RoDiL: Giving Route Directions with Landmarks by Robots

Kanta Tachikawa[1][a], Shota Akahori[1][b], Kohei Okuoka[1][c], Mitsuhiko Kimoto[2][d]
and Michita Imai[1][e]
[1]*Department of Engineering Informatics, Keio University, Tokyo, Japan*
[2]*Department of Mechanical Engineering Informatics, Meiji University, Tokyo, Japan*

Keywords: Route Guiding, Human-Robot-Interaction, Large Language Model.

Abstract: For social robots, a critical aspect is the design of mechanisms for providing information that is understandable to a human recipient. In tasks such as giving route directions, robots must explain the route clearly to ensure that the user can reach the destination. However, most studies on guiding robots have assumed that the robot will only present a route from its current location without considering the route's complexity. In this study, we propose a robot guiding system, RoDiL (Route Directions with Landmarks), that aims to guide users along a simple route by leveraging their knowledge of a city, especially when the route from the current location to the destination is complex. Specifically, within the context of user interaction, RoDiL comprehends which landmarks are familiar to the user. Subsequently, RoDiL initiates giving route directions using the landmark familiar to the user as the starting point. We conducted an experimental comparison between landmark-based guidance and non-landmark-based guidance with 100 participants. Landmark-based guidance was evaluated significantly more highly when the direct route from the current location was complex. In contrast, when the route from the current location was simple, non-landmark-based guidance was preferred. These results confirm the efficacy of the RoDiL design criteria.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

1 INTRODUCTION

Social robots capable of performing tasks on behalf of humans are becoming increasingly prevalent (Kato et al., 2004)(Kanda et al., 2010)(Wallén, 2008)(Kim and Yoon, 2014)(Pitsch and Wrede, 2014). By taking over tasks traditionally performed by humans, robots can reduce the burden on people. One example of the increasingly utilized social robots is the guiding robot. Guiding robots are robots that perform the task of giving route directions to users (Luna et al., 2018)(Triebel et al., 2016). This study investigates such social robots tasked with giving route directions.

However, the task of giving route directions presents a significant challenge for robots. When it comes to tasks that convey routes to humans, the critical aspect is how to make the route understandable to the recipient. For example, in cases where the route is characterized by a high number of turns, long distances, and complexity, merely explaining the route may not be effective, as humans may not remember the explained route in its entirety. In other words, the robot must carefully craft its utterances and methods of route selection to ensure that the user fully understands the entire route.

The theme of guiding robot has already been researched in previous studies. For example, in the study on methods of route presentation, Oßwald et al. proposed a method for robots to provide natural route descriptions by imitating the sentences that humans explain routes (Oßwald et al., 2014). Furthermore, in the study of route generation, Russo et al. proposed an algorithm that uses visible landmarks as nodes in indoor environments to generate low-level routes and simplified text (Russo et al., 2014).

These studies demonstrated efficient methods for explaining simple routes which is characterized by a minimal number of turns, short distances, and simplicity. However, most previous studies on robots giving route directions assumed that the robot would guide the user from the current location without considering the route's complexity, as shown on the left in Figure 1. There is a lack of investigation into more realistic approaches for effectively explaining longer distances. In other words, those studies did not investigate how to transform complex routes into simple, comprehensible route directions.

---

[a] https://orcid.org/0009-0007-3674-5636
[b] https://orcid.org/0009-0009-1434-3868
[c] https://orcid.org/0000-0002-5569-3356
[d] https://orcid.org/0000-0001-8441-8815
[e] https://orcid.org/0000-0002-2825-1560

Tachikawa, K., Akahori, S., Okuoka, K., Kimoto, M. and Imai, M.
RoDiL: Giving Route Directions with Landmarks by Robots.
DOI: 10.5220/0013131300003890
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 17th International Conference on Agents and Artificial Intelligence (ICAART 2025) - Volume 1*, pages 298-309
ISBN: 978-989-758-737-5; ISSN: 2184-433X
Proceedings Copyright © 2025 by SCITEPRESS – Science and Technology Publications, Lda.

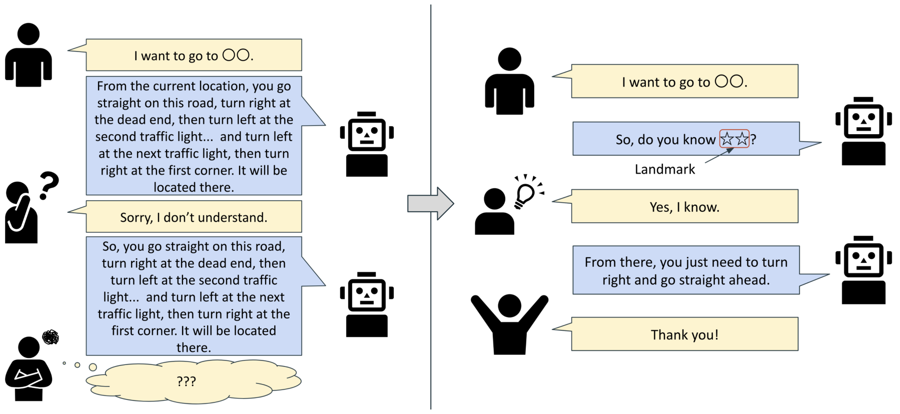

I want to go to 〇〇.
From the current location, you go straight on this road, turn right at the dead end, then turn left at the second traffic light... and turn left at the next traffic light, then turn right at the first corner. It will be located there.
Sorry, I don't understand.
So, you go straight on this road, turn right at the dead end, then turn left at the second traffic light... and turn left at the next traffic light, then turn right at the first corner. It will be located there.
???
I want to go to 〇〇.
So, do you know ☆☆?
Landmark
Yes, I know.
From there, you just need to turn right and go straight ahead.
Thank you!

Figure 1: Examples of dialogues in previous studies and with RoDiL.

Therefore, we propose a robot system, called RoDiL (Route Directions with Landmarks), that aims to guide users along a simpler route by leveraging their knowledge of a city, especially when the direct route to the destination is complex. In an interaction, RoDiL selects a landmark and inquires whether the user is familiar with it. If the user is familiar with the landmark, RoDiL proceeds to give route directions to the user with a route from that landmark to the destination (right side of Figure 1). In other words, RoDiL does not initially possess a database of landmarks known to the user; instead, it seeks to discover landmarks familiar to the user through the interaction. In this context, the route from the landmark to the destination must be simpler when compared to the route from the current location to the destination. Furthermore, considering the user's travel distance, the landmark should not be located significantly off the route from the current location to the destination. In addition, the effectiveness of a landmark is nullified if unknown to the user; thereby, RoDiL is obligated to suggest landmarks of a high degree of familiarity. In summary, the landmarks suggested by RoDiL must be located along the route from the current location to the destination, have a simple route to the destination, and be well known. To extract landmarks that fulfill specific conditions, RoDiL considers the distance and the number of turns of the route from the landmark to the destination, and the number of reviews the landmark has received on Google Maps for calculation. Finally, RoDiL suggests the landmark with the highest score first. Through user interaction, RoDiL selects routes starting from landmarks that the user is familiar with. Due to the previously mentioned calculations, the chosen route is simpler than the route from the current location. Consequently, this reduces the robot's utterances required for guidance and enhances the clarity of the route direction.

In summary, the contribution of this study lies in the development of a robot system that acquires users' knowledge of a city through interaction and provides understandable explanations of complex routes.

2 RELATED WORK

2.1 Guiding Robots

Previous studies have been conducted on robots capable of giving route directions to users. For example, Richter et al. proposed a process for adapting route guidance to match environmental and route characteristics (Richter and Klippel, 2005)(Richter, 2008). Oßwald et al. introduced an algorithm that enables robots to learn human-written route descriptions using inverse reinforcement learning, thus imitating human descriptions (Oßwald et al., 2014). Their study significantly enhanced the naturalness of robot-generated route descriptions. In addition, Shah et al. proposed a guiding system called LM-Nav (Shah et al., 2023), which was created using the pretrained models ViNG (Shah et al., 2021), CLIP (Radford et al., 2021), and GPT-3 (Brown et al., 2020). LM-Nav provides a high-level interface to users in a vision-based setting where destinations are specified using images. Waldhart et al. introduced the Shared Visual Perspective planner, which gives route directions based on the shared visual perspectives of humans and robots (Waldhart et al., 2019). Morales et al. modeled route concepts based on human cognitive processes as a three-layer model comprising memory, survey, and path layers, thereby developing a system that generates route directions that are understandable by humans (Morales et al., 2014). Inclusion of the path layer allows robots to perform at a level equivalent to that of human expert.

In evaluating guiding robots, Rosenthal et al. demonstrated that concise route descriptions given by

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

robots significantly reduce human attention toward the robots (Rosenthal et al., 2022). Okuno et al. indicated that it is beneficial for robots to perform gestures in sync with their verbal instructions at appropriate times during guidance (Okuno et al., 2009). In another study, Heikkilä showed that the characteristics making human-provided guidance intuitive and effective have a similar impact when humanoid robots perform guiding tasks (Heikkilä et al., 2019).

Other previous studies on guiding robots focused on generating routes. Russo et al. developed an algorithm that generates low-level route directions and simplified text by using visible landmarks along the route in indoor environments (Russo et al., 2014). Zheng et al. proposed a robot that uses panoramic representations to acquire route descriptions from route views captured during trial movements (Zheng and Tsuji, 1992). These descriptions are then used for giving route directions. Moreover, Zhang et al. developed a robot guiding system aimed at visually impaired individuals (Zhang and Ye, 2019). This system estimates the robot's posture by using RGB-D camera images, depth data, and inertial data from an inertial measurement unit (IMU), and the estimated results are applied to route planning.

In addition, other studies demonstrated the utility of robots performing gestures. For example, Okuno et al. demonstrated that appropriateness significantly enhances the utility of gestures in guiding scenarios (Okuno et al., 2009). Salem et al. revealed that nonverbal behaviors such as hand and arm gestures, when performed in conjunction with verbal communication, lead users to positively evaluate a robot (Salem et al., 2011). Kim et al. discovered a correlation between robots' coordinated gestures and factors such as intimacy, homogeneity, and engagement (Kim et al., 2012).

2.2 Communication Robots Using Natural Language Models

Studies on using large language models, such as GPT (Radford et al., 2018), to generate robot utterances include a study by Billing et al. (Billing et al., 2023). They integrated GPT-3 with the Aldebaran Pepper and Nao robots. Levoka et al. applied GPT as an IoT service and proposed a natural language dialogue model for interaction with humanoid robots (Lekova et al., 2022). Zhang et al. introduced Mani-GPT, a model that comprehends environments through object information, understands human intentions through dialogue, generates natural language responses to human inputs, and formulates appropriate action plans to assist humans (Zhang et al., 2023). Mani-GPT thereby enables generation of suitable strategies for individual support.

2.3 Limitations of Previous Studies

Most previous studies on route explanation and generation by robots only considered scenes in which the routes themselves do not exhibit any complexity. If the route from the current location to the destination is complex, it is difficult for users to understand the route correctly, because these previous methods directly explain the complex route as it is. We should thus reduce the quantity of utterances in robots' route explanations to ensure that they are comprehensible to users.

3 RODIL

In this study, we propose RoDiL (Route Directions with Landmarks), a robot guiding system that gives route directions by confirming whether a user has city knowledge and then using landmarks as starting points for guidance when the route from the current location to the destination is complex. Figure 2 shows an operational diagram of RoDiL.

Upon receiving user input specifying a destination, RoDiL queries Google Maps to obtain the route from the current location to the destination and buildings along the route as landmark candidates. The candidates are then sorted by their values of an evaluation function, and the top three landmarks with the highest values are selected. RoDiL compares the routes from the selected landmark candidates to the destination with the direct route from the current location in terms of the number of turns. If there is at least one appropriate landmark, RoDiL suggests landmarks to the user in order of the fewest turns.

3.1 Input and Output

The user's utterance is the entire input to RoDiL. The utterance is converted to text through speech recognition and input to GPT-4 (OpenAI, 2023). Then, the output text is vocalized and delivered as the robot's utterance.

3.2 Search Module

Upon receiving the user's destination, RoDiL passes the current location and the destination to Google Maps. Then, by using the Directions API of Google Maps, which outputs routes between two specified

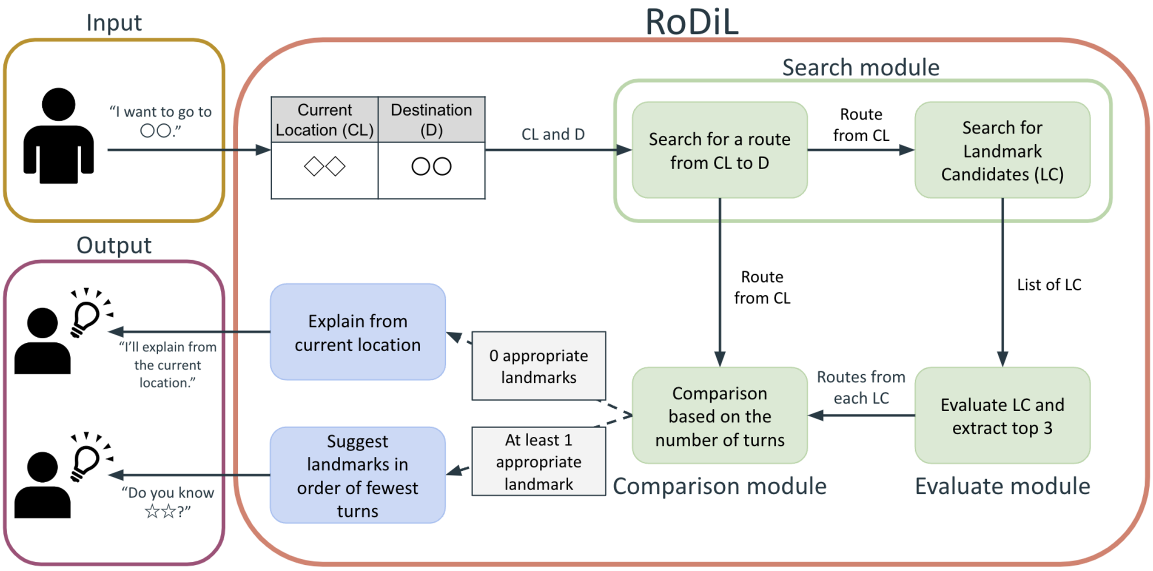

Input
"I want to go to ○○."
RoDiL
Current Location (CL)
Destination (D)
◇◇
○○
CL and D
Search module
Search for a route from CL to D
Route from CL
Search for Landmark Candidates (LC)
Output
Route from CL
List of LC
Explain from current location
"I'll explain from the current location."
0 appropriate landmarks
Comparison based on the number of turns
Routes from each LC
Evaluate LC and extract top 3
Suggest landmarks in order of fewest turns
"Do you know ☆☆?"
At least 1 appropriate landmark
Comparison module
Evaluate module

Figure 2: Operational diagram of RoDiL.

points, RoDiL obtains the route from the current location to the destination.

Next, RoDiL searches for buildings along the obtained route that could potentially serve as landmarks. Buildings that are listed on Google Maps with at least one review are eligible landmark candidates. The rationale behind requiring at least one review is to exclude buildings, such as private homes or commercial buildings, that users are presumed to be unfamiliar with. The Places API is leveraged to search for buildings within a circular area centered on a specified point. By moving the circular search area along the route, RoDiL searches for landmark candidates. The search radius is set to 40 meters to include buildings on adjacent streets as landmark candidates. A similar search is also conducted within a 200-meter radius centered on the destination. This extended search enables the inclusion of buildings that are near the destination, even if they are not located directly along the route from the current location to the destination.

Figure 3 shows an example of a landmark candidate search. The green triangle represents the current location, the red triangle represents the destination, the blue line indicates the route, and the red circles represent the search areas. The dark purple rectangles represent landmark candidates that are subject to search, whereas the light purple rectangles represent landmark candidates that are not searched. RoDiL outputs a list of all the identified landmark candidates.

3.3 Evaluation Module

By sorting the landmark candidates, RoDiL can efficiently propose well-known buildings that the user may be familiar with. Specifically, RoDiL sorts the list of candidates by their values of an evaluation function defined in Eq(1) below, where each landmark is assigned a single value. Here, R denotes the number of reviews, W is a weight, D is the distance (in meters) from the landmark to the destination, and T is the number of turns from the landmark to the destination. The weight W is set to 1 if the landmark is within a circle of radius 40 m, or to 3 if it is outside that circle but within a circle of radius 200 m.

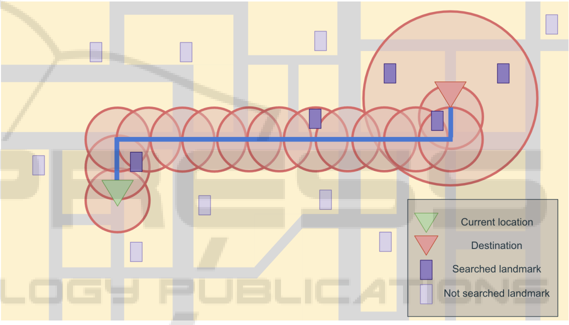


Figure 3: Example of searching landmark candidates.

$$F(Landmark) = 3R - WD - 30T \tag{1}$$

RoDiL extracts the top three landmark candidates with the highest values of the evaluation function, thus limiting the actual proposed landmarks to a maximum of three.

3.4 Comparison Module

Next, RoDiL compares routes from the proposed landmarks with the route from the current location. If

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

the route from the current location is simpler, then it is explained to the user. On the other hand, if the route from the current location is more complex, then giving route directions from a landmark is used to guide the user on a simpler route. In RoDiL, a route's complexity is determined by the number of turns. For the three landmark candidates, RoDiL compares the corresponding routes with the route from the current location in terms of the number of turns, to determine an appropriate number of landmarks to use. The algorithm operates as Algorithm 1. If there are no appropriate landmarks, RoDiL gives route directions from the current location; otherwise, if there is at least one appropriate landmark, RoDiL gives route directions from those landmarks in ascending order of T. In this way, it can suggest landmarks with simpler routes first.

```
Data: landmark candidates $l_1, l_2, l_3$
Data: current location $c$
Result: list of appropriate landmarks $a$
initialization;
if $CountTurn(c) > 3$ then
    for $i \leftarrow 1$ to 3 do
        if $CountTurn(l_i) < CountTurn(c)/2$ then
            $a.append(l_i)$
        end
    end
end
```

Algorithm 1: Determine appropriate landmarks.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

3.5 Gestures

In our study, RoDiL uses the Pepper robot to give route directions. For the robot to visualize guidance, we prepared gestures that were shown to be effective for guiding utterances in previous studies. Specifically, we prepared the three gestures shown in Figure 4: pointing left, pointing right, and extending an arm forward, which respectively correspond to turning left, turning right, and going straight ahead. We assume that the user and robot are interacting face to face; therefore, the gestures for "left" and "right" were designed as mirror images. The robot's utterance is segmented according to punctuation. For each segment containing the word "left," "right," or "straight," the corresponding gesture is triggered simultaneously with the utterance segment.

3.6 Inquiring User's Familiarity of Landmarks

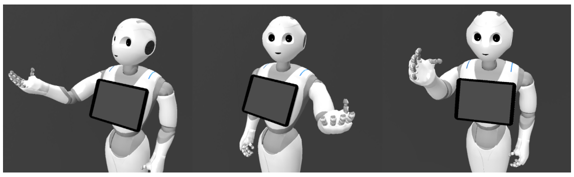
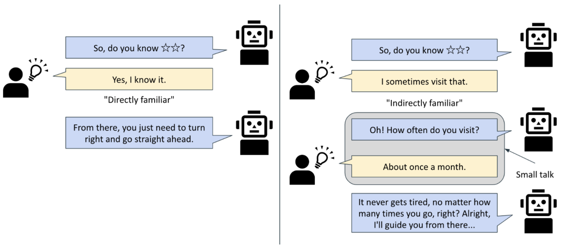


Figure 4: The robot's gestures: "left," "right," and "straight.

Here, we assume that at least one appropriate landmark has been identified. After suggesting a landmark to the user, the user's response is placed in one of two categories: "familiar" and "unfamiliar." Within the "familiar" category, however, there are two conceivable subtypes: direct acknowledgment of familiarity, and indirect indication of familiarity without explicitly stating so. Accordingly, RoDiL generates different utterances for each of these three response patterns: "directly familiar," "indirectly familiar," and "unfamiliar." If the response is "directly familiar," the robot gives route directions from the landmark. For an "indirectly familiar" response, the robot performs small talk about the suggested landmark before giving the directions. The purpose of small talk here is to smooth the flow of communication. Lastly, if the response is "unfamiliar," the robot proposes the next landmark, or it starts explaining from the current location if no further landmarks exist. Figure 5 shows conversation examples for "directly familiar" and "indirectly familiar" responses. RoDiL determines these three patterns using GPT-4 (OpenAI, 2023). GPT processes the user's utterances as input and outputs one of the specified patterns. If the user's response does not fit any of the patterns, i.e. if the user does not respond about recognition of landmarks, GPT outputs "irrelevant". For example, inputting "I know it." results in the output "directly familiar", while inputs such as "I go there often." or "The hamburgers there are tasty, aren't they?" yield the output "indirectly familiar".

Figure 6 illustrates the entire procedure in a flowchart. Following a user's response, the robot's subsequent utterances are modified accordingly.



So, do you know ☆☆?
Yes, I know it.
"Directly familiar"
From there, you just need to turn right and go straight ahead.
So, do you know ☆☆?
I sometimes visit that.
"Indirectly familiar"
Oh! How often do you visit?
About once a month.
Small talk
It never gets tired, no matter how many times you go, right? Alright, I'll guide you from there...

Figure 5: Differences in the robot's utterances due to variation in the user's responses.

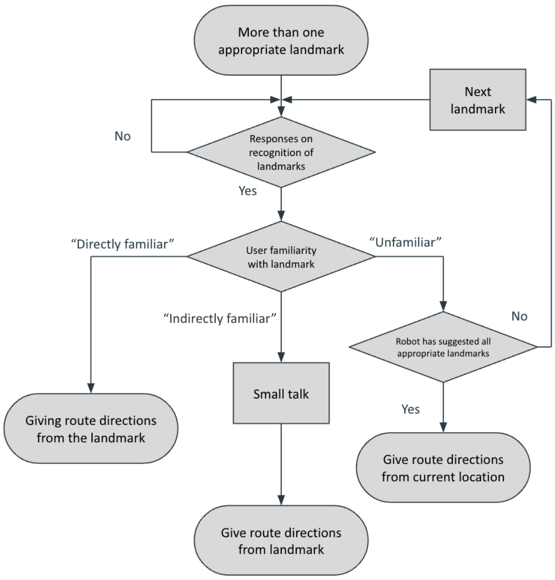

More than one appropriate landmark
Next landmark
No
Responses on recognition of landmarks
Yes
"Directly familiar"
User familiarity with landmark
"Unfamiliar"
"Indirectly familiar"
No
Robot has suggested all appropriate landmarks
Small talk
Giving route directions from the landmark
Yes
Give route directions from current location
Give route directions from landmark

Figure 6: Flowchart of the robot's utterances.

4 EXPERIMENT

4.1 Task Settings

To examine RoDiL's efficacy, we conducted an experiment comparing its guiding with and without landmarks.

We chose Shibuya Station in Shibuya Ward, Tokyo, as the location for giving route directions. Users asked a robot, which was present near Shibuya Station, for route directions to a destination. However, we conducted the interactions with the robot not in the field but within the confines of a laboratory setting.

4.2 Conditions

We considered the following factors: scene and guide. Both scene and guide factors were treated as within-subjects factors, resulting in each participant experiencing four conditions that combined two levels of scene with two levels of guide.

4.2.1 Scene Factor

Scene factor has two levels: Simple and Complex.
Simple. The route from the current location to the destination is simple. The experimental route was from Shibuya Scramble Crossing to the Tower Records Shibuya store. The number of turns was zero.
Complex. The route from the current location to the destination is complex. The experimental route was from the south exit of Shibuya Station to the Tower Records Shibuya store. The number of turns was seven.

4.2.2 Guide Factor

Guide factor has two levels: Without Landmarks and With Landmarks.
Without Landmarks. Regardless of the complexity of the route, the robot always gives route directions from the current location.
With Landmarks. Regardless of the complexity of the route, the robot always gives route directions from a landmark.

4.3 Participants

We used crowdsourcing to recruit 100 participants (63 man, 36 woman, and 1 other, mean age = 42.57 years, *SD* = 9.53 years) for the experiment.

4.4 Video Preparation

In this experiment, we used videos. We recorded four interaction videos, which were created by combining specific scenes with various conditions, as listed in Table 1. Figure 7 shows an image captured from one of the videos. RoDiL comprises two approaches: Simple-Without Landmarks and Complex-With Landmarks. In the With Landmarks condition, it is assumed that RoDiL identifies appropriate landmarks, and when the robot suggested landmarks, the user responded by indicating that he was "directly familiar" with them.

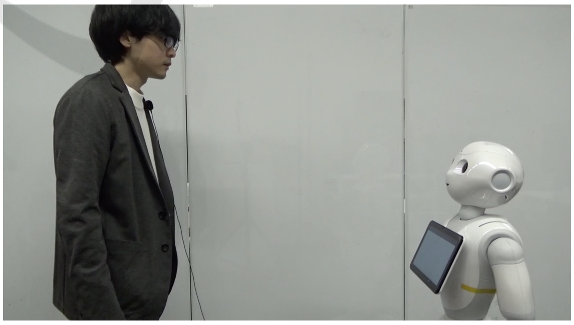


Figure 7: An image captured from one of the scenes in the videos.

4.5 Procedures

The experimental participants were first informed that they would view videos in which the robot gives route directions. After viewing each video, the participants answered a questionnaire for the video and an additional questionnaire for each scene.

Table 1: Recorded Videos.

Title of Video	Content
Simple - Without Landmarks	The robot gives route directions from current location in the Simple scene.
Simple - With Landmarks	The robot gives route directions from a landmark in the Simple scene.
Complex - Without Landmarks	The robot gives route directions from current location in the Complex scene.
Complex - With Landmarks	The robot gives route directions from a landmark in the Complex scene.

4.6 Measurements

The participants responded to a total of three surveys. Surveys 1 and 2 used a seven-point response format (where 1 is the most negative and 7 is the most positive) and were administered after the viewing of each video. Survey 1 is outlined in Table 2, while Survey 2 comprised items from the Godspeed questionnaire (Bartneck et al., 2009). In Survey 3, the participants were asked to choose, in a binary single-answer format, which robot they would prefer to use in each scene. Accordingly, Survey 1 aimed to evaluate the guiding skill, Survey 2 assessed the robot impression, and Survey 3 was meant for an overall evaluation.

Table 2: Survey 1 Contents.

	Question
Q1	Was the guidance clear?
Q2	Could this guidance potentially lead you to reach the destination?
Q3	Was the guidance smooth?
Q4	Was the robot skillful?
Q5	Would you like to try using this robot?

4.7 Hypotheses and Predictions

For scenes where the route from the current location to the destination was complex, we predicted that the With Landmarks condition, using landmarks for simple route guiding, would enhance the guidance clarity and thus improve the participants' impression of the robot. We thus anticipated that, for a complex route from the current location, the With Landmarks condition would receive higher scores on both Survey 1 and Survey 2. Moreover, we expected that a greater number of participants would express a preference for the With Landmarks condition on Survey 3. Conversely, for scenes where the route from the current location to the destination was simple, we predicted that the Without Landmarks condition, giving route directions without any unnecessary conversation about landmarks, would improve both the guidance clarity and the impression of the robot. Therefore, for a simple route from the current location, we expected that the Without Landmarks condition would garner higher evaluations across the surveys. From these predictions, we formulated the following two hypotheses.

H1. When the route from the current location to the destination is simple, the Without Landmarks condition will be perceived more favorably and evaluated more highly.

H2. When the route from the current location to the destination is complex, the With Landmarks condition will be perceived more favorably and evaluated more highly.

5 RESULTS

We conducted a two-way ANOVA on the users' responses to Surveys 1 and 2, with the significance level set to 5%. When a significant interaction was observed, we conducted multiple comparisons by the Bonferroni method. In addition, we conducted a Chi-squared test on the users' responses to Survey 3, with the significance level set to 5%.

5.1 Survey 1: Guiding Skill

Figure 8 shows the results of Survey 1: Guiding Skill. We observed a significant interaction effect across all items, including questions Q1 to Q5 and the overall average (Q1: $F(1,396) = 28.13$, $p < 0.001$, $\eta^2 = 0.058$; Q2: $F(1,396) = 29.27$, $p < 0.001$, $\eta^2 = 0.060$; Q3: $F(1,396) = 26.25$, $p < 0.001$, $\eta^2 = 0.057$; Q4: $F(1,396) = 12.83$, $p < 0.001$, $\eta^2 = 0.030$; Q5: $F(1,396) = 16.35$, $p < 0.001$, $\eta^2 = 0.037$; Average: $F(1,396) = 28.58$, $p < 0.001$, $\eta^2 = 0.062$).

The results from multiple comparisons indicated significant differences for Q3 between Simple-Without Landmarks and Simple-With Landmarks ($p < 0.001$), between Simple-Without Landmarks and Complex-Without Landmarks ($p < 0.001$), and between Simple-Without Landmarks and Complex-With Landmarks ($p < 0.001$). For the other items, we observed significant differences across all items between Complex-Without Landmarks and Simple-Without Landmarks (Q1: $p < 0.001$; Q2: $p < 0.001$; Q4: $p < 0.001$; Q5: $p < 0.001$; Average: $p < 0.001$), between Complex-Without Landmarks and Simple-With Landmarks (Q1: $p < 0.001$; Q2: $p < 0.001$; Q4: $p < 0.001$; Q5: $p < 0.001$; Average: $p < 0.001$), and between Complex-Without Landmarks and Complex-With Landmarks (Q1: $p < 0.001$; Q2: $p < 0.001$; Q4: $p < 0.001$; Q5: $p < 0.001$; Average: $p < 0.001$).

5.2 Survey 2: Robot Impression

Next, Figure 9 shows the results of Survey 2: Robot Impression. From the two-way ANOVA, we found a significant interaction effect only for the item Godspeed III: Likeability ($F(1,396) = 4.535$, $p = 0.034$, $\eta^2 = 0.011$). The results from multiple comparisons indicated significant differences between Complex-Without Landmarks and each of the other three patterns: Simple-Without Landmarks ($p < 0.001$), Simple-With Landmarks ($p < 0.001$), and Complex-With Landmarks ($p = 0.002$)).

5.3 Survey 3: Overall

Lastly, Table 3 shows the results of Survey 3: Overall. The findings indicate that, when the route from the current location was simple, half the participants preferred the robot under the Without Landmarks condition, while the other half preferred the robot under the With Landmarks condition. In contrast, when the route from the current location was complex, 80% of the participants preferred the robot under the With Landmarks condition. We conducted the chi-square test, we found that there was a significant association between the scene factor and the guide factor ($\chi^2(1) = 19.53$, $p < 0.001$, $\phi = 0.31$).

6 DISCUSSION

6.1 Survey 1: Guiding Skill

On Survey 1: Guiding Skill, as shown in Figure 8, we observed significant differences between Complex-Without Landmarks and the other three patterns for Q1, Q2, Q4, Q5, and the average. Of particular importance is the finding that Complex-With Landmarks was evaluated significantly more highly than Complex-Without Landmarks was. This result supports H2, as the With Landmarks condition was evaluated more highly when the route from the current location was complex. Participants reported that "route directions from known locations were more understandable," suggesting that simple route guidance using landmarks was effective. Conversely, in the case of Q3 (smoothness), Simple-Without Landmarks was evaluated significantly more highly than the other three patterns were. We believe that this was because participants perceived superfluous questions about landmarks as hindering smooth guidance.

Furthermore, the fact that Simple-Without Landmarks was evaluated significantly more highly than Simple-With Landmarks was supports H1, as the Without Landmarks condition was perceived more favorably when the route from the current location was simple. Therefore, on Survey 1, H1 was supported by the results of Q3, while H2 was supported by the results of all the questions besides Q3.

6.2 Survey 2: Robot Impression

On Survey 2: Robot Impression, as shown in Figure 9, we only observed a significant difference for Godspeed III: Likeability between the Complex-Without Landmarks pattern and the other three patterns. Participants expressed concerns about delays in the robot's response and the loudness of its motor. We also received comments such as "The robot was cute," and "The robot's behavior was charming." These comments suggest that some participants evaluated the robot in terms of its gestures and attitude, regardless of the guiding content. In this experiment, we did not differentiate the robot's behavior under each condition, which we consider to have yielded results without significant differences. On the other hand, for Likeability, Complex-With Landmarks received significantly higher ratings than Complex-Without Landmarks did. We speculate that this was because many participants felt distrust or discomfort with complex guidance under the Without Landmarks condition. From these results, we conclude that H2 was supported by the results for Likeability on Survey 2.

6.3 Survey 3: Overall

As seen in Table 3, when the route from the current location was complex, 80% of the participants preferred the robot under the With Landmarks condition. These results corroborate that the findings are not attributed to error, as evidenced by the Chi-square test. These results support H2.

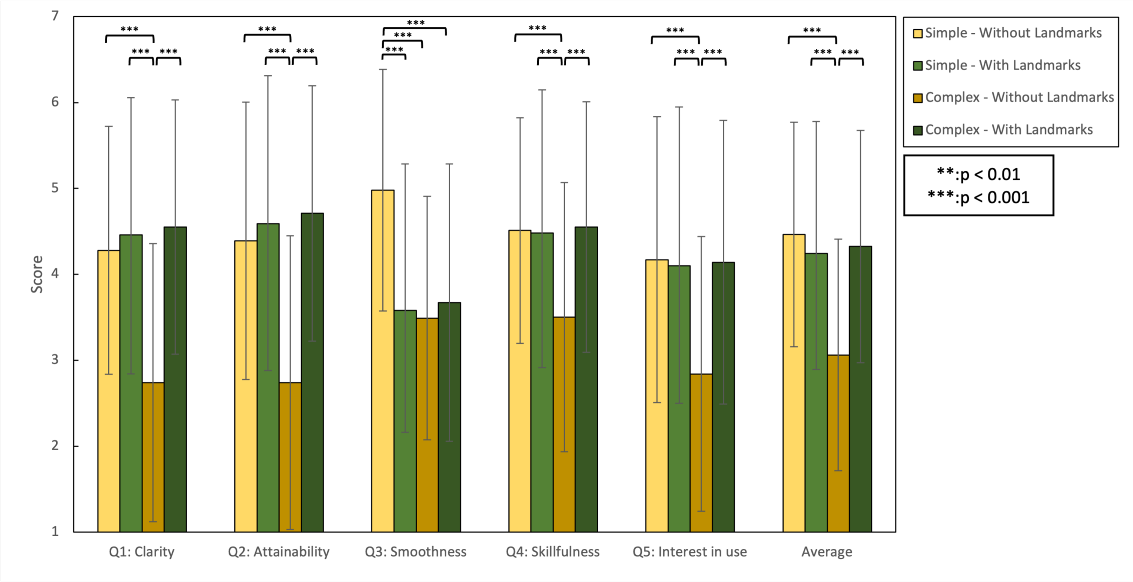
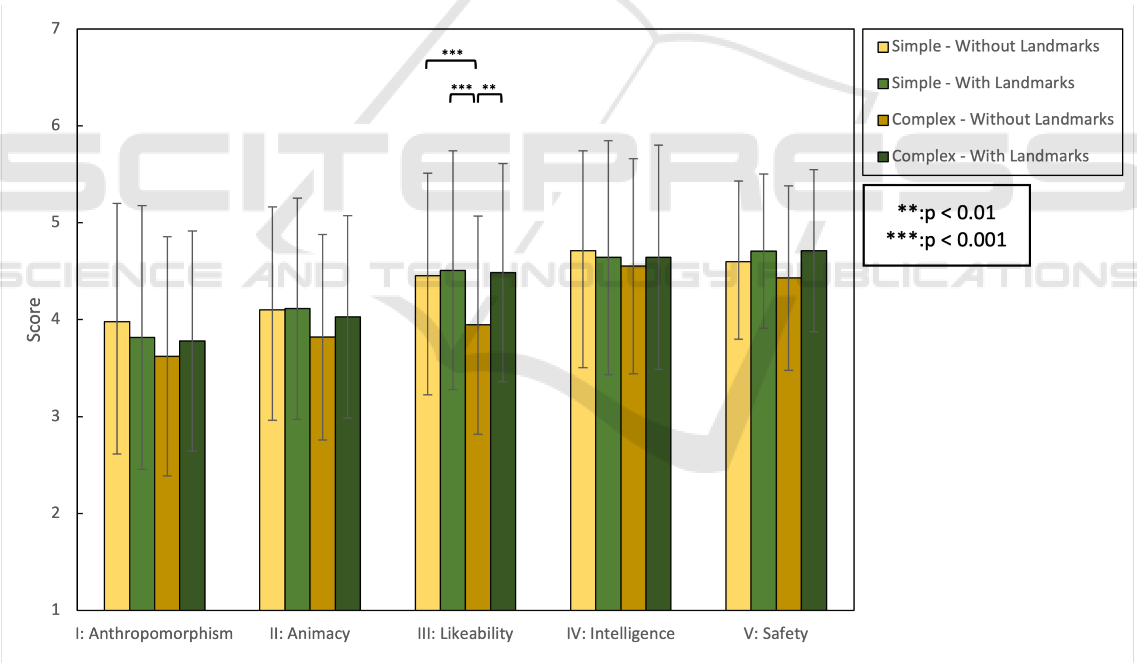


Figure 8: Results of Survey 1. The error bars represent the standard deviations.



SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

Figure 9: Results of Survey 2. The error bars represent the standard deviations.

6.4 Summary

Upon consolidating the results of the three surveys, we concluded that both H1 and H2 were supported. The validation of these hypotheses indicates that, when the route from the current location to the destination is simple, giving route directions from the current location without using landmarks is effective. On the other hand, when the route from the current location is complex, the use of landmarks to give route directions proves effective. RoDiL gives route directions without using landmarks when the route from the current location is simple and with landmarks when the route is complex. Thus, through this experiment, we have demonstrated the efficacy of RoDiL.

Table 3: Result of Survey 3.

		Guide Factor	
		Without landmarks	With landmarks
Scene Factor	Simple	51	49
	Complex	21	79

6.5 Guidelines

RoDiL extracts notable buildings located near the route from the current location to the destination. We hypothesize that in a city with numerous such buildings, RoDiL would be able to identify at least one appropriate landmark regardless of the destination or route chosen. In addition, we presuppose that RoDiL conducts giving route directions within a walking distance of 15 minutes. Hence, we believe that RoDiL would be most significantly effective in urban areas that are densely packed with notable buildings, including, but not limited to, cities like Tokyo and New York.

6.6 Contribution

The key contribution of this study, especially for those involved in HRI studies, is showcasing how interactions with robots can be significantly enhanced through LLMs in specified dialogue areas. One notable area is guiding assistance, which is uniquely suited to robots. Humanoid robots, being capable of active perception and communication, enable people to naturally ask them for directions. For example, using a disembodied computer to seek route directions might be difficult for individuals with declining cognitive abilities, such as the elderly. In essence, the physical presence of a robot aids in the natural interaction needed for sharing directions, thereby enhancing communication between humans and machines. This study envisions potential future interaction experiences between humans and robots. RoDiL demonstrates the ability to conduct navigation dialogue through robot interaction and to deliver clear routes. Consequently, this study offers valuable insights for developing voice dialogue-based HRI methods.

7 CASE STUDY

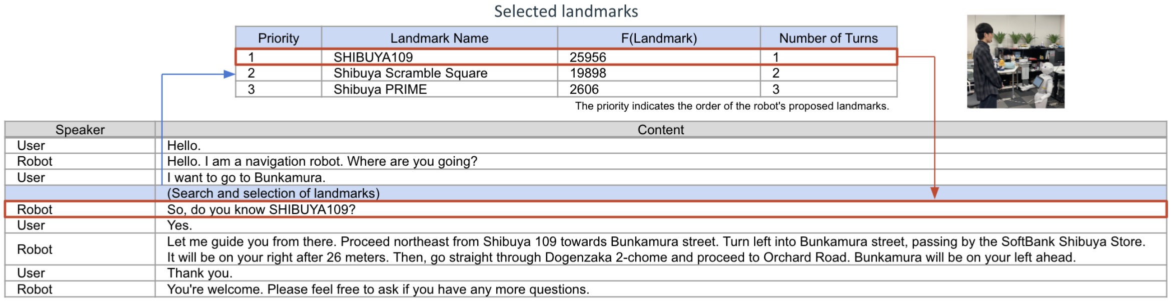
We also conducted a case study to investigate how users would actually use RoDiL. We set the current location to the south exit of Shibuya Station in Shibuya Ward, Tokyo, although we again conducted the case study in a laboratory. We used Pepper again, and the participants conversed with Pepper through voice recognition. Three participants took part in this case study. They were first instructed to choose a destination from the Tower Records Shibuya store, Bunkamura, or Yoyogi Park. After making their choice, they began interacting with the robot. During the interaction, the participants spoke to the robot, which used RoDiL, and asked for route directions to their chosen destination. These interactions were conducted in Japanese.

Figure 10 shows an example from the case study results. The participant first greeted the robot and then asked for directions to Bunkamura. Upon detecting the participant's request, RoDiL searched for a route from the south exit of Shibuya Station to Bunkamura and for landmark candidates along the way. As a result, it selected the landmarks shown in Figure 10 as appropriate. The robot suggested SHIBUYA109, the landmark with the highest priority, to the participant. The participant responded with "directly familiar," which prompted the robot to begin giving route directions without engaging in small talk. The interaction concluded when the robot had finished explaining the directions and the participant had expressed his gratitude.

8 FUTURE WORK

There are several areas in this study that still require further exploration. The directions currently use absolute terms like north and south because of reliance on Google Maps. To enhance user comprehension, it would be better to use relative directions like left and right. Also, only three participants interacted with the robot during this study. It would be beneficial to include more participants in a case study to gather additional feedback and perspectives. Although the experiments took place in a controlled laboratory environment, it is worth considering conducting field studies to test the robot in real-world settings. To suggest more appropriate landmarks, refining the evaluation function is essential. Presently, we recommend a maximum of three routes with the highest evaluation function scores due to computational constraints, which might overlook other possible solutions. This is an area that needs more research. The concept of

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS



Selected landmarks

Priority	Landmark Name	F(Landmark)	Number of Turns
1	SHIBUYA109	25956	1
2	Shibuya Scramble Square	19898	2
3	Shibuya PRIME	2606	3

The priority indicates the order of the robot's proposed landmarks.

Speaker	Content
User	Hello.
Robot	Hello. I am a navigation robot. Where are you going?
User	I want to go to Bunkamura.
	(Search and selection of landmarks)
Robot	So, do you know SHIBUYA109?
User	Yes.
Robot	Let me guide you from there. Proceed northeast from Shibuya 109 towards Bunkamura street. Turn left into Bunkamura street, passing by the SoftBank Shibuya Store. It will be on your right after 26 meters. Then, go straight through Dogenzaka 2-chome and proceed to Orchard Road. Bunkamura will be on your left ahead.
User	Thank you.
Robot	You're welcome. Please feel free to ask if you have any more questions.

Figure 10: Example interaction from the case study.

"indirectly familiar" has not yet been thoroughly investigated for practical use. We need more experiments to understand to what extent the small talks generated by "indirectly familiar" responses can help users feel a sense of familiarity and connection.

9 CONCLUSION

In this study, we proposed a robot guiding system called RoDiL (Route Directions with Landmarks) that gives route directions using landmarks to provide a simpler route when the route from the user's current location to the destination is complex. RoDiL explores the user's knowledge of a city, and utilizing familiar locations as landmarks for the user. Experimental results have substantiated the effectiveness of RoDiL.

This is the first study to demonstrate a significant reduction in the quantity of guidance instructions by utilizing landmarks located along a route as a starting point. In addition, this study demonstrated that robots could acquire users' knowledge through interaction, even in the absence of pre-existing users' knowledge. This result is likely to make a significant contribution to the generation of dialogue in HRI. There are three possible future works. The first involves improving the landmark selection mechanism, especially the fitting of the evaluation function. The second involves investigating the usefulness of a system that decides whether to incorporate small talk according to the user's responses. The third future work entails evaluating RoDiL's usability by having users interact with a robot using RoDiL in a real-world setting.

ACKNOWLEDGEMENTS

This study was supported by JST CREST Grant Number JPMJCR19A1, Japan.

REFERENCES

Bartneck, C., Kulić, D., Croft, E., and Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1:71–81.

Billing, E., Rosén, J., and Lamb, M. (2023). Language models for human- robot interaction. In *HRI '23: Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pages 905–906.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, pages 1877–1901.

Heikkilä, P., Lammi, H., Niemelä, M., Belhassein, K., Sarthou, G., Tammela, A., Clodic, A., and Alami, R. (2019). Should a robot guide like a human? a qualitative four-phase study of a shopping mall robot. In *International Conference on Social Robotics*, pages 548–557.

Kanda, T., Shiomi, M., Miyashita, Z., Ishiguro, H., and Hagita, N. (2010). A communication robot in a shopping mall. *IEEE Transactions on Robotics*, 26(5):897–913.

Kato, S., Ohshiro, S., Itoh, H., and Kitamura, K. (2004). Development of a communication robot ifbot. In *2004 IEEE International Conference on Robotics and Automation (ICRA)*, pages 697–702.

Kim, A., Kum, H., Roh, O., You, S., and Lee, S. (2012). Robot gesture and user acceptance of information in human-robot interaction. In *HRI '12: Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 279–280.

Kim, Y. and Yoon, W. C. (2014). Generating task-oriented interactions of service robots. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(8):981–994.

Lekova, A., Tsvetkova, P., Tanev, T., Mitrouchev, P., and

SCIENCE AND TECHNOLOGY PUBLICATIONS

Kostova, S. (2022). Making humanoid robots teaching assistants by using natural language processing (nlp) cloud-based services. *Journal of Mechatronics and Artificial Intelligence in Engineering*, 3(1):30–39.

Luna, K. L., Palacios, E. R., and Marin, A. (2018). A fuzzy speed controller for a guide robot using an hri approach. *IEEE Latin America Transactions*, 16(8):2102–2107.

Morales, Y., Satake, S., Kanda, T., and Hagita, N. (2014). Building a model of the environment from a route perspective for human–robot interaction. *International Journal of Social Robotics*, 7:165–181.

Okuno, Y., Kanda, T., Imai, M., Ishiguro, H., and Hagita, N. (2009). Rproviding route directions: Design of robot's utterance, gesture, and timing. In *HRI '09: Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*, pages 53–60.

OpenAI (2023). *GPT-4 Technical Report*.

Oßwald, S., Kretzschmar, H., Burgard, W., and Stachniss, C. (2014). Learning to give route directions from human demonstrations. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3303–3308.

Pitsch, K. and Wrede, S. (2014). When a robot orients visitors to an exhibit. referential practices and interactional dynamics in real world hri. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 36–42.

Radford, A., Kim, J. W., amd A. Ramesh, C. H., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, pages 8748–8763.

Radford, A., Narasimhan, K., Salimans, T., and Sutskever, I. (2018). *Improving Language Understanding by Generative Pre-Training*.

Richter, K. (2008). Context-specific route directions. *KI*, 22:39–40.

Richter, K. and Klippel, A. (2005). A model for context-specific route directions. In *Spatial Cognition IV. Reasoning, Action, Interaction*, pages 58–78.

Rosenthal, S., Vichivanives, P., and Carter, E. (2022). The impact of route descriptions on human expectations for robot navigation. *ACM Transactions on Human-Robot Interactions*, 11(35):1–19.

Russo, D., Zlatanova, S., and Clementini, E. (2014). Route directions generation using visible landmarks. In *ISA '14: Proceedings of the 6th ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness*, pages 1–8.

Salem, M., Rohlfing, K., Kopp, S., and Joublin, F. (2011). A friendly gesture: Investigating the effect of multimodal robot behavior in human-robot interaction. In *2011 20th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 247–252.

Shah, D., Eysenbach, B., Kahn, G., Rhinehart, N., and Levine, S. (2021). Ving: Learning open-world navigation with visual goals. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 13215–13222.

Shah, D., Osiński, B., Ichter, B., and Levine, S. (2023). Lmnav: Robotic navigation with large pre-trained models of language, vision, and action. In *Proceedings of the 6th Conference on Robot Learning*, pages 492–504.

Triebel, R., Arras, K., Alami, R., Beyer, L., Breuers, S., Chatila, R., Chetouani, M., Cremers, D., Evers, V., Fiore, M., Hung, H., Ramírez, O. A. I., Joosse, M., Khambhaita, H., Kucner, T., Leibe, B., Lilienthal, A. J., Linder, T., Magnusson, M., Okal, B., Palmieri, L., Rafi, U., van Rooij, M., and Zhang, L. (2016). Spencer: A socially aware service robot for passenger guidance and help in busy airports. In *Field and Service Robotics: Results of the 10th International Conference*, pages 607–622.

Waldhart, J., Clodic, A., and Alami, R. (2019). Reasoning on shared visual perspective to improve route directions. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1–8.

Wallén, J. (2008). *The History of the Industrial Robot*. Linköping: Linköping University Electronic Press.

Zhang, H. and Ye, C. (2019). Human-robot interaction for assisted wayfinding of a robotic navigation aid for the blind. In *2019 12th International Conference on Human System Interaction (HSI)*, pages 137–142.

Zhang, Z., Chai, W., and Wang, J. (2023). Mani-gpt: A generative model for interactive robotic manipulation. *Procedia Computer Science*, 226:149–156.

Zheng, J. Y. and Tsuji, S. (1992). Panoramic representation for route recognition by a mobile robot. *International Journal of Computer Vision*, 9:55–76.