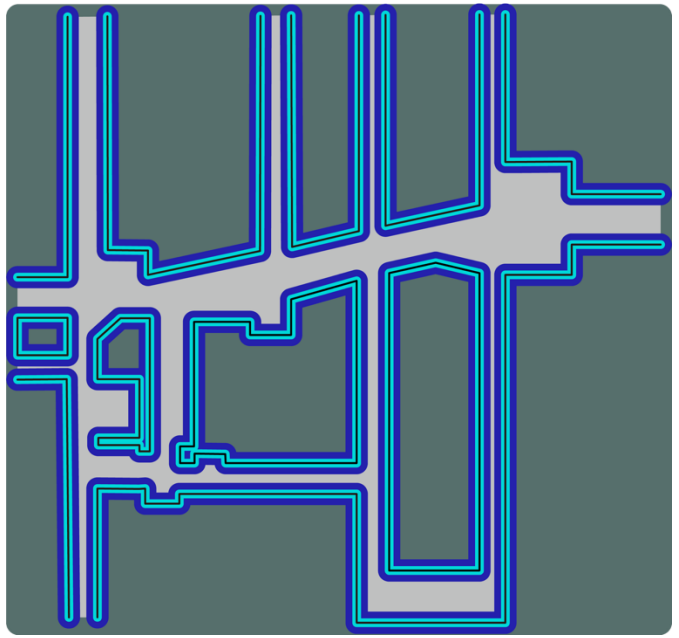
Infrastructure



Static Route Map



LiDAR Map



Map-less Navigation System

To Realize a System that **Blind**

People Can Use Anywhere

Develop a System **Without**

Pre-installed Maps or Infrastructure

Today's Presentation

Robotic System





VLN Agent



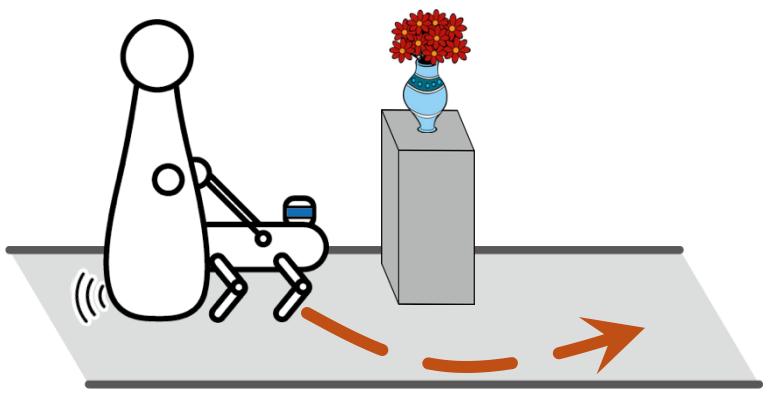
Smartphone System



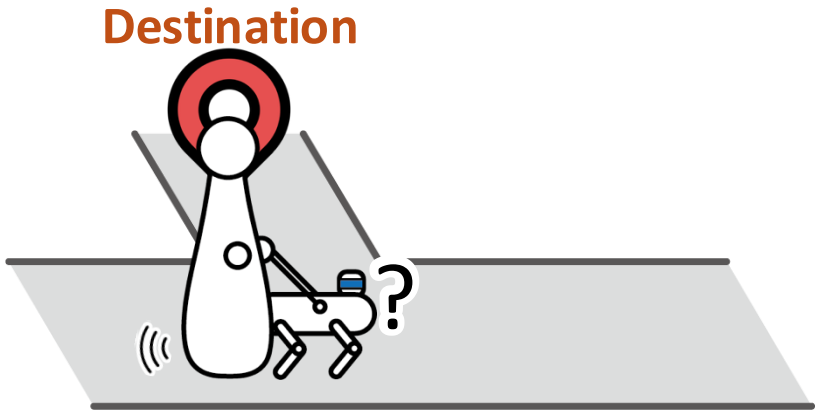


Challenge: Map-less Means Robot Doesn't Know Where To Go

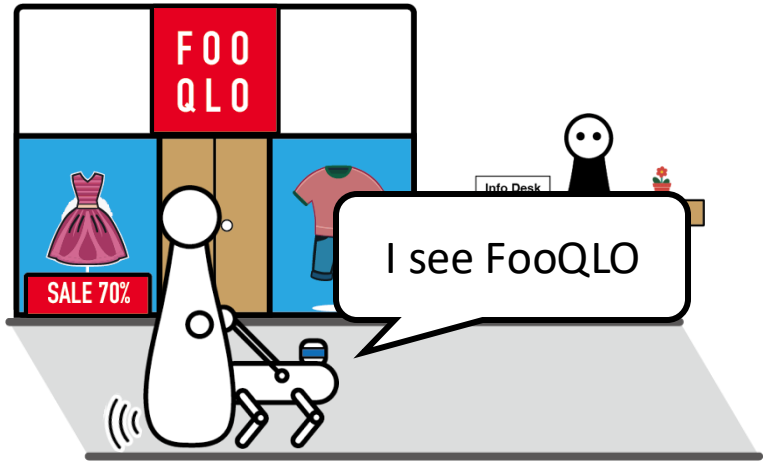
Can avoid obstacles
if the destination is set



Cannot determine
which way to go



Need to recognize
environmental information

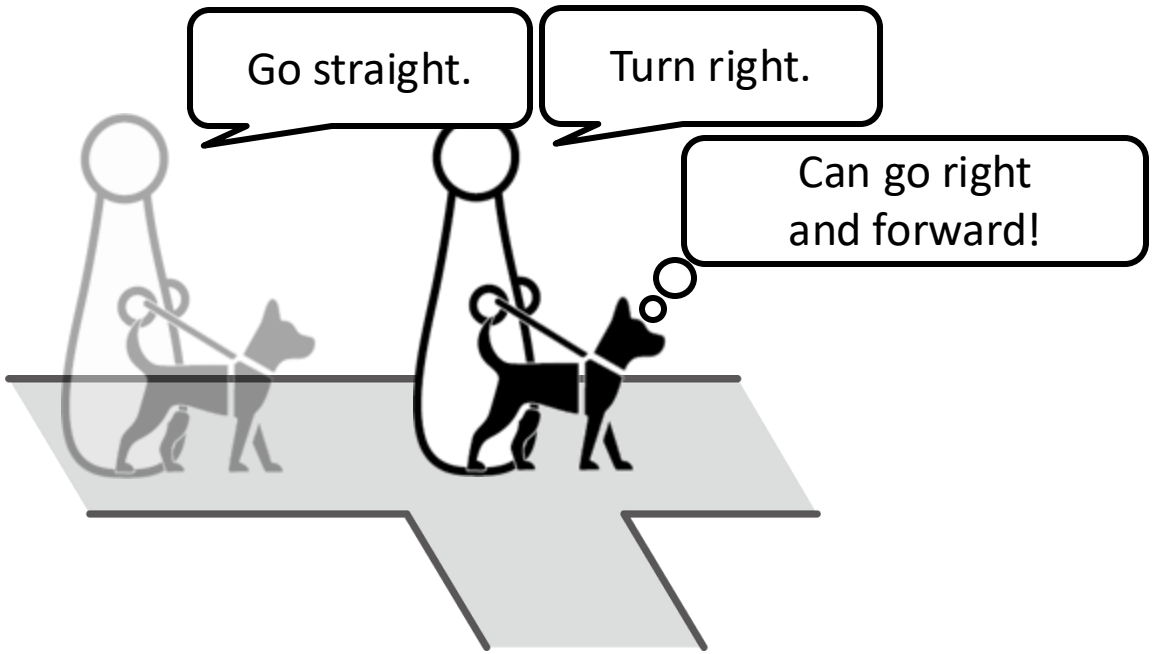


Key Idea: Imitate Interactions of Guide Dogs and Go Beyond

Adopt collaborative interaction[1] and enhance user decision making with AI

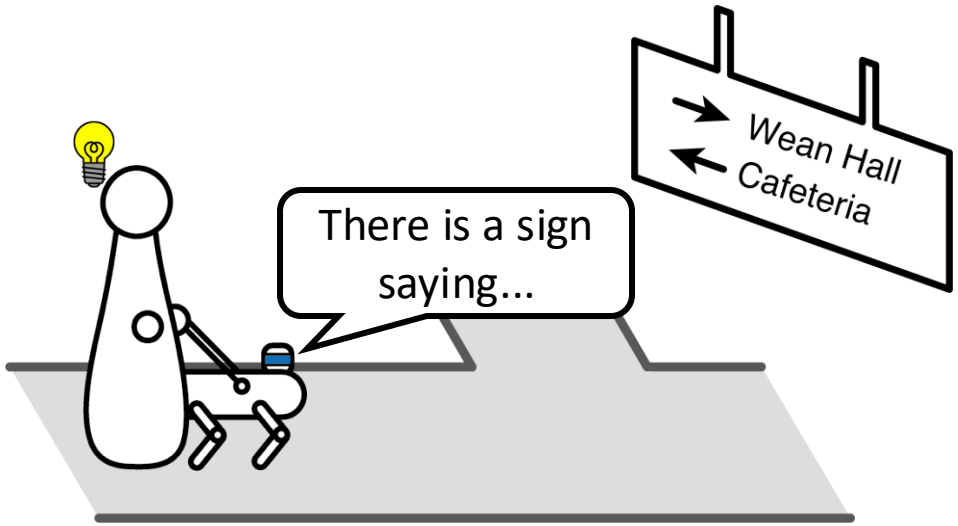
Collaborative Navigation

User determine direction and guide dogs guide user safely



Utilization of Sensors and AI

Use sensors and recognition technology to convey information more than guide dogs

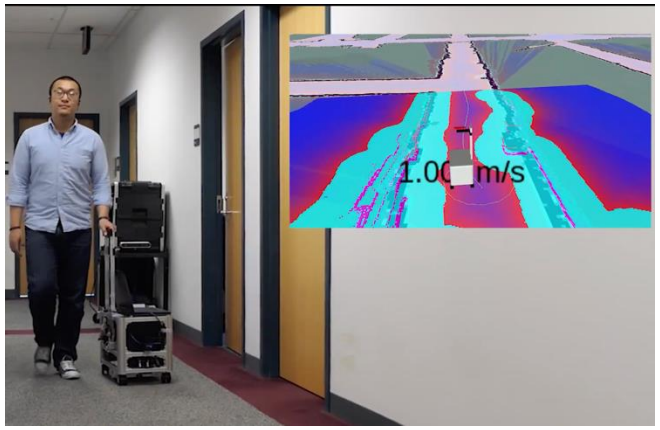


[1] Hwang et al., "Towards Robotic Companions: Understanding Handler-Guide Dog Interactions for Informed Guide Dog Robot Design," CHI2024

RQ: How Can We Design Map-less System?

Existing: Users Follow System Instruction[1,2]

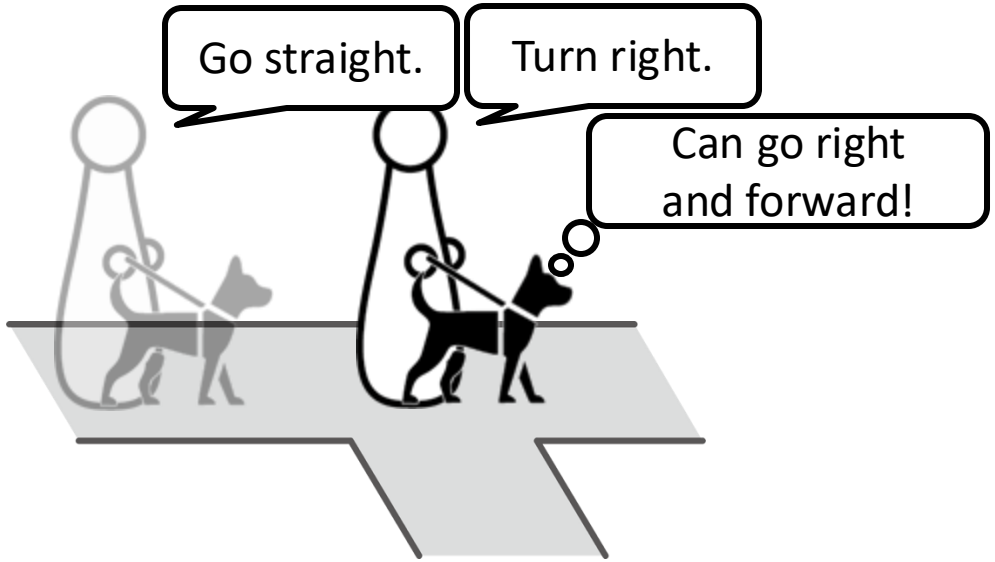
Prior focus was mainly on how much the system can assist using newest technology





Ours: Users Engage in Decision-Making

How much can blind people complement system capability? What information do blind people need for decision-making?



Focus on Navigation and Exploration task, and adopt participatory design approach[3]

[1] Guerreiro et al., “CaBot: Designing and Evaluating an Autonomous Navigation Robot for Blind People,” ASSETS2019
[2] Kaniwa et al., “ChitChatGuide: Conversational Interaction Using Large Language Models for Assisting People with Visual Impairments to Explore a Shopping Mall,” MobileHCI2024
[3] Delgado et al. “The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice,” EAAMO2023

Research Projects

Platform: AI-Suitcase

Robot Advantage

- Can take over mobility
- **Users can focus on environment perception and decision-making**

Suitcase Pro/Con[1]

- Seamless and can blend into environment
- Cannot go up the stairs



[1] Takagi et al, "Field Trials of Autonomous Navigation Robot for Visually Impaired People," CHI2025



PathFinder: Designing a Map-less Navigation System for Blind People in Unfamiliar Buildings

Masaki Kuribayashi[1], Tatsuya Ishihara[2], Daisuke Sato[3], Jayakorn Vonkulhbisal[2],

Karnik Ram[3], Seita Kayukawa[1], Hironobu Takagi[2], Shigeo Morishima[4], Chieko Asakawa[3,5]

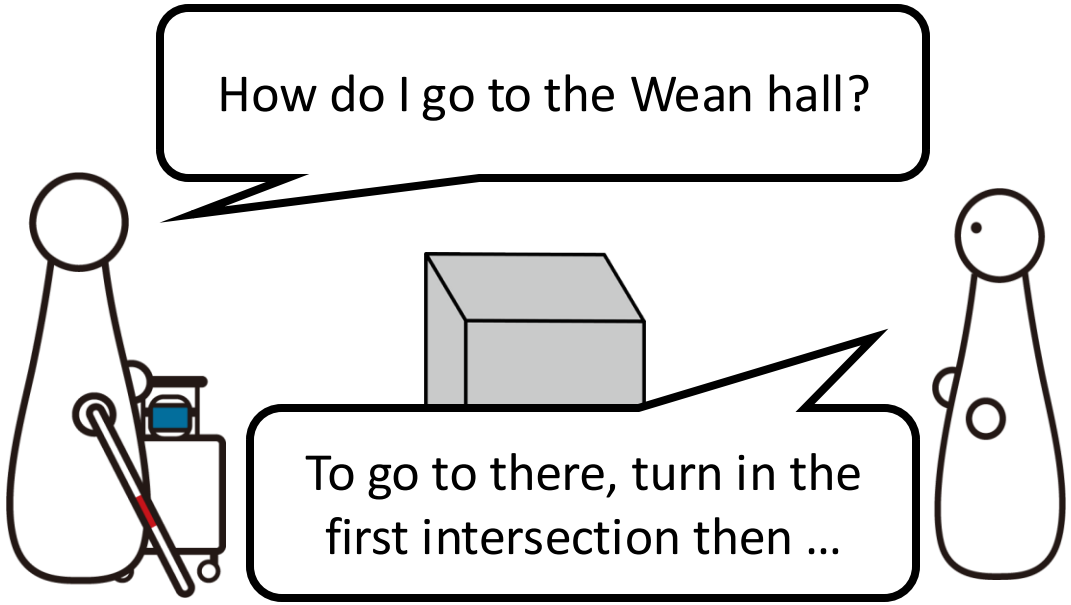
[1]Waseda University, [2]IBM Research - Japan, [3]Carnegie Mellon University,

[4]Waseda Research Institute of Science and Engineering, [5]IBM Research

CHI23

Situation to Use PathFinder

Unfamiliar building where a blind person have a destination and acquired route description from sighted people[1] (*e.g.*, entrance or passersby)



[1] Engel et al., "Travelling More Independently: A Requirements Analysis for Accessible Journeys to Unknown Buildings for People with Visual Impairments"

Design Procedure of PathFinder

RQ: What information and function need is there for PathFinder, and how does it compare with existing with-map system?

Step 1: Investigation of Route Description and Useful information

Interview 10 sighted people of route description and 5 blind participants of useful information in unfamiliar building given the route description.

Step 2: Prototype PathFinder

Prototyped PathFinder by implementing intersection detection and sign recognition functionalities.

Step 3: Improve Interface and Functionality

5 blind participants used the initial prototype of PathFinder and gathered comments for improvements.

Step 4: Conduct Main Study to Compare with With-Map System

Compared with with-map system and investigated performance and pros/cons of PathFinder

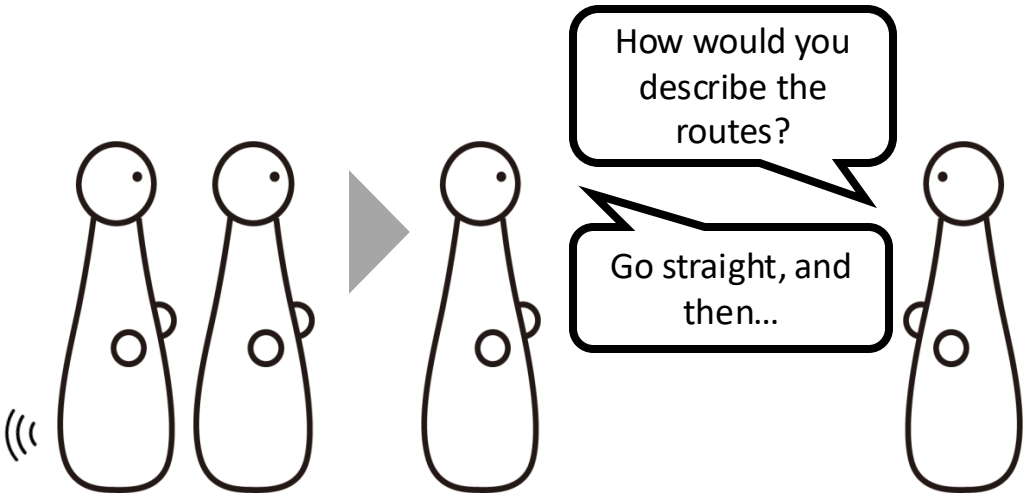
Investigation of Route Description and Useful Information

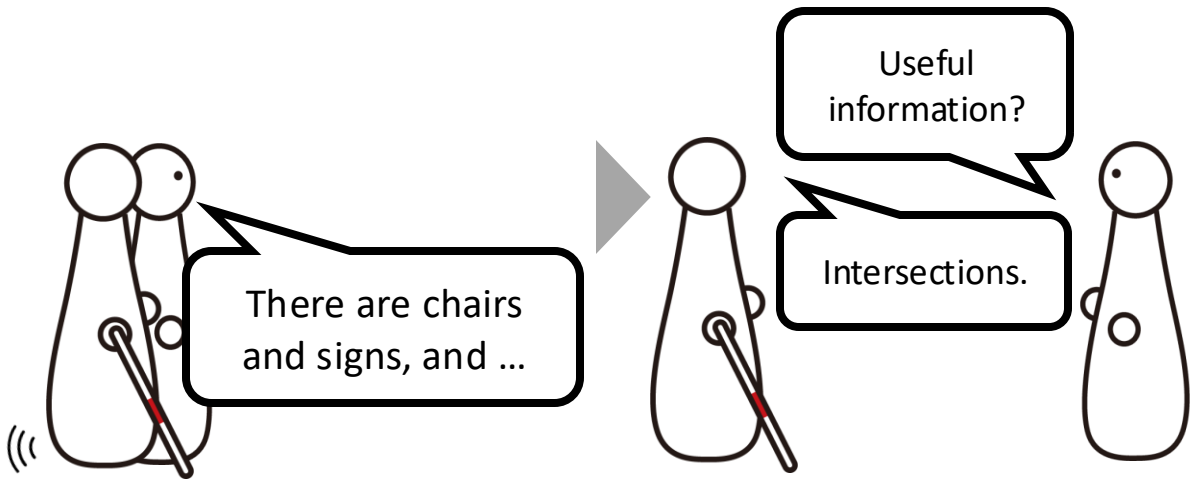
Route Description Investigation with 10 Sighted Participants

1. Navigate through two routes
2. Describe the routes for blind people

Useful Information Investigation with 5 Blind Participants

1. Listen to the obtained route description
2. Navigate the route as the experimenter points out visible features
3. Rate how useful each information was

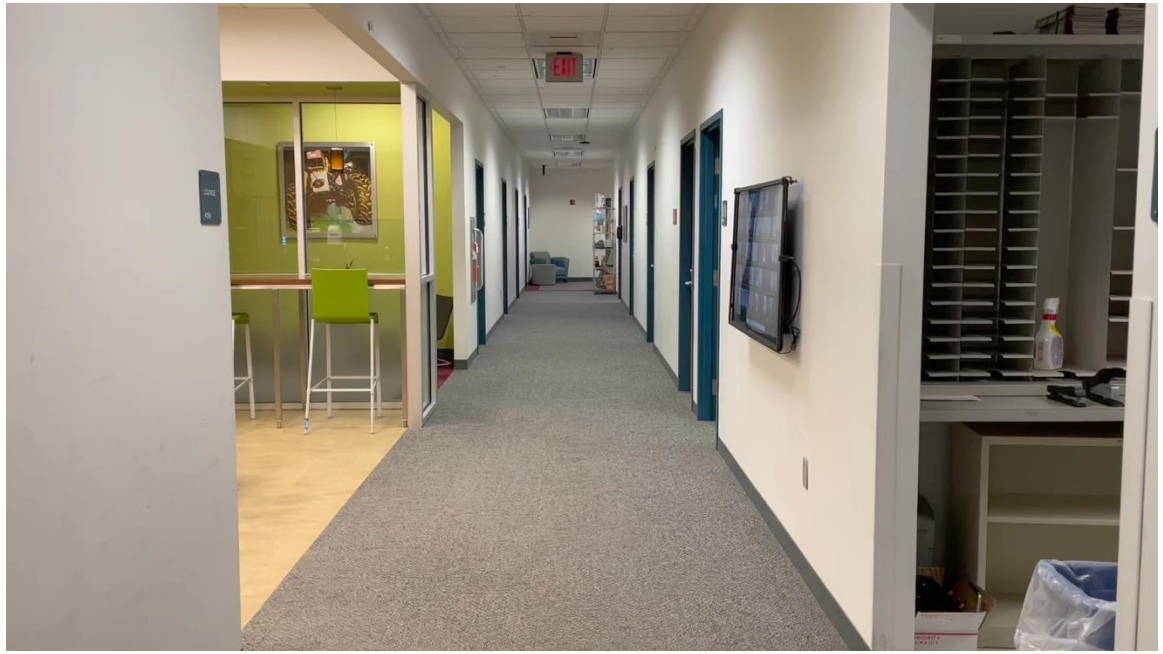




Route Description Investigation With 10 Sighted Passersby

Intersections and distance to walk were the most described

Route 1 (46m)

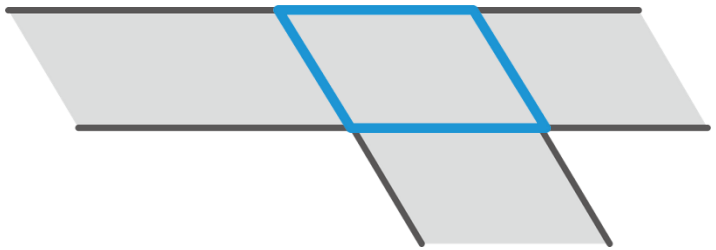


Route 2 (166m)



Three Most Useful Information For Blind People in Unfamiliar Buildings

Intersections
Help users determine when to turn



Directional Signs
Help users to be confident that they are going the correct way



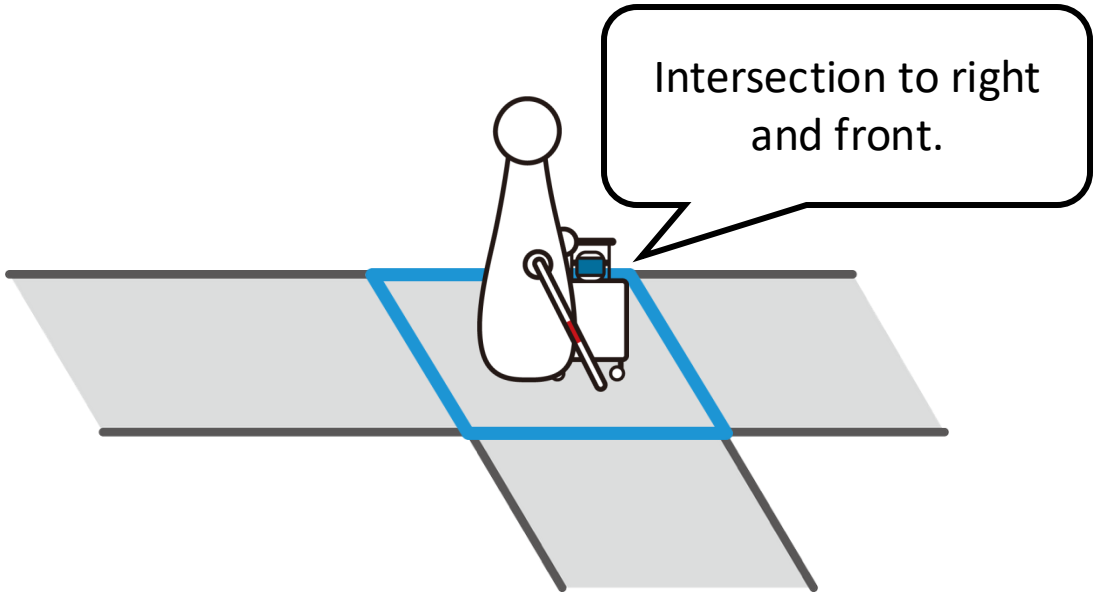
Textual Signs
Helps users to know that they have reached the destination



System Design of PathFinder

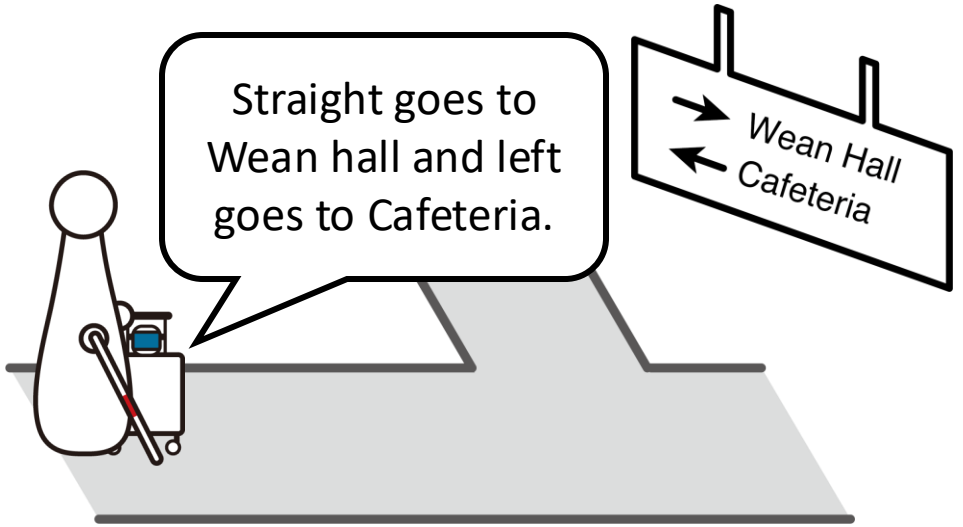
Intersection Detection

Recognize location and shape of intersections



Sign Recognition

Read out textual and directional signs to help users to be more confident



Interface of PathFinder

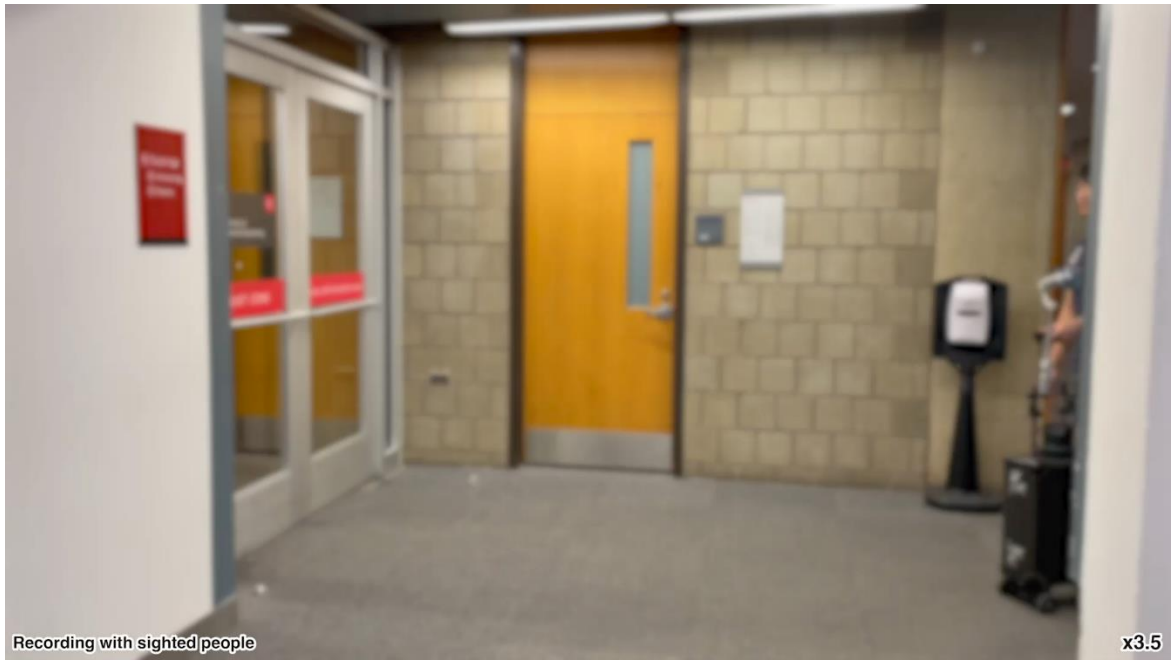
Intersection Detection

User can press front button to go to next intersection and left/right button to specify direction to go



Sign Recognition

User can press back button to instruct the robot to perform sign recognition



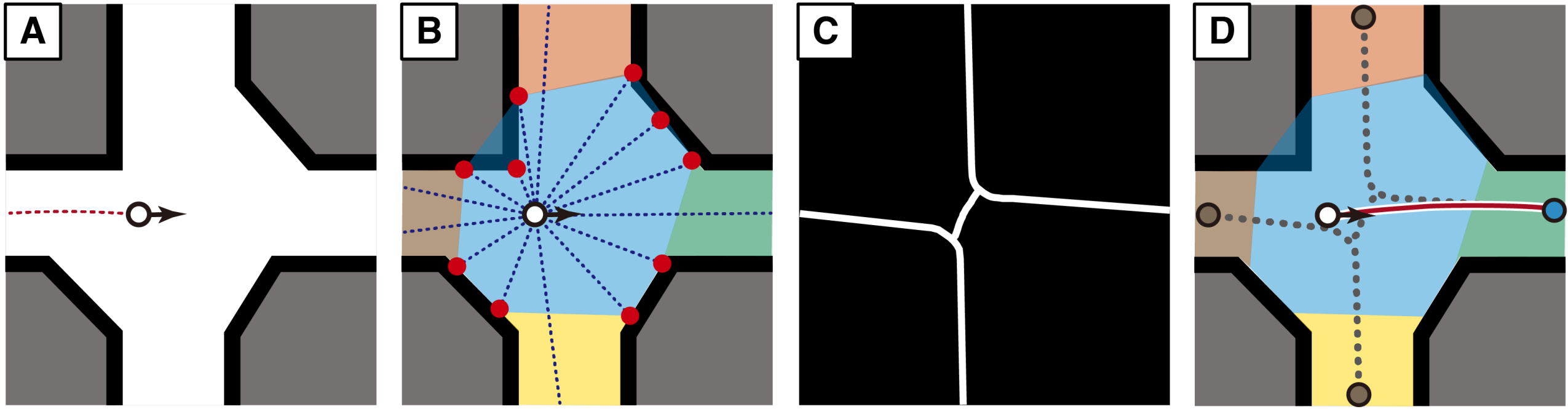
Applying Underground Intersection Detection[1] to PathFinder

1. Get 20 m × 20 m LiDAR Map from SLAM

2. Cast rays and calculate convex hull

3. Skeletonize LiDAR Map

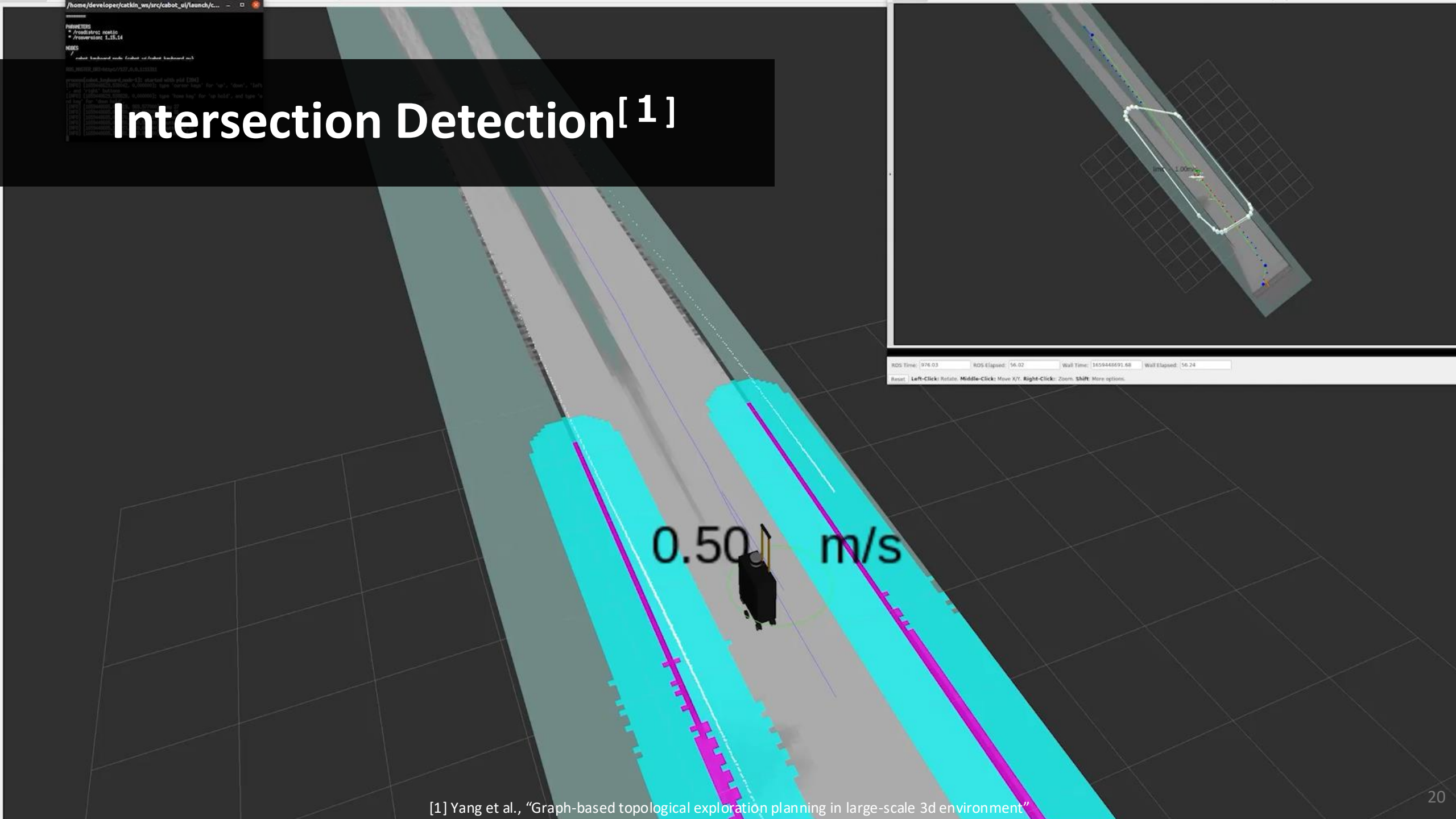
4. Set mid-point of the path as destination



*36 Rays are casted and only ones within 8m are sampled

[1] Yang et al., "Graph-based topological exploration planning in large-scale 3d environment"

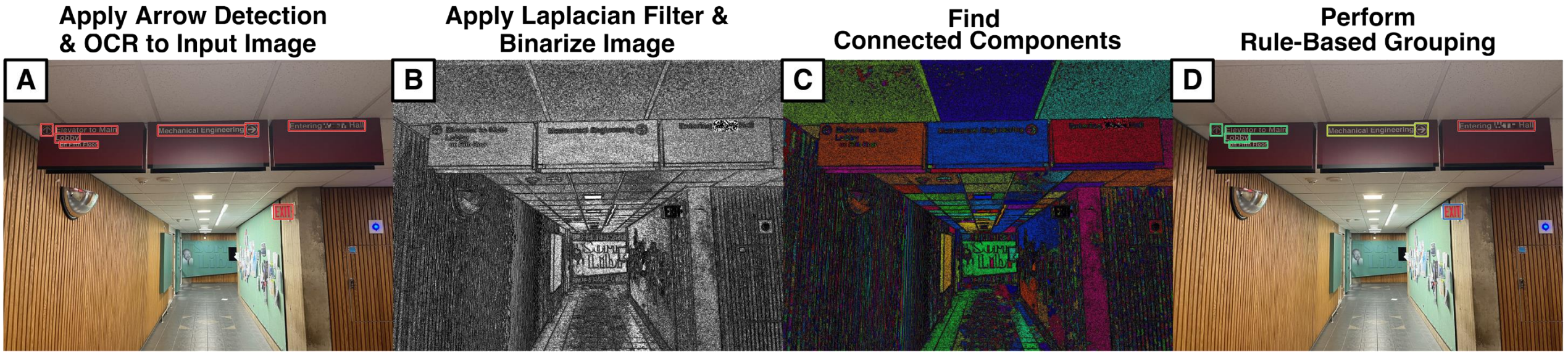
Intersection Detection[1]



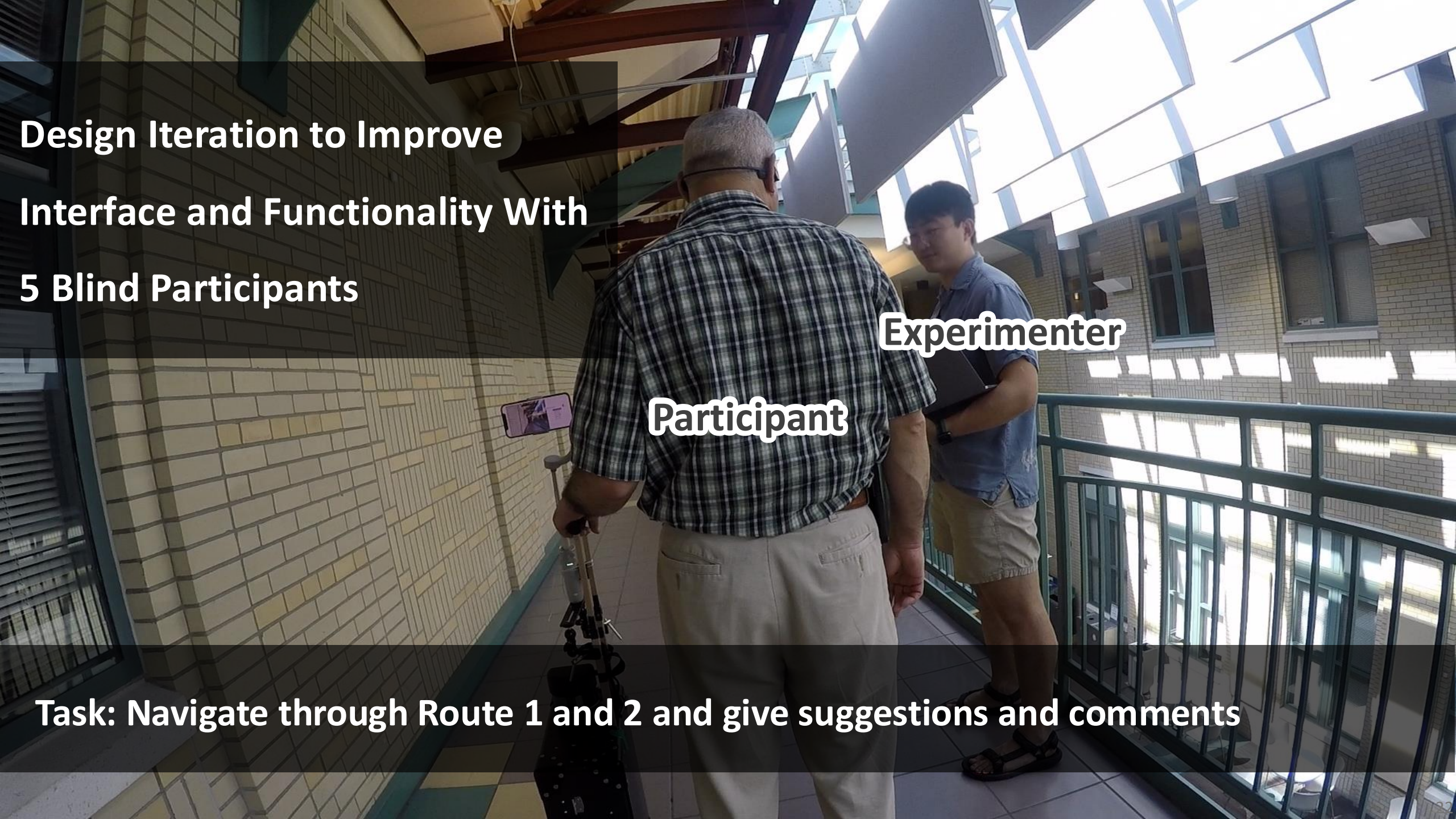
[1] Yang et al., “Graph-based topological exploration planning in large-scale 3d environment”

Directional and Textual Sign Recognition

Result Feedback: **Front leads to Elevator to main lobby**, **right to mechanical engineering**.
Also, one sign saying **Entering Wean Hall**



Text: EasyOCR
Arrow: YOLOv3



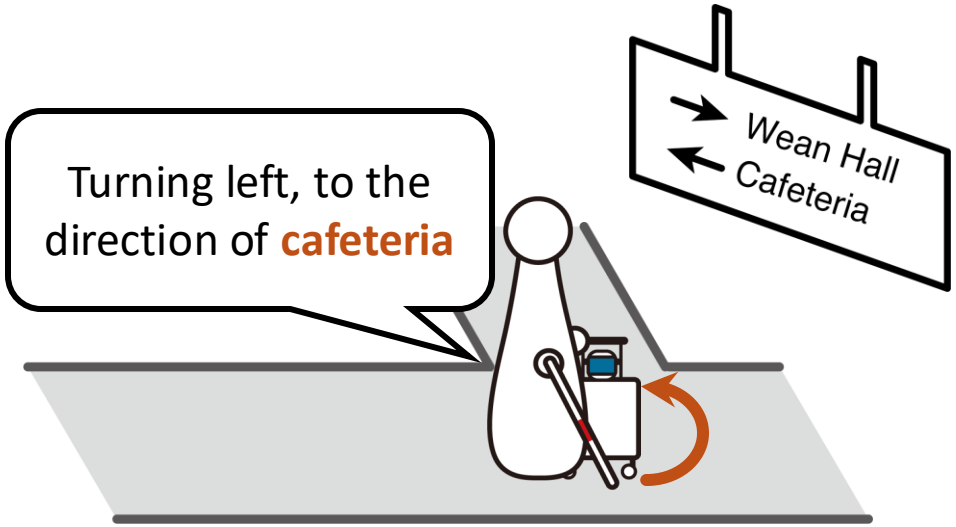
Design Iteration to Improve Interface and Functionality With 5 Blind Participants

Task: Navigate through Route 1 and 2 and give suggestions and comments

Examples of Improvement of Interface and Functionality

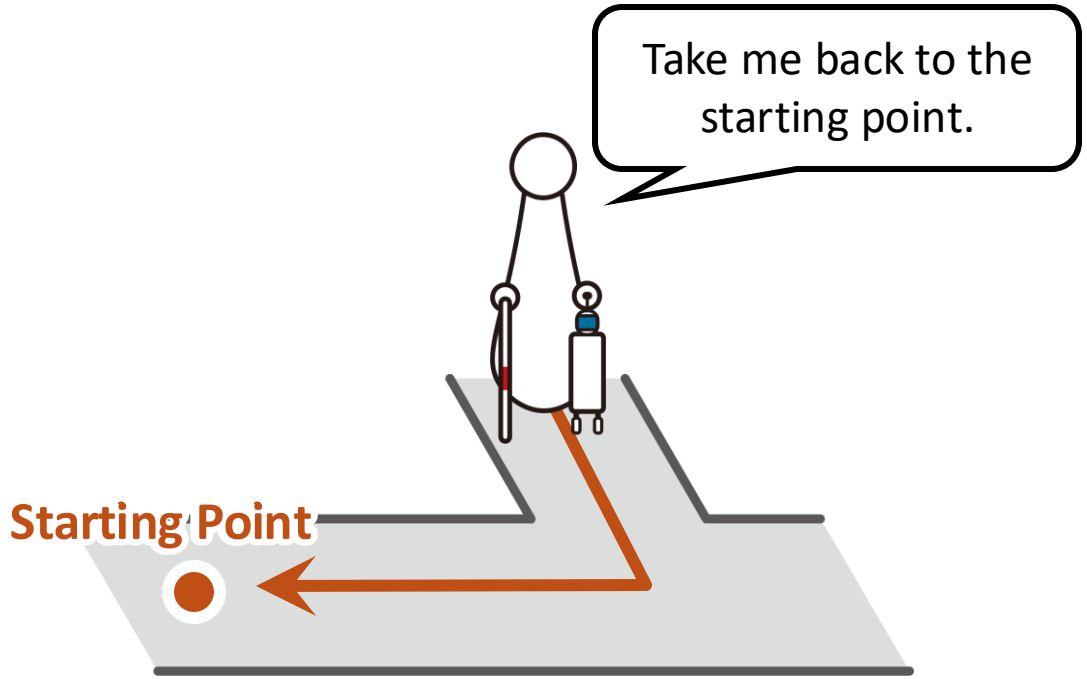
Refer to Signs When Turing

One participant wanted the system to refer to signs while turning when they are available



Additional "Take-Me-Back" Functionality

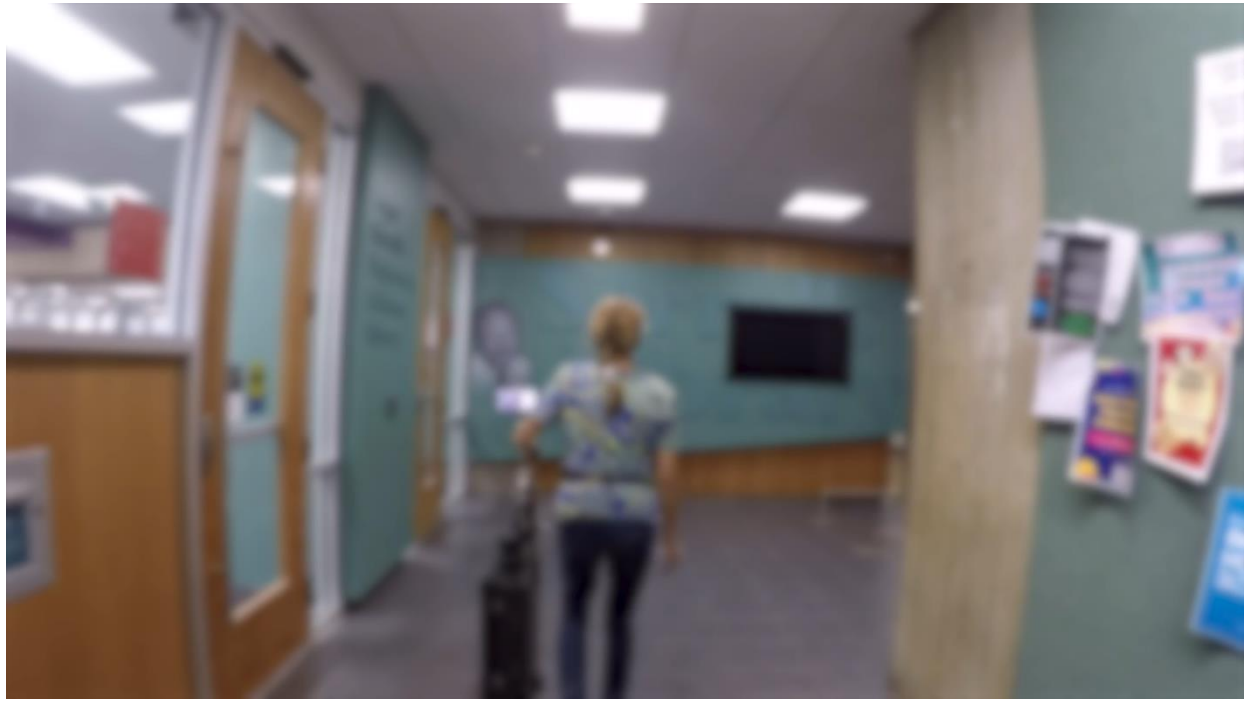
One participant requested for a functionality to go back to where they started



User Study with 7 Blind Participants

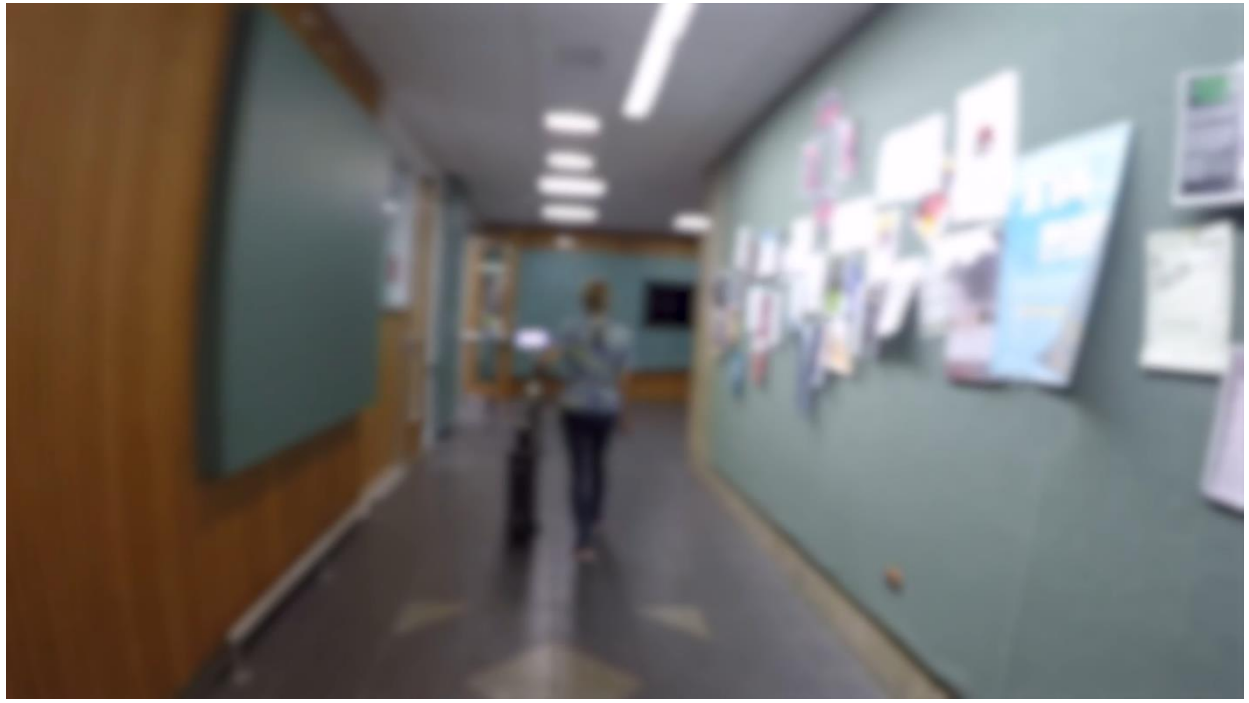
PathFinder

A navigation robot **without prebuilt map**



CaBot[1]

A navigation robot **with prebuilt map**

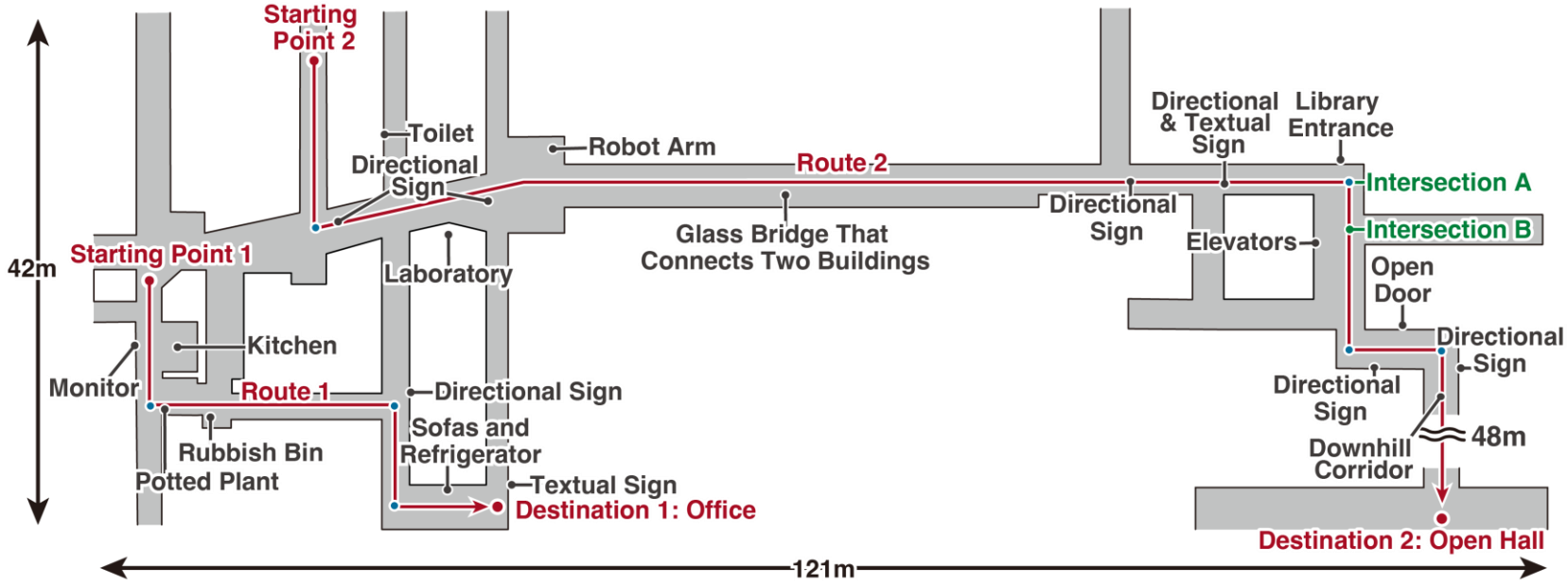


Task: Navigate through the Routes based on route description and then go back

[1] Guerreiro et al. “CaBot: Designing and Evaluating an Autonomous Navigation Robot for Blind People,” ASSETS2019

Experiment Detail

- Participants were allowed to ask routes if needed (*e.g.*, forgot the route)
  - Participants asked routes 1.29±1.70 times for Route 1 and 3.43±1.80 times for Route 2
- Experimenter intervened if participants were lost, and recovered back to the route



Route 1 description example

"Go straight, turn on the first left. Then go straight down the hallway until you meet the right corner. Basically, that's an end. Turn right there. Go straight. When you see a corner, turn left. If you walk straight for around 12 steps, you will see the destination with a specific room number in front of you."

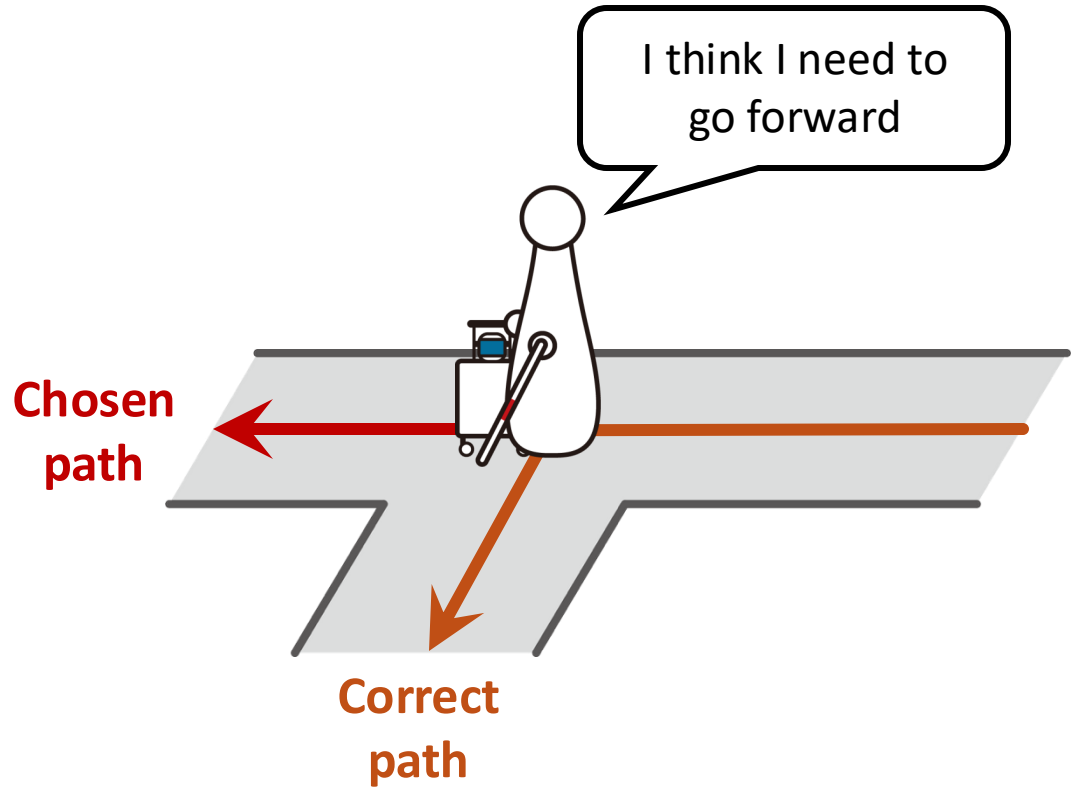
Route 2 description example

"Go straight, turn on the first left. Then go straight down the hallway until you meet the end which leads to the right. Turn right there. Go to the next end which leads to left. Turn left there. Walk ten steps, turn right. Go all the way down the hallway. Once you reach an open space, you have reached the destination."

Task Success Rate of PathFinder was 71%

Failure: Experimenter intervened at least once because participants were lost

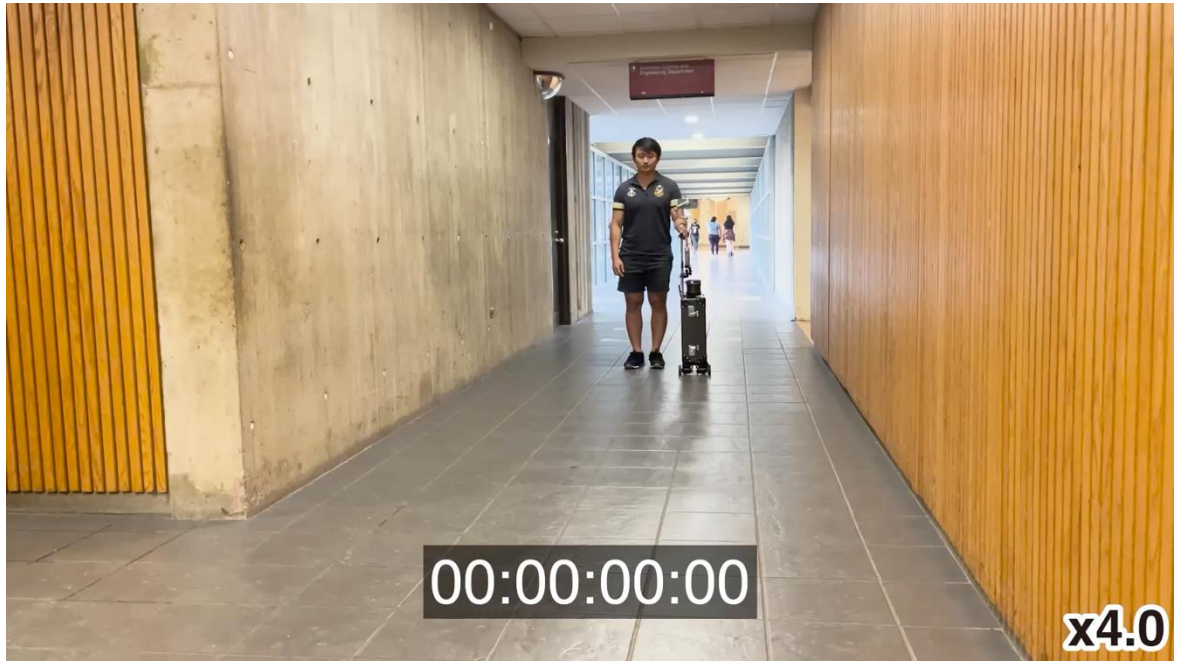
4 occurrence where users gave wrong instruction at an intersection



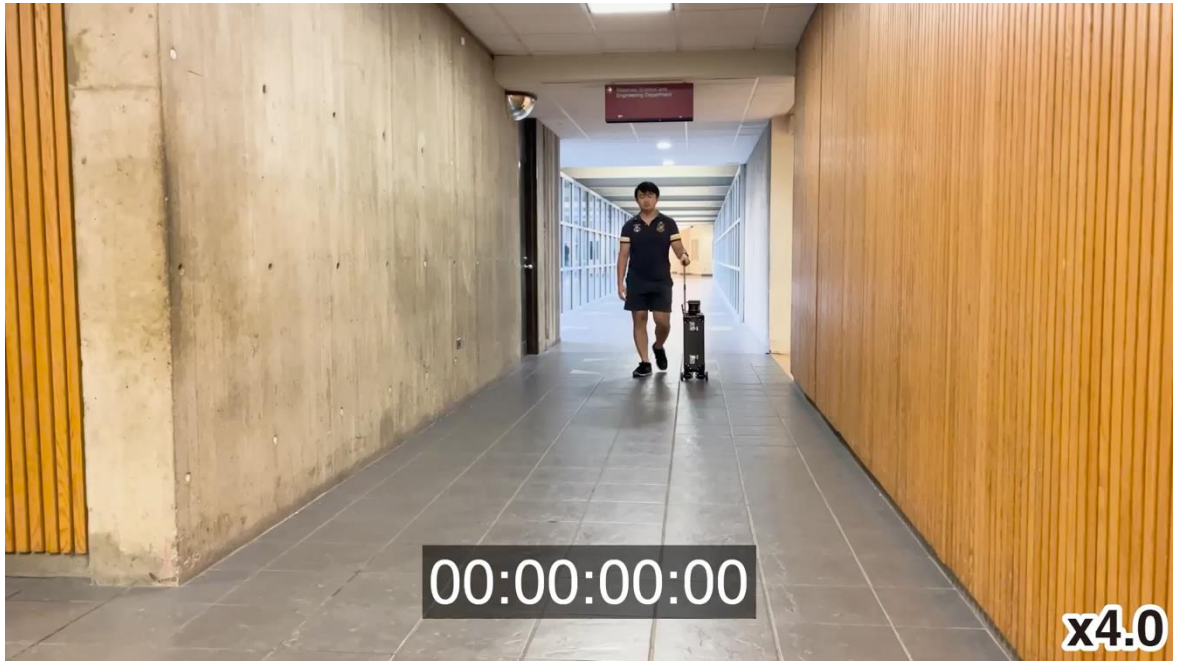
Task Completion Time Increased When Using PathFinder

Users needed to instruct PathFinder which way to go at every intersection

PathFinder



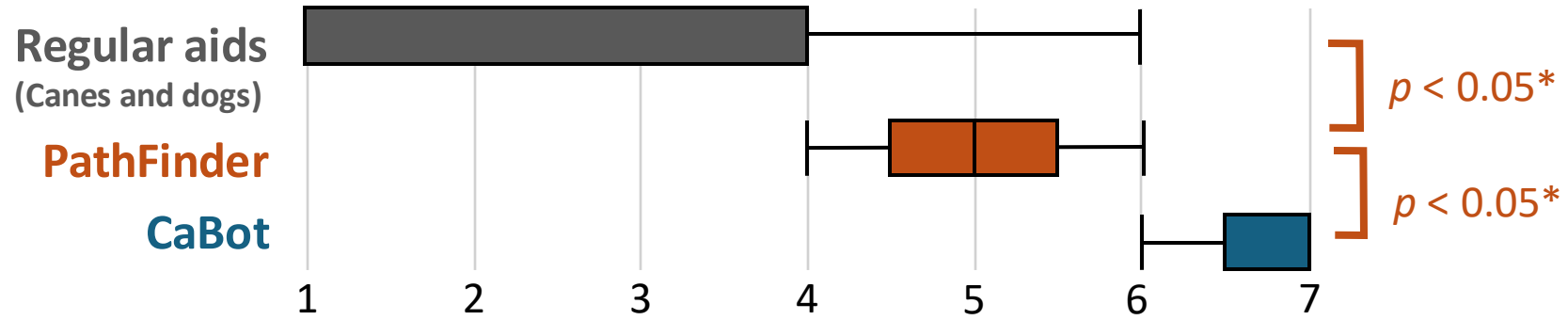
CaBot



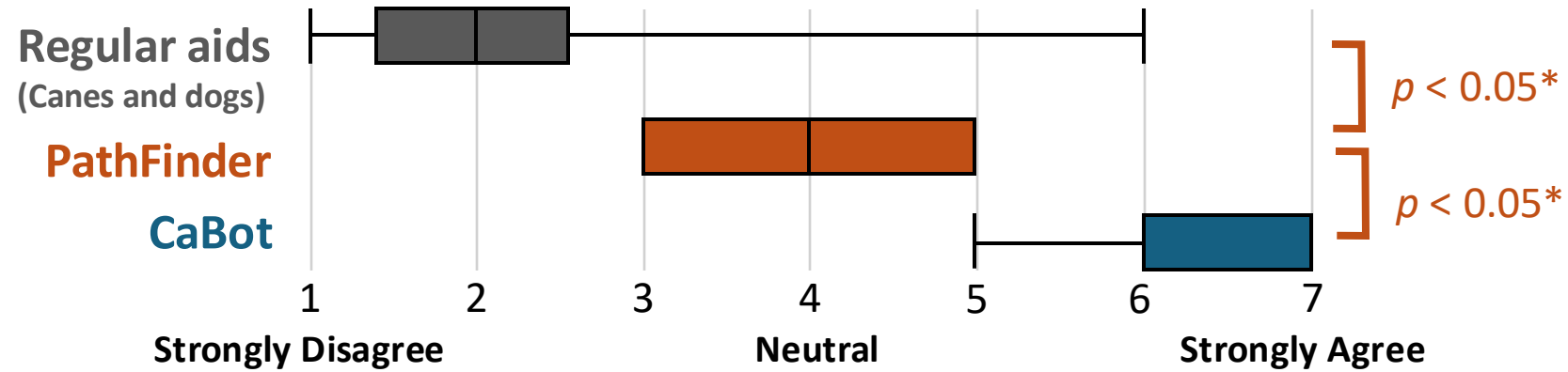
Confidence & Cognitive Load

PathFinder was rated in between regular aid and Cabot

More Confidence



Less Cognitive Load



*Used Wilcoxon Signed Rank test for statical analysis

All Participants Appreciated the Map-less Aspect of PathFinder



P07

"If maps are available, it is definitely the most useful. But maps are not available every where. However, ***even without map information, the robot is very useful because it reads out information around me and lets me find my way****."*

PathFinder: Designing a Map-less Navigation System for Blind People in Unfamiliar Buildings

- Blind people can determine their way in unfamiliar buildings based on route description by using navigation robot that integrates sign recognition and intersection detection
- While it took more time and effort to use PathFinder, they were still willing to use the system because it has potential to be used in various locations
- For future work, we aim to investigate how we can navigate blind users with PathFinder multi floor routes by integrating various algorithm

Navigation To Exploration

Exploration involves navigation as well as learning
Recent MLLM has potential to tackle exploration

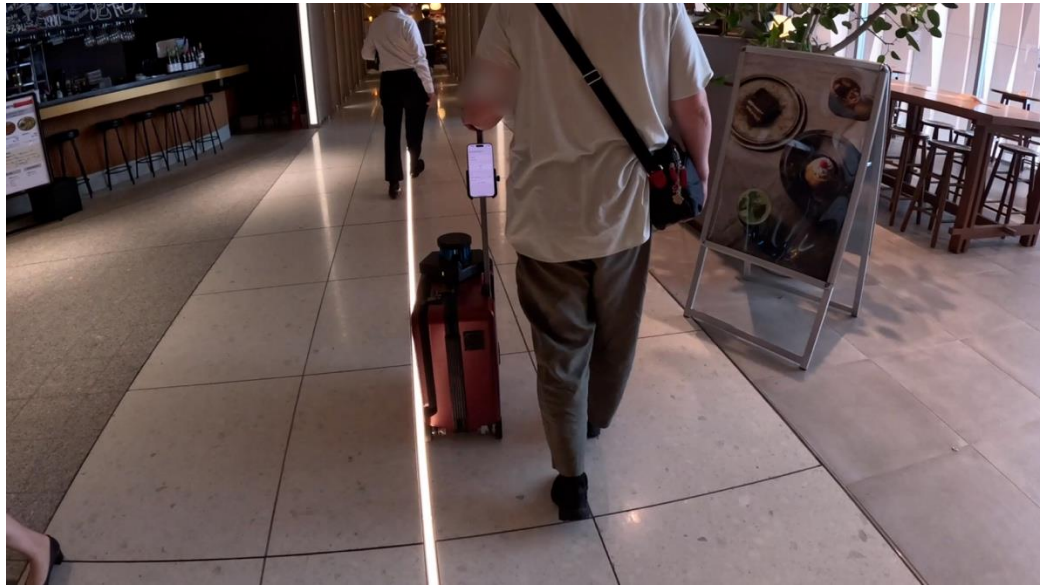
PathFinder

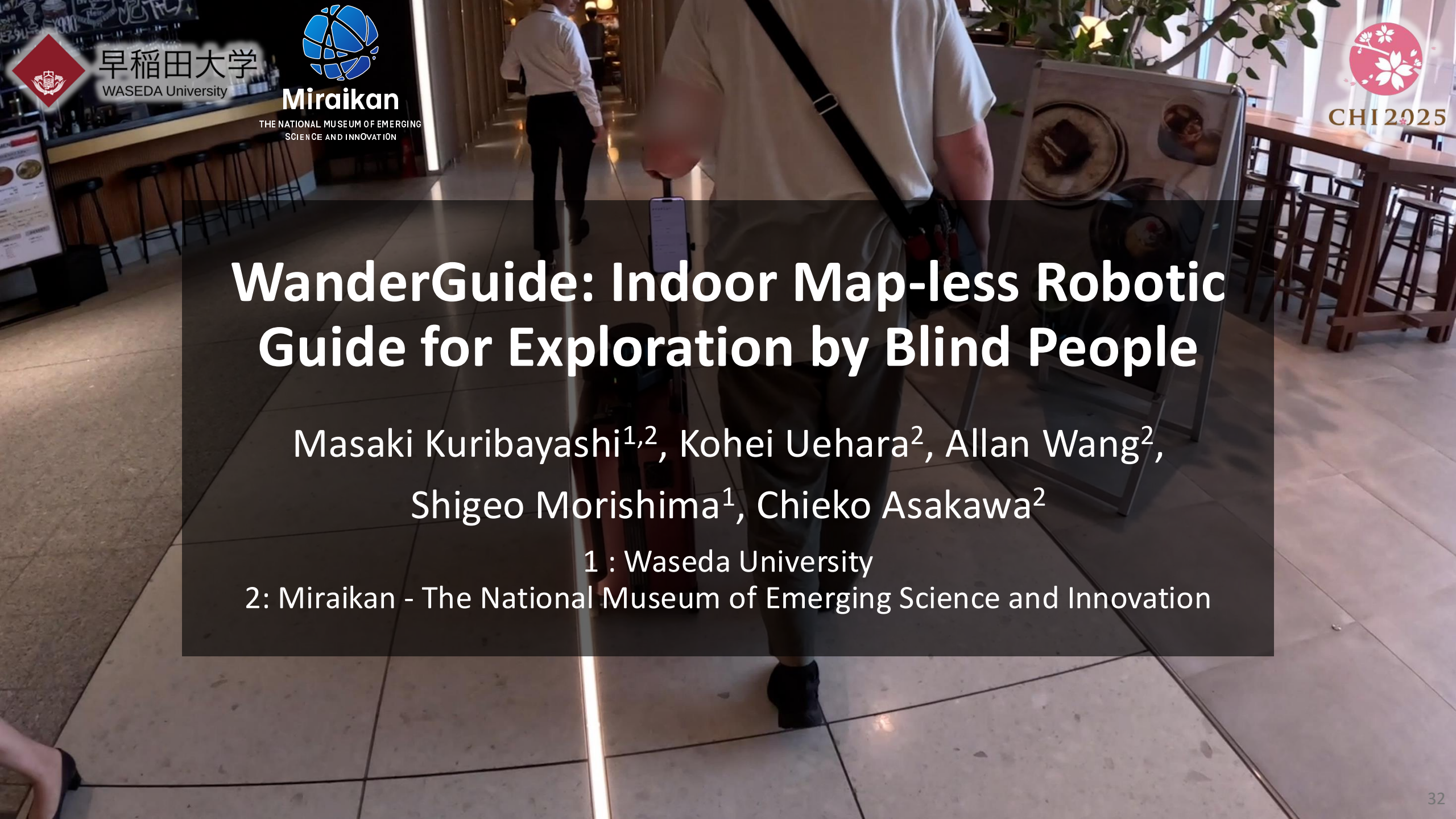
Navigation System



WanderGuide

Exploration System





WanderGuide: Indoor Map-less Robotic Guide for Exploration by Blind People

Masaki Kuribayashi[1,2], Kohei Uehara[2], Allan Wang[2],

Shigeo Morishima[1], Chieko Asakawa[2]

1 : Waseda University

2: Miraikan - The National Museum of Emerging Science and Innovation

Design Procedure of WanderGuide

RQ: What is the system requirement for WanderGuide?

Step 1: Prototype WanderGuide

Extended PathFinder with more general intersection detection algorithm and MLLM-powered scene description

Step 2: Improve Based on Formative Study

Conducted need finding study with 5 blind participants at Toranomon-Hills and Miraikan and revealed that blind people appreciate MLLM-powered description

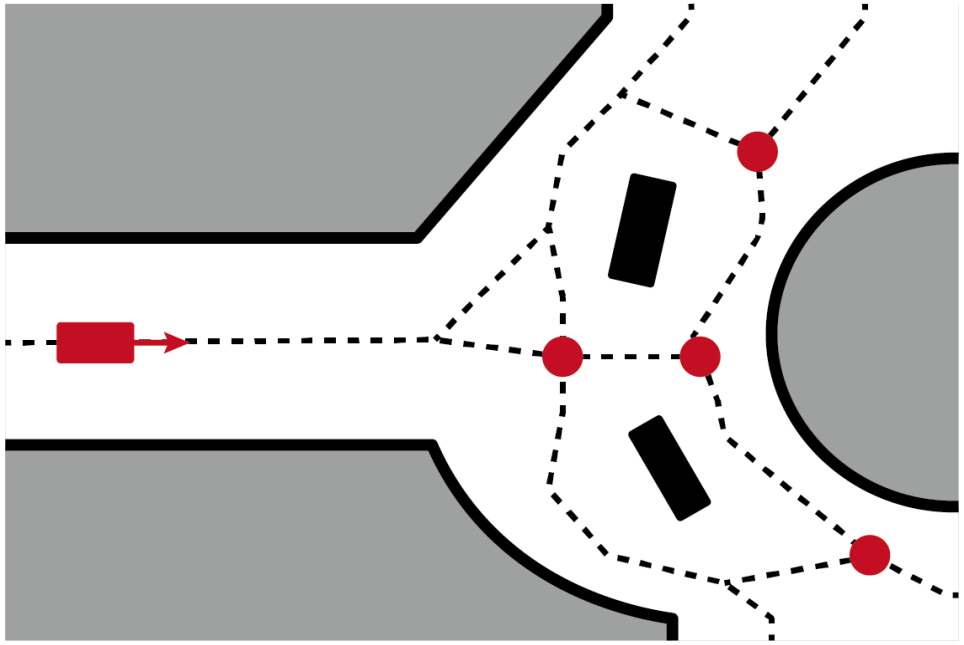
Step 3: Iterative Study for Obtaining Further Improvement

Conducted main study in Miraikan and revealed that users were able to find interesting place with the system and future improvement is in audio recognition

Initial System Design of WanderGuide

Waypoint Detection

More generalized intersection detection



MLLM-Based Scene Description & QA

Describe various information



System Overview with MLLM-powered Surrounding Explanation



Conversation Function



The video is demonstration. The sound and performance of the system may differ.

Surrounding Description & Conversation by GPT-4o

- Based on three image inputs, issues description of the surrounding every 5s
- Prompt was designed based on description guideline[1]
  (*e.g.*, describe general scene and then left, front and right)

Example: This is a futuristic exhibition hall that has vibrant displays. To your left, there is a uniquely shaped wooden table and archway. Ahead, you can see a curved blue sofa and a white sign that reads "Entrance." On your right, large colorful panels line the wall, displaying information about the future and health.







[1] Jain et al, “"I Want to Figure Things Out": Supporting Exploration in Navigation for People with Visual Impairments,” CSCW2023

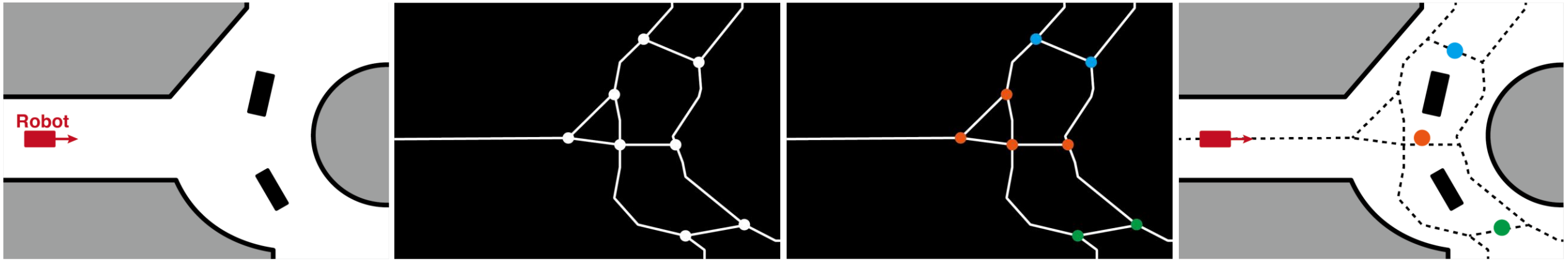
Waypoint Detection and Map-less Navigation Algorithm

Obtain cost map

Apply skeletonization and candidate point

Cluster points with DBSCAN

Take mid point of cluster as way point



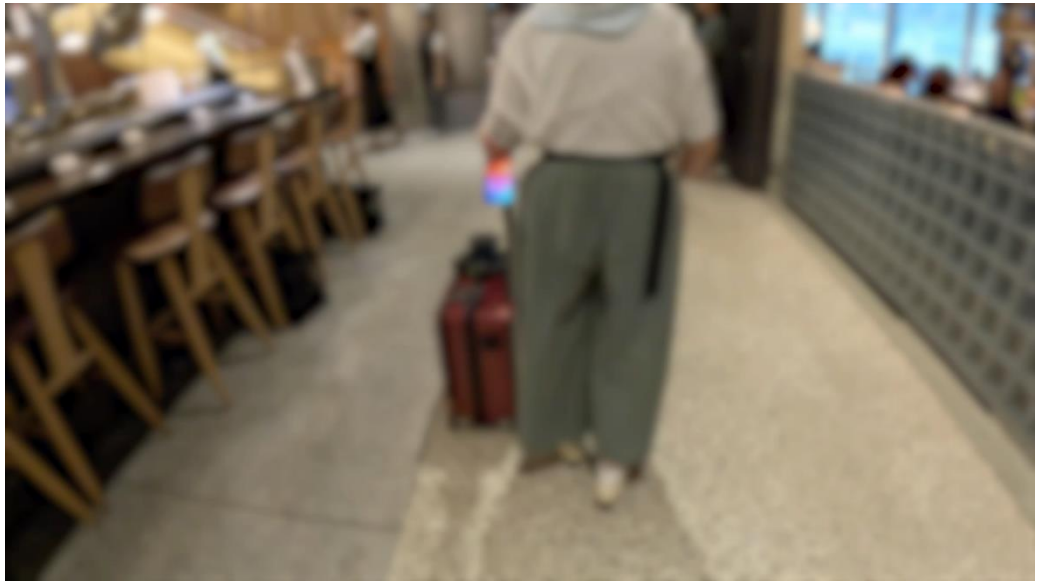
Minimum points for cluster=1
Neighborhood radius=1.25m

Formative Study with 10 Blind Participants

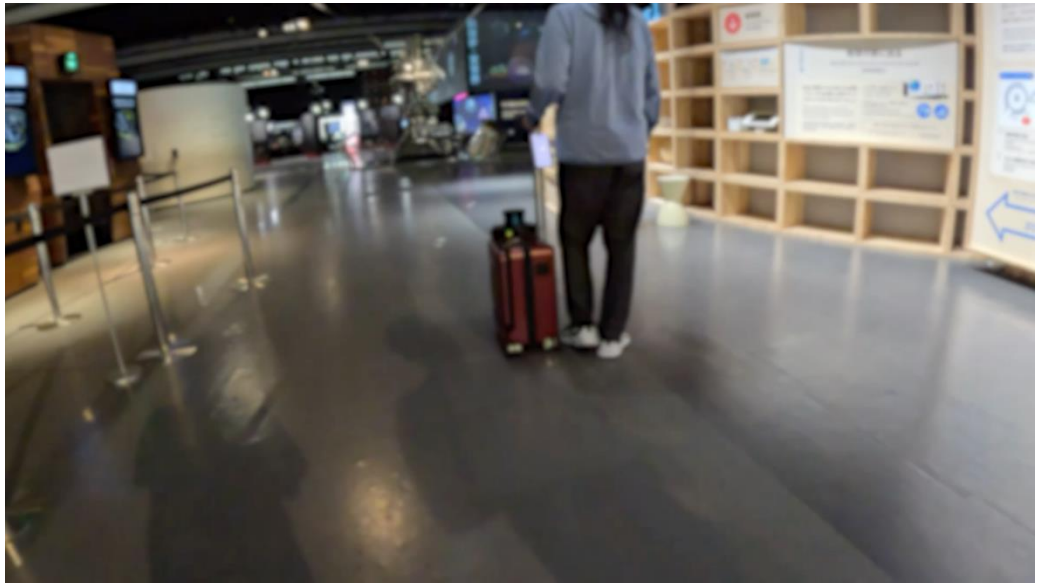
RQ1: How should the MLLM describe environment?

RQ2: What other functionality should WanderGuide have?

Shopping Mall



Science Museum



Design Improvement From The Formative Study

Improvement 1: Need of Varying Level of Description

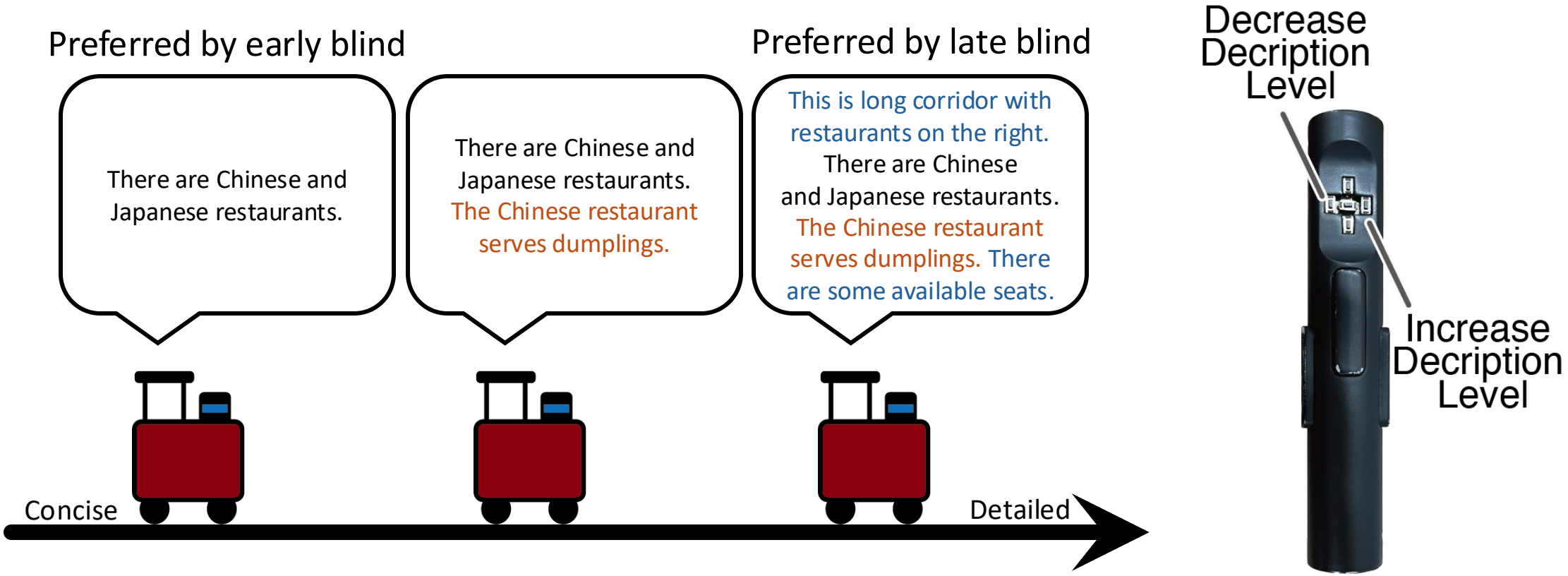
Preferences of the length in described descriptions was observed

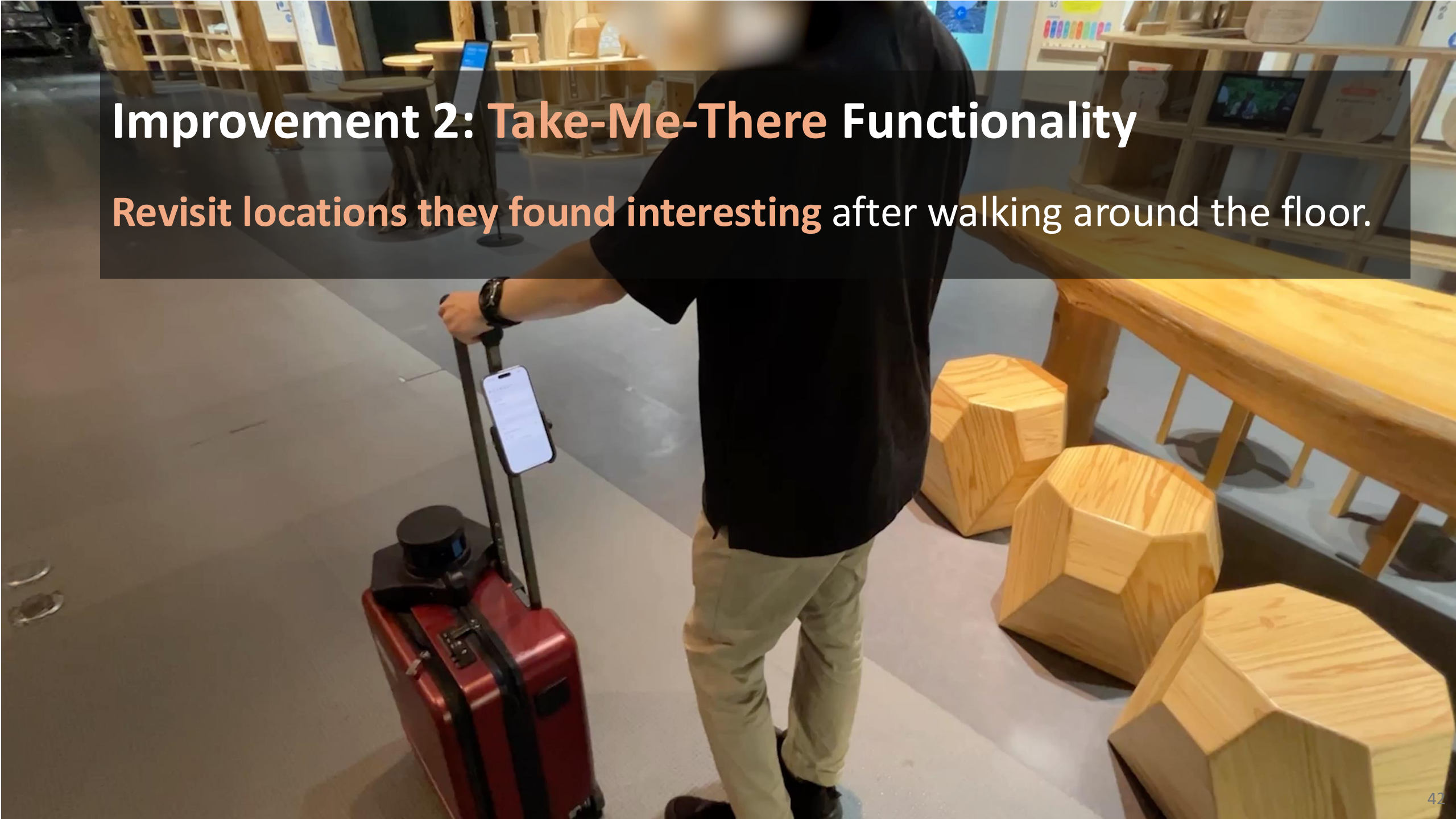
Improvement 2: More Concrete Information

Add high-res camera and modify prompting

Improvement 1: Varying Level of Description

Adjust the detail level by using
three different prompts specifying sentence length.





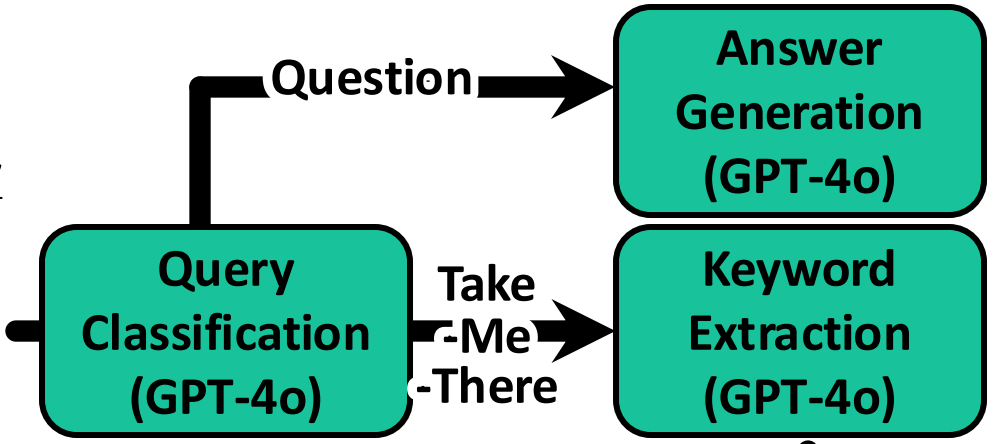
Improvement 2: Take-Me-There Functionality

Revisit locations they found interesting after walking around the floor.

Implementation: Take-Me-There Functionality

Input: User Query

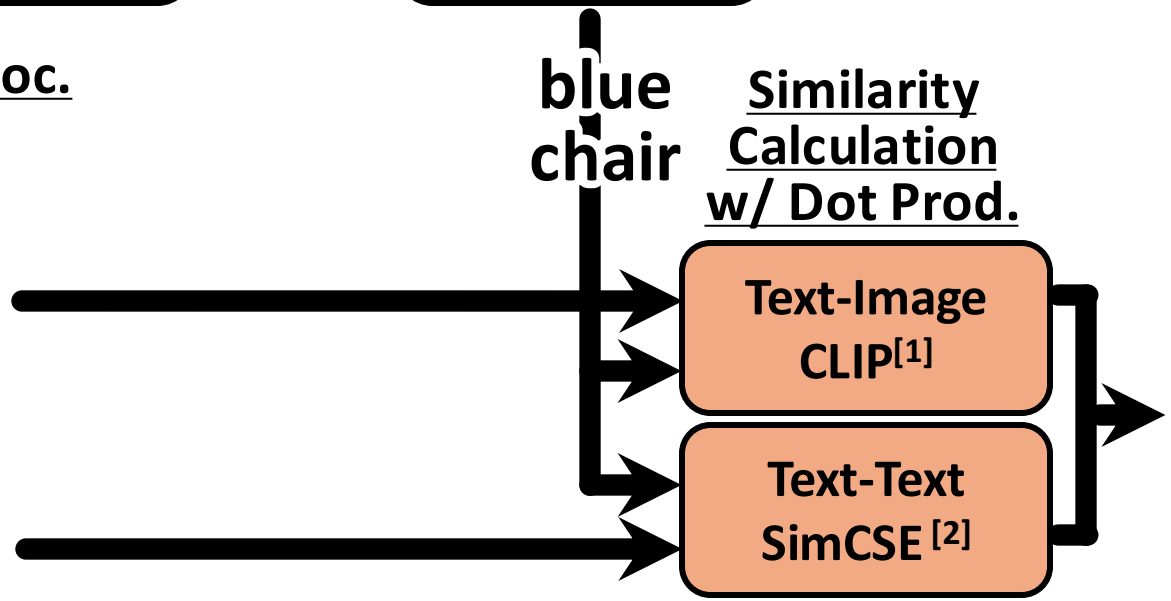
I want to go to the blue chair, and take some rest



Semantic Map for Visited Loc.

Location	Image
(x0, y0)	
(x1, y1)	…

Location	Text
(x2, y2)	There is a big book…
(x3, y3)	Left, wooden desk…



Location Selection

Location	Similarity
(x0, y0)	0.95
(x1, y1)	0.89
(x2, y2)	0.01
(x3, y3)	0.12

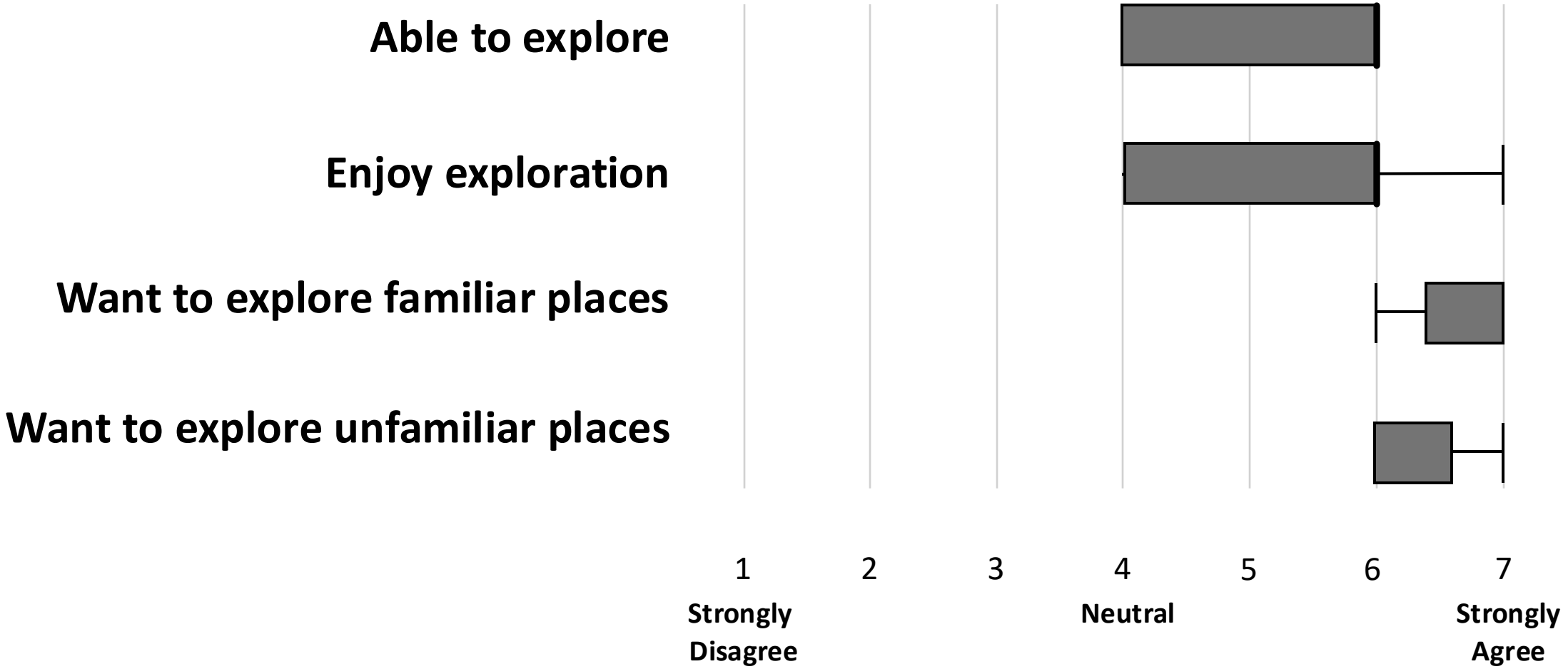
[1] Radford et al, "Learning transferable visual models from natural language supervision," ICML2021
[2] Gao et al, "SimCSE: Simple Contrastive Learning of Sentence Embeddings," EMNLP2021

Iterative User Study with 5 Blind Participants

Task: Use WanderGuide for 20 minutes and explore the Science Museum Freely



Participants Enjoyed Exploration and Expressed a Desire to Use It in Various Locations



Participants Varied Description Level Also Based on Context

Proportion of time spent using each description level

	Concise	Balanced	Detailed
P11	0.10%	**76.46%**	**23.43%**
P12	**15.22%**	**29.88%**	**54.91%**
P13	0.19%	**49.14%**	**50.66%**
P14	0.15%	**87.94%**	**11.91%**
P15	0.14%	**99.86%**	0.00%

Conversation Analysis

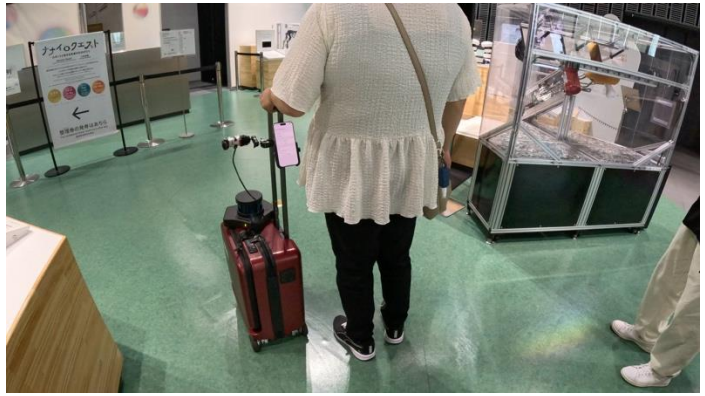
General Query: Request for general scene description
Specific Query: Request for specific detail (e.g., content written in the panel)
Command Query: Request to go to specific destination (e.g., Take-Me-There function)

	General	Specific	Command Query				Count
			Go to Specific Location	Go to Sound Source	Go to Initial Location	Others (e.g., Go along the wall)	
P11	11%	46%	**14%**	11%	0%	17%	35
P12	30%	10%	**50%**	0%	10%	0%	10
P13	62%	38%	**0%**	0%	0%	0%	13
P14	14%	46%	**25%**	0%	0%	14%	28
P15	8%	25%	**46%**	0%	4%	17%	24

Success Scenario: Finding Point of Interest

Users were able to find interesting place based on MLLM-based description and used Take-Me-There functionality to further explore

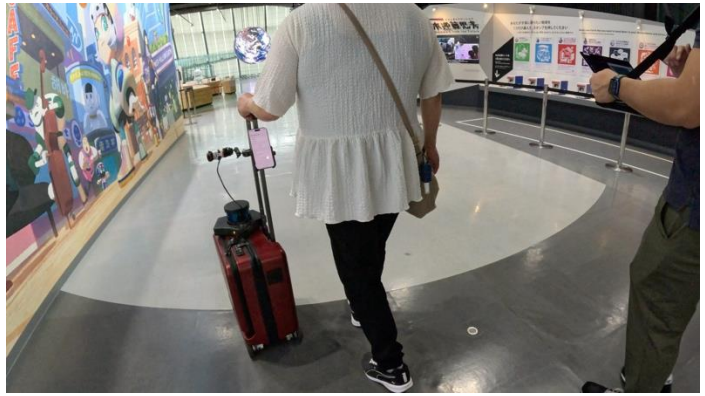
1. Robot goes near “Nanairo-Quest” and records it



2. User finish walking around the floor and asks “I want to go to something quest”



3. The robot takes user near the “Nanairo-Quest” with Take-Me-There functionality



Future Work: Need of Audio Recognition

Occurrence: Users asked about the source of the sound the heard and even specified the sound source and the place they want to explore



P13

“At a place like the exhibition hall we're visiting this time, various sounds are coming from all directions. This prompts questions like, “What's happening at that sound over there?” ***It would be nice if we could ask specific questions like, “What's that sound coming from the right?”*** *”*

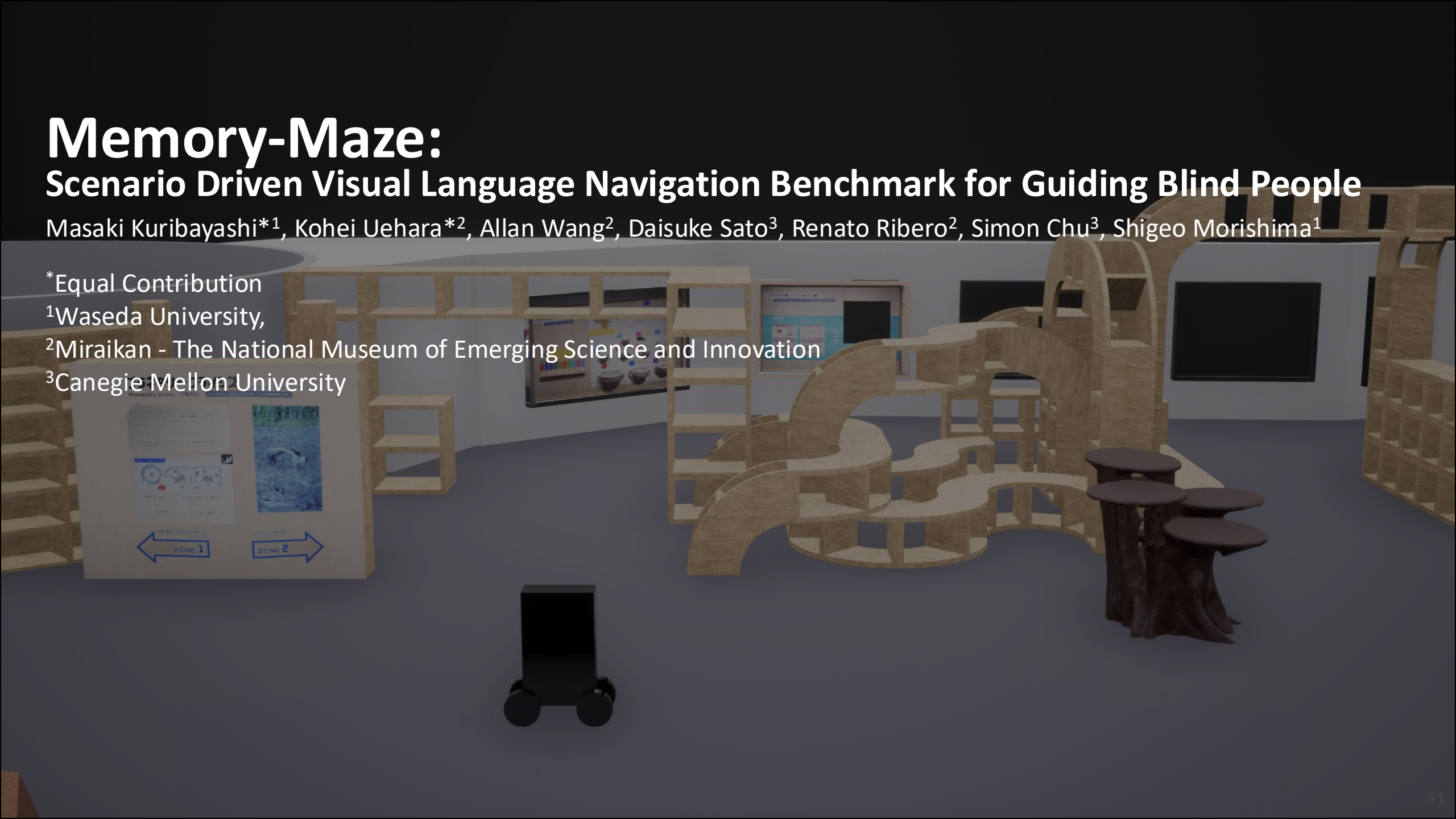
WanderGuide: Indoor Map-less Robotic Guide for Exploration by Blind People

- We revealed that blind people can enjoy exploration by using WanderGuide
- Through the formative study, we revealed several system requirement such as varying level of description and “Take-Me-There” functionality
- Through the study, we revealed the potential of blind people determining their destination during exploration by listening to MLLM-based description
- For future work, we plan to improve MLLM-based description and integrate audio recognition functionality

Memory-Maze:
Scenario Driven Visual Language Navigation Benchmark for Guiding Blind People

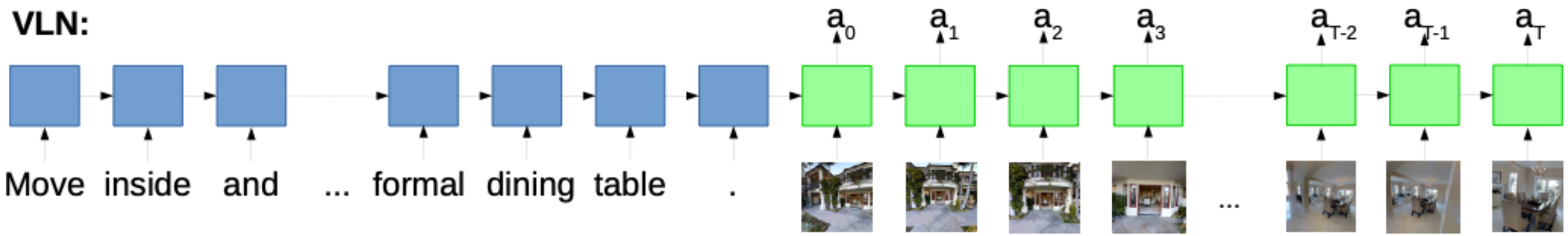
Masaki Kuribayashi*[1], Kohei Uehara*[2], Allan Wang[2], Daisuke Sato[3], Renato Ribero[2], Simon Chu[3], Shigeo Morishima[1]

*Equal Contribution
[1]Waseda University,
[2]Miraikan - The National Museum of Emerging Science and Innovation
[3]Canegie Mellon University



What is Visual Language Navigation (VLN)?

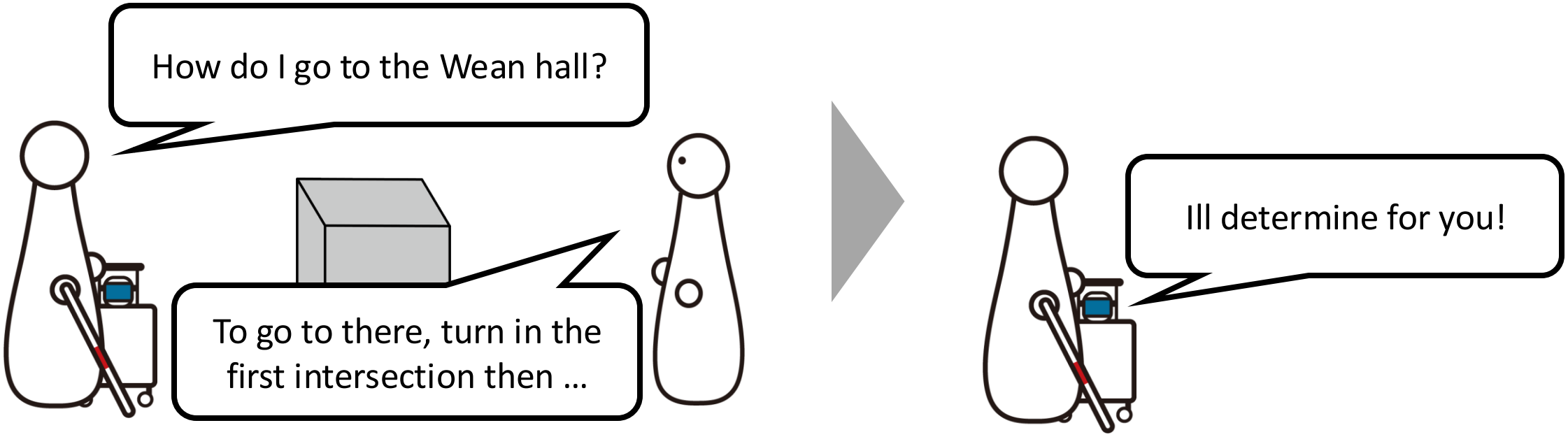
Agent navigates to language specified destination with vision[1]



[1] Anderson et al, "Vision-and-language Navigation: Interpreting Visually-grounded Navigation Instructions in Real Environments," CVPR2018

Application of VLN for Map-less Navigation

VLN agents can benefit blind people in the PathFinder situation by determining which way to go instead of the users



Prior VLN Benchmarks Overlook Blind Navigation Scenario

Environment Difference

Existing Scenario[1]
Static House

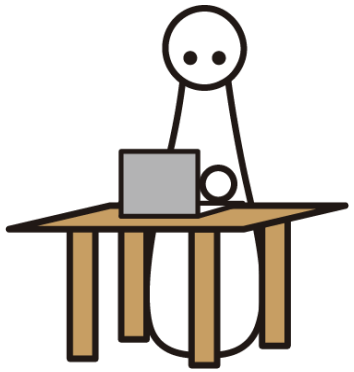


Blind Nav. Scenario
Public Buildings

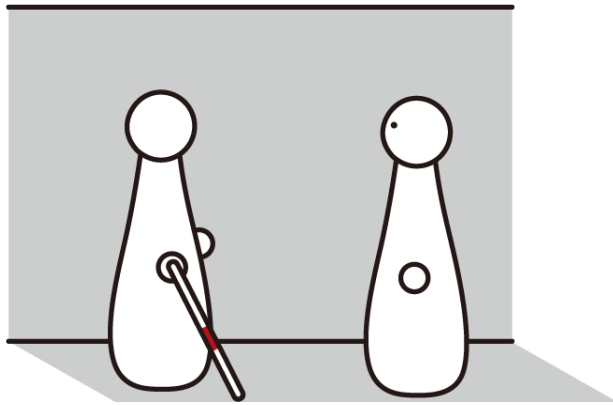


Instruction Difference

Existing Scenario[1]
Type Online



Blind Nav. Scenario
Describe Onsite



[1] Anderson et al, "Vision-and-language Navigation: Interpreting Visually-grounded Navigation Instructions in Real Environments," CVPR2018

Memory-Maze Benchmark

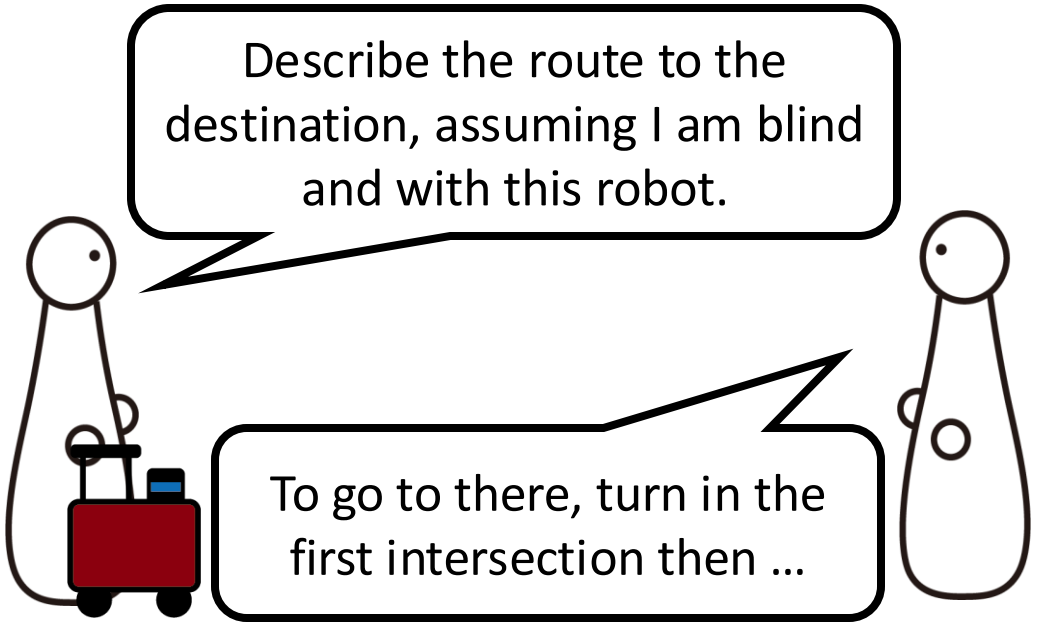
Contains virtual environment of public buildings

and memory-based instructions through user studies

Memory and Observation-Based Instruction Collection

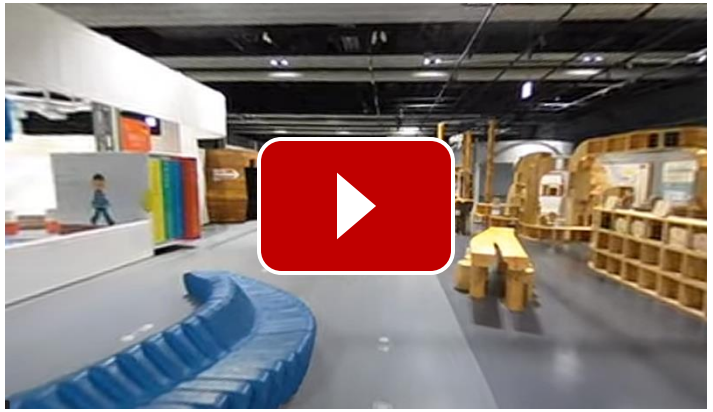
Onsite Study for Memory-based Instructions

Describe route to destination from memory
Collected 252 Instructions



Online Study for Observation-based Instructions

Watch first person video and type in instructions
Collected 312 Instructions



Describe the route to the destination.

First, turn left and go…

Memory and Observation-Based Instruction Comparison

Memory-based Instructions

92.17 Words (x1.15)
33.76% Errors (x4.19)
30.17% Alt. Routes

"So go straight for **20 meters and turn left. And then turn left.** There will be a hall connecting **Scott Hall and this is... Wait, this is... Newell Simon Hall.**[...] **If you turn right, there is a small... You can cross it**, but it's **probably** like one to two meters opening there. [...] And then cross that hall. And then **I think** you can just go straight downstairs, **probably like 10 steps**, and then you will be there, the cafe."

Observation-based Instructions

80.35 Words
8.06% Errors

"Go a little way down this road, then continue straight after turning left. Along the way, you will pass through a path lined with glass on both sides. After that, turn right at the dead end and follow the road, then turn right again before the stairs. Continue straight to reach your destination."

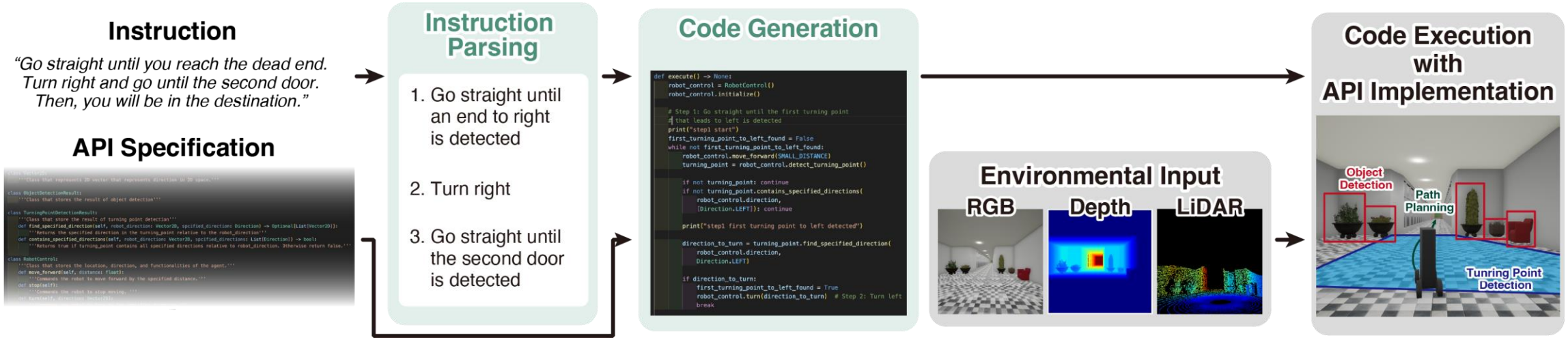
LLM-Based VLN Model Requirement

No Additional Training

Foundation model (GPT-4) is used for navigation in unseen environments, where training-based methods often fail.

Minimum Latency

Open-loop code generation with a robot navigation API is used to reduce delay during navigation.



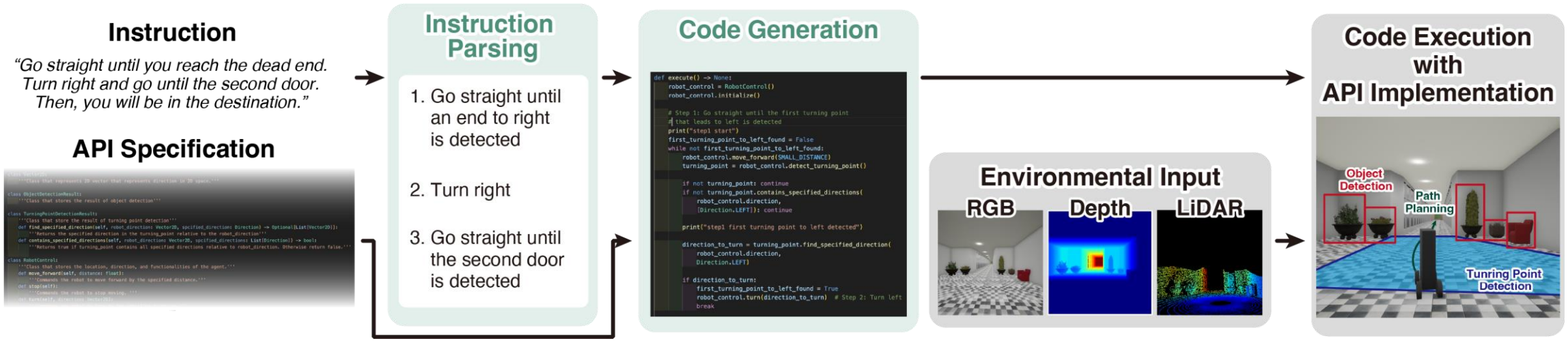
LLM-Based VLN Model Detail

Rule Designing[1] Based on Training Set (10% of samples)

- Ex 1. To specify the shape and number of intersection if possible
- Ex 2. To ignore repetitive route description

Robot Control API

- Designed and formatted based on similar approach in object navigation[2]
- Ex. move_forward(distance: float), detect_from_RGB_image(object: str), turn(direction: str)



[1] Shah et al, "LM-Nav: Robotic Navigation with Large Pre-trained Models of Language, Vision, and Action," CoRL2023
[2] Biggie et al, "Tell Me Where to Go: A Composable Framework for Context-Aware Embodied Robot Navigation," CoRL2023

VLN Metrics

- **Success Rate (SR)**: Robot stopped within 3m at the destination
- **Oracle Success Rate (OSR)**: Robot was within a 3m from the target location
- **Shortest Path Distance (SPD)**: Distance between the robot's final location and the goal
- **Coverage Weighted by Length Score (CLS)**:
  How closely an agent's trajectory follows the reference path

Condition	Online Study Data				Onsite Study Data			
	SR↑	OSR↑	SPD↓	CLS↑	SR↑	OSR↑	SPD↓	CLS↑
NavGPT	0.01	0.03	67.37	0.04	0.00	0.02	64.19	0.05
NaVid	0.00	0.00	70.83	0.02	0.00	0.00	71.32	0.02
Proposed	0.09	0.15	45.45	0.34	0.06	0.09	51.75	0.32
Proposed (+Parser)	**0.11**	**0.22**	**37.29**	**0.42**	**0.08**	**0.12**	**40.39**	**0.44**

Result: LLM-Based Method Improves Moderately Over Prior Methods

1. Prior methods exhibited vey poor performance
2. Parser contributed to the performance improvement
3. Onsite study was more difficult than online study

Condition	Online Study Data				Onsite Study Data			
	SR↑	OSR↑	SPD↓	CLS↑	SR↑	OSR↑	SPD↓	CLS↑
NavGPT	0.01	0.03	67.37	0.04	0.00	0.02	64.19	0.05
NaVid	0.00	0.00	70.83	0.02	0.00	0.00	71.32	0.02
Proposed	0.09	0.15	45.45	0.34	0.06	0.09	51.75	0.32
Proposed (+Parser)	**0.11**	**0.22**	**37.29**	**0.42**	**0.08**	**0.12**	**40.39**	**0.44**

Future Work: Interactive Decision Making

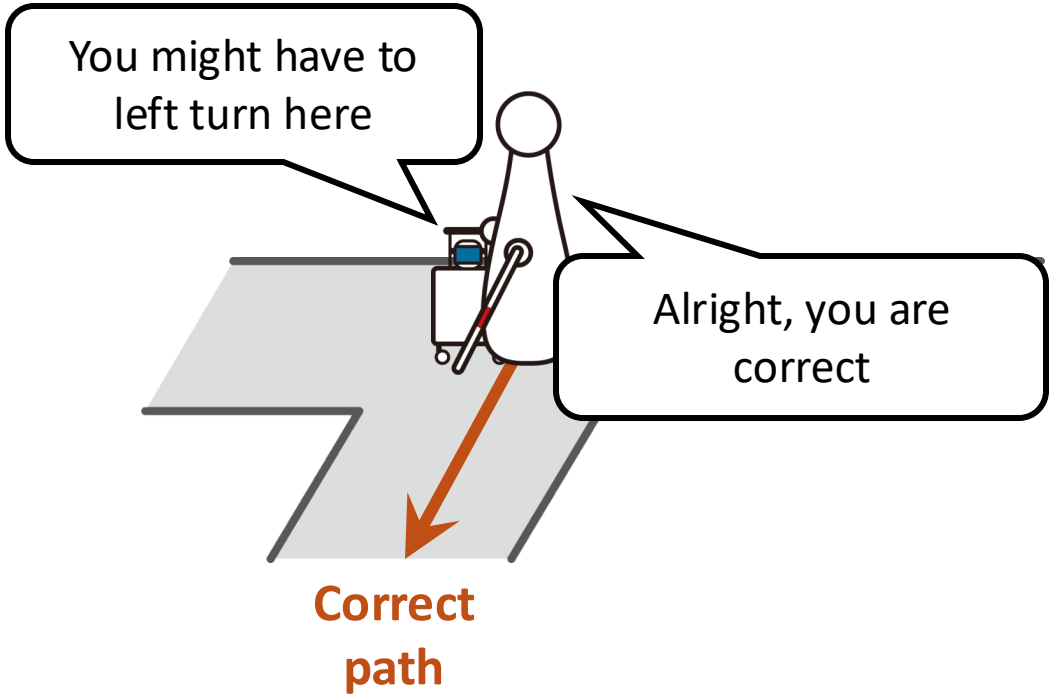
Related Work: Interactive VLN[1]

Model asks for help when it is not confident about next action



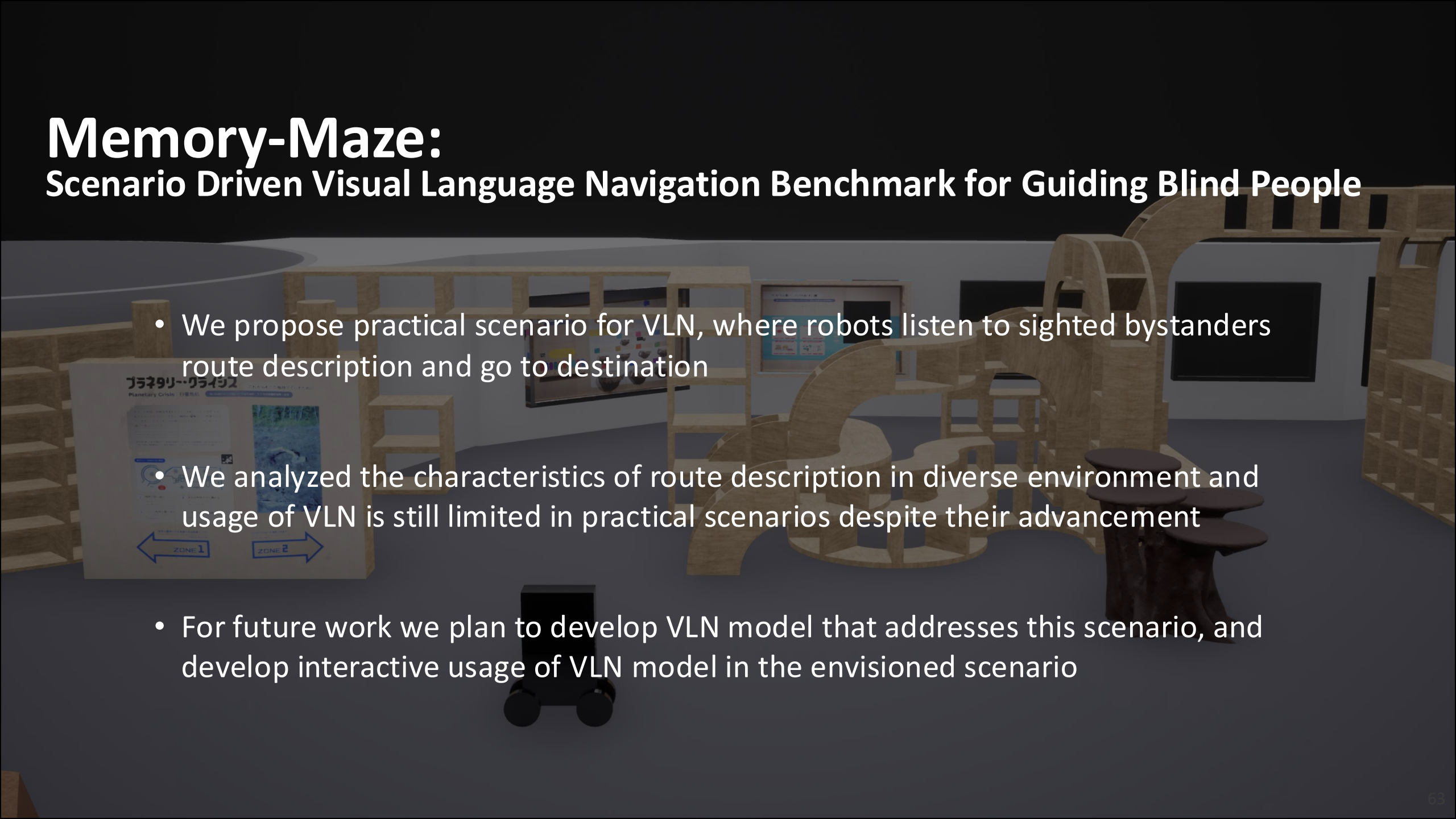
Future Work: Adopt Hybrid Approach

Model ask blind user for help



[1] Chi et al, "Just Ask: An Interactive Learning Framework for Vision and Language Navigation," AAAI2020

Memory-Maze:
Scenario Driven Visual Language Navigation Benchmark for Guiding Blind People



- We propose practical scenario for VLN, where robots listen to sighted bystanders route description and go to destination
- We analyzed the characteristics of route description in diverse environment and usage of VLN is still limited in practical scenarios despite their advancement
- For future work we plan to develop VLN model that addresses this scenario, and develop interactive usage of VLN model in the envisioned scenario

Conclusion: Map-less Navigation for Blind People

Initial Investigation of Collaborative Interaction in Navigation Context

Despite having need to interact more, users appreciated map-less system.
Showed collaborative interactions how to use of user capabilities and systems functions.

Extension to Exploration Task

Tacked exploration problem which goes beyond navigation. Showed the design implications of how to used current AI for better exploration activity.

Finding Interdisciplinary Needs

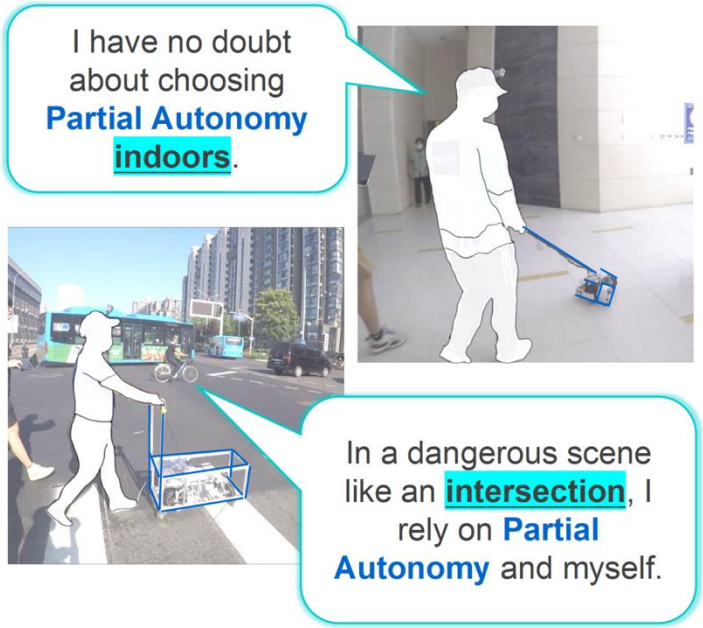
To realize the system, the scope went beyond HCI. The map-less system requirement led to finding overlooked aspect in VLN.

Future Work: Stronger Intervention to Robot

Collaborative interaction requires **blind users' decision making based on system feedback** for **complementing systems' knowledge lack.**

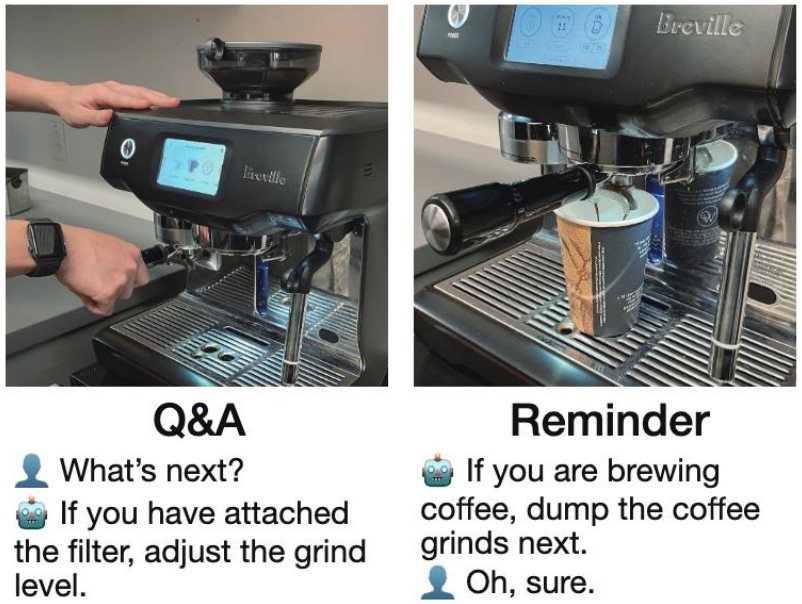
HRI: Shared Control[1]

Users intervene to robot navigation behavior for **better user autonomy, agency, and control.**



HAI: Mixed Initiative[2]

Both human or AI initiate interaction for **complementing AI's imperfection**



[1] Zhang et al., ""I am the follower, also the boss": Exploring Different Levels of Autonomy and Machine Forms of Guiding Robots for the Visually Impaired," CHI2024

[2] Arakawa et al., "Scaling Context-Aware Task Assistants that Learn from Demonstration and Adapt through Mixed-Initiative Dialogue" UIST2025

Crowd/Obstacle Scenario: Delegation of “Excuse Me”

Long-term study revealed that users changed their preference in interaction overtime, while others consistently relied on robot or initiated interaction themselves

