Exploring Factors Influencing Hospitality in Mobile Robot Guidance: A Wizard-of-Oz Study with a Teleoperated Humanoid Robot

Ao Guo, Shota Mochizuki, Sanae Yamashita, Saya Nikaido,
Tomoko Isomura, Ryuichiro Higashinaka
Graduate School of Informatics, Nagoya University, Japan
guo.ao.i6@f.mail.nagoya-u.ac.jp, mochizuki.shota@nagoya-u.jp
{yamashita.sanae.w7, nikaido.saya.v2}@s.mail.nagoya-u.ac.jp
{isomura,higashinaka}@i.nagoya-u.ac.jp

Abstract

Developing mobile robots that can provide guidance with high hospitality remains challenging, as it requires the coordination of spoken interaction, physical navigation, and user engagement. To gain insights that contribute to the development of such robots, we conducted a Wizard-of-Oz (WOZ) study using Teleco, a teleoperated humanoid robot, to explore the factors influencing hospitality in mobile robot guidance. Specifically, we enrolled 30 participants as visitors and two trained operators, who teleoperated the Teleco robot to provide mobile guidance to the participants. A total of 120 dialogue sessions were collected, along with evaluations from both the participants and the operators regarding the hospitality of each interaction. To identify the factors that influence hospitality in mobile guidance, we analyzed the collected dialogues from two perspectives: linguistic usage and multimodal robot behaviors. We first clustered system utterances and analyzed the frequency of categories in high- and low-satisfaction dialogues. The results showed that short responses appeared more frequently in high-satisfaction dialogues. Moreover, we observed a general increase in participant satisfaction over successive sessions, along with shifts in linguistic usage, suggesting a mutual adaptation effect between operators and participants. We also conducted a time-series analysis of multimodal robot behaviors to explore behavioral patterns potentially linked to hospitable interactions.

1 Introduction

Recent advancements in mobile robotics and human-robot interaction have enabled the deployment of robots for mobile guidance tasks in various environments (Sharkawy, 2021; Dahiya et al., 2023; Robinson et al., 2023). For example, Vásquez and Matía (2020) and Yuguchi et al. (2022) developed mobile robots capable of object detection, environmental sensing, and emotion expression. Despite these advances, developing mobile robots that can provide guidance with high hospitality remains challenging, as it requires the consideration of spoken interaction, physical navigation, and user engagement.

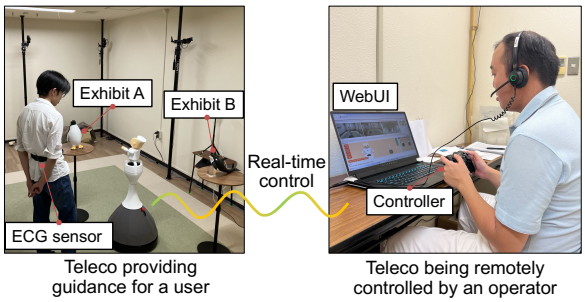

Figure 1: Teleco providing mobile guidance under teleoperation by a remote operator.

To address these challenges, we conducted a Wizard-of-Oz (WOZ) study using Teleco, a teleoperated humanoid robot, to explore the factors influencing hospitality in mobile robot guidance. Specifically, we enrolled 30 participants as visitors to an experimental exhibition space simulating an aquarium-like environment and two trained operators, who teleoperated Teleco to provide mobile guidance to the participants, as shown in Fig. 1. A total of 120 dialogue sessions spanning over ten hours of interaction were collected, along with evaluations from both participants and operators regarding the hospitality of each interaction. Note that, in this study, we define hospitality as the overall quality of the guidance experience, including how welcoming, engaging, and informative the interaction felt.

We analyzed the collected dialogue data from two perspectives: linguistic usage and multimodal robot behaviors. We first clustered system utterances and analyzed the frequency of categories in high- and low-satisfaction dialogues. The results showed that short responses appeared more fre-

Proceedings of the 26th Annual Meeting of SIGDIAL, pages 461–470
Aug 25–27, 2025.

quently in high-satisfaction dialogues. Moreover, we observed a general increase in participant satisfaction over successive sessions, along with shifts in linguistic usage, suggesting a mutual adaptation effect between operators and participants. We also conducted time-series analysis of multimodal behaviors and found that sessions with active engagement from both users and operators, through both speech and nonverbal actions, were associated with higher hospitality compared to sessions dominated by one-sided interactions.

2 Related Work

Mobile robots have been used in a variety of human-interaction scenarios, especially for guiding visitors in museums, aquariums, and other public spaces (Sheridan, 2016; Ajoudani et al., 2018; Rubio et al., 2019). Early work such as MINERVA (Thrun et al., 2000) and RHINO (Burgard et al., 1999) has demonstrated the capability of robots for museum tour guidance using probabilistic navigation methods. Honda's ASIMO (Nakano et al., 2005) further advanced robot mobility, enabling autonomous walking and utilization in receptionist and information guidance roles.

The recently developed Pepper robot is equipped with capabilities for emotional expression and wheeled navigation, targeting customer service (Pandey and Gelin, 2018; Tuomi et al., 2021). Vásquez and Matía (2020) developed Doris, a tour guide robot capable of autonomously navigating predefined waypoints and conveying emotional expressions through dialogue templates. Similarly, Yuguchi et al. (2022) incorporated environmental recognition, enabling autonomous navigation by detecting objects and surroundings in indoor environments. Iio et al. (2020) proposed a human-like guide robot capable of proactively approaching visitors and providing exhibit explanations by combining human-tracking and gaze estimation. More recently, Kondo et al. (2023) introduced a multi-radio Wi-Fi system for teleoperated mobile robots and demonstrated its effectiveness through long-term field deployment.

While these studies have advanced the technical and functional capabilities of mobile guidance robots, little attention has been paid to hospitality in the context of mobile robot interaction. To address this gap, this study investigates factors influencing perceived hospitality during mobile robot guidance through a WOZ experiment.

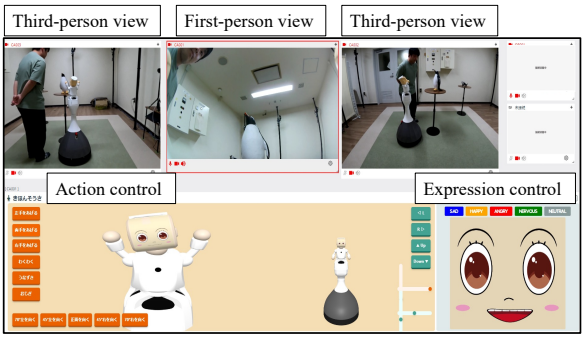

Figure 2: WebUI used for teleoperating the robot. The upper part shows three camera views. The bottom left is the Action Control panel for triggering predefined gestures, and the bottom right is the Expression Control panel for selecting facial expressions.

3 Wizard-of-Oz Data Collection for Mobile Robot Guidance

To investigate the interaction factors that influence the perception of hospitality, we conducted a WOZ experiment using Teleco[1], a humanoid mobile robot equipped with an OLED display and wheeled mobility, as shown in Fig. 1.

The experiment was conducted in an indoor space designed to simulate an aquarium-like exhibition environment, where the robot guided participants by introducing animal figures and providing related explanations. During the experiment, we collected multimodal data and questionnaire responses from both participants and operators to analyze how different interaction factors affected perceived hospitality. The study was approved by our institute's ethics committee and was conducted in accordance with ethical guidelines.

3.1 Robot System

This subsection describes the Teleco robot system used in the experiment, including the teleoperation interface for remote control and the multimodal data collection setup.

3.1.1 Teleoperation Interface

As shown in Fig. 2, the operator utilized a web-based user interface (WebUI) to remotely control Teleco and monitor the interaction. The WebUI displayed both a first-person view from Teleco's chest-mounted camera and third-person views of the environment, allowing the operator to maintain spatial awareness and situational context.

[1]https://www.vstone.co.jp/english/

Using a gamepad-style controller, the operator manually guided Teleco's movements while providing real-time responses to participants. The operator was located in a separate soundproof room, ensuring that the operator's voice was transmitted solely through Teleco. To further enhance the naturalness of the interaction and reduce any sense of dissonance, the operator's voice was processed using Retrieval-based Voice Conversion (RVC)[2] to match Teleco's built-in Text-to-Speech voice.

In addition to controlling movement, the WebUI allowed the operator to trigger a range of predefined nonverbal behaviors, including both physical actions and facial expressions. Specifically, 11 behavioral actions (e.g., raising the left hand) and five facial expressions (happy, sad, angry, nervous, neutral) were available and could be activated manually during interactions. These behaviors were not automatically linked to specific utterances but rather selected in real time by the operator to complement spoken dialogue.

3.1.2 Devices for Multimodal Data Collection

In parallel with teleoperation, multiple devices were utilized to collect multimodal data during the mobile guidance interactions. Three types of cameras were deployed to capture the interactions from different perspectives, as shown in Fig. 2: a first-person view from Teleco's chest-mounted camera, third-person views from GoPro cameras, and full-room views from Sony RX0M2 cameras installed throughout the room. Audio was recorded on the WebUI side, where both the operator's transmitted voice and the participant's voice (via the chest-mounted camera's ambient microphone) were synchronized. This setup ensured that both local and remote audio streams were captured and temporally aligned for analysis.

Participants' heartbeat signals were measured using a BIOPAC system[3], which recorded high-resolution electrocardiography (ECG) signals. We also recorded Teleco's LiDAR data and the operator's control logs obtained from the WebUI. The logs included both triggered behavioral actions and selected emotional expressions.

3.2 Data Collection Procedure

Two trained operators alternated across different days to provide mobile guidance to participants. The overall experimental procedure is illustrated in Fig. 3. Each session began with a pre-experiment questionnaire to collect participants' demographic information, including age, gender, and prior experience with robot interactions.

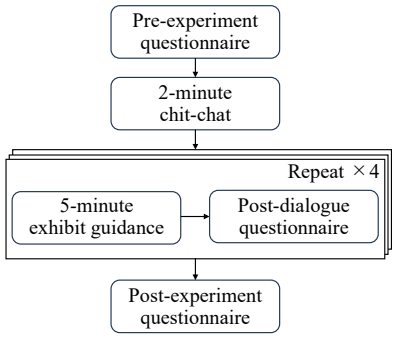

Figure 3: Flowchart of experimental procedure.

The interaction with Teleco comprised two stages: a 2-minute chit-chat session followed by four exhibit guidance sessions, each lasting approximately five minutes. Before the interaction began, participants were instructed to interact naturally with the robot, freely ask questions, and follow its guidance throughout the sessions. The initial chit-chat session was designed to build rapport and reduce initial tension, helping participants become comfortable interacting with the robot before the formal guidance began.

In each guidance session, Teleco introduced two marine animals, resulting in a total of eight animals (e.g., dolphins, orcas, polar bears, and manta rays) presented throughout the experiment. The order of the animals to be presented was randomized for each participant.

After each session, both the participant and the operator completed a post-dialogue questionnaire to assess the perceived hospitality of the interaction. The participant questionnaire consisted of 11 items rated on a 5-point Likert scale, as listed in Table 1. The items were categorized into three key dimensions: (1) subjective experience, including satisfaction, engagement, interest, and perceived familiarity (Q1–Q4); (2) informational quality of the guidance, focusing on the clarity and informativeness of the robot's speech (Q5–Q6); and (3) behavioral and spatial coordination, assessing multimodal synchronization and shared attention between the user and the robot (Q7–Q11). The operator also responded to a corresponding post-dialogue questionnaire, the items of which are presented in Table 2. At the end of the experi-

[2]https://github.com/RVC-Project/Retrieval-based-Voice-Conversion-WebUI/

[3]https://www.biopac.com/product/bionomadix-smart-sys/

ID	User Questionnaire Item
Q1	I was satisfied with the conversation.
Q2	I actively participated in the conversation.
Q3	My interest and curiosity toward the introduced creatures deepened through the conversation.
Q4	I felt a sense of closeness with the robot.
Q5	The robot's speech was informative.
Q6	The robot's guidance was easy to understand.
Q7	The robot maintained an appropriate sense of distance.
Q8	The robot was looking at the same things as me.
Q9	The robot's speech and actions were consistent.
Q10	The robot's speech was appropriate for the situation.
Q11	The robot's actions were appropriate for the situation.

Table 1: User's post-dialogue questionnaire (items translated from Japanese), rated on a 5-point Likert scale.

ID	Operator Questionnaire Item
Q1	I was able to remotely control the robot to effectively engage in conversation.
Q2	The visitor was satisfied with the interaction.
Q3	I felt that my communication was appropriate.
Q4	Through our conversation, I successfully deepened my interest in the visitor's exhibit.
Q5	The robot maintained an appropriate distance.
Q6	The robot was looking at the same exhibit as the visitor.
Q7	My speech matched the robot's actions.
Q8	I was able to operate the robot as intended using the controller.
Q9	The camera feed provided sufficient information for controlling the robot.
Q10	I effectively used commands to control the robot's movements and facial expressions.

Table 2: Operator's post-dialogue questionnaire (items translated from Japanese), rated on a 5-point Likert scale.

ment, participants completed a final questionnaire reflecting on their overall experience.

3.3 Statistics of Collected Data

During the WOZ mobile guidance experiment, we collected 120 guidance dialogues from 30 participants, along with dialogue data from the two operators. All data streams were recorded independently and synchronized using an analog synchronization signal. After data collection, all modalities were temporally aligned, and audio recordings were manually transcribed for subsequent analysis. Table 3 summarizes the basic statistics of the dialogues. On average, both participants (as users) and operators produced a similar number of utterances per dialogue. However, operator utterances were generally longer in both content and duration, resulting in a higher number of spoken seconds per minute. This reflects the operator's role as an active provider of information and responses to participants' questions.

Item	User	Operator
No. of people	30	2
Total utterances	4,870	4,820
Avg. utt. len (words)	8.3±8.0	19.2±13.4
Avg. duration per utt. (sec)	1.5±1.5	3.8±3.1
Avg. utt. per dialogue	40.6	40.2
Speaking sec per minute	11.9	30.1

Table 3: Guidance dialogue statistics.

4 Results and Analysis

To identify factors that influence perceived hospitality in mobile robot guidance, we conducted a multi-perspective analysis focusing on both linguistic usage and multimodal robot behaviors. Our approach was informed by prior studies demonstrating the impact of language patterns on subjective perceptions in human-robot interaction (Yang et al., 2012), as well as the use of time-series analysis to uncover behavioral dynamics in interaction settings (Zhou et al., 2024).

We first examined the participant questionnaire results to evaluate the overall level of perceived hospitality and identify general trends across sessions. We then analyzed the robot's linguistic patterns using embedding-based clustering to identify language features associated with higher hospitality ratings. We further explored multimodal behavioral patterns using time-series clustering of robot behaviors to uncover interaction dynamics linked to high perceptions of hospitality.

4.1 Questionnaire Results

We first evaluate the overall perceived hospitality based on the questionnaire scores and then examine how participants' perceptions evolved across the four guidance sessions.

4.1.1 Evaluation of Perceived Hospitality

We analyzed participant questionnaire responses from 120 guidance sessions to assess perceived hospitality. The means and standard deviations for the 11 evaluated items are summarized in Table 4. Most items received mean ratings above 4.0 (on a 5-point Likert scale), indicating generally favorable impressions of the robot guidance. However, three items (guidance satisfaction, active participation, and guidance informativeness) showed

User Questionnaire Item	Score
Q1. Guidance Satisfaction	3.96±0.89
Q2. Active Participation	3.91±0.91
Q3. Interest in Exhibit	4.24±0.81
Q4. Familiarity with Robot	4.13±1.09
Q5. Guidance Informativeness	3.98±1.04
Q6. Guidance Clarity	4.06±0.88
Q7. Distance Appropriateness	4.18±0.93
Q8. Shared Attention on Exhibit	4.47±0.76
Q9. Speech–Action Consistency	4.33±0.71
Q10. Speech Appropriateness	4.20±0.84
Q11. Action Appropriateness	4.19±0.89

Table 4: User questionnaire statistics. Scores with underlines indicate mean values below 4.0.

Operator Questionnaire Item	Score
Q1. Guidance Satisfaction	3.33±0.99
Q2. Active Participation	2.74±1.28
Q3. Interest in Exhibit	3.47±0.93
Q4. Familiarity with Robot	3.20±1.10
Q5. Guidance Informativeness	3.22±1.27
Q6. Guidance Clarity	4.23±0.93
Q7. Distance Appropriateness	3.52±0.95
Q8. Shared Attention on Exhibit	3.13±1.12
Q9. Speech–Action Consistency	3.38±1.31
Q10. Speech Appropriateness	2.87±1.14

Table 5: Statistics of operator questionnaire. Scores with underlines indicate mean values below 3.0.

Dialogue Excerpt with Guidance Satisfaction of 2	
Speaker	**Utterance [Gesture/Emotion]**
Operator	Penguins eat fish, squid, and ocean creatures.
Operator	Its exploration depth can increase to over 200 meters. [*Emotion: Happy*]
User	Why so deep?
Operator	I wonder why. Sorry, I'm not sure about that. [*Emotion: Sad*]
User	I see. Huh.
Operator	I'll look it up by next time.
User	Please do.

Dialogue Excerpt with Guidance Satisfaction of 5	
Speaker	**Utterance [Gesture/Emotion]**
Operator	It's called a spotted eagle ray.
User	It has spotted patterns, right?
Operator	Yeah, exactly. Take a look at the back; the tail is really long, right? [*Emotion: Happy*]
User	Yeah, it's long.
Operator	That is the venomous part. [*Emotion: Neutral*]
User	Oh, I see.
Operator	So, make sure not to touch it. [*Gesture: Raise left hand*]

Table 6: Dialogue excerpts with high and low guidance satisfaction (translated from Japanese).

mean scores below 4.0, suggesting potential areas for improvement. In contrast to participant responses, operators generally gave lower hospitality ratings to the same sessions, as shown in Table 5. This suggests that they were more self-critical about the quality of their guidance and perceived greater room for improvement in their interactions. Dialogue excerpts representing high and low levels of guidance satisfaction are presented in Table 6. Specifically, the operator in the low-satisfaction excerpt failed to address the user's question, whereas in the high-satisfaction excerpt, the operator provided clear answers, effective interaction, and safety-focused guidance.

4.1.2 Trends in Perceived Hospitality Across Sessions

Since each participant engaged in four guidance sessions, we calculated the average score for each item by session to examine potential changes over time, as shown in Table 7.

We found that the minimum scores for most items tended to appear in the earlier sessions, while the maximum scores were more frequently observed in the later sessions. This indicates a general upward trend in perceived hospitality as the sessions progressed. A similar trend was also observed when comparing sessions guided by operators in their early and late stages, with the improvement more evident in the latter half through gaining experience, supporting the overall pattern of increased perceived hospitality. We presume that this improvement is due to gradual adaptation between participants and operators. Participants likely adjusted their expectations and questioning behavior as they became more familiar with the robot's capabilities, while operators refined their interaction style based on prior sessions to better accommodate user needs. Our linguistic analysis also indicated that operators increasingly used short responses in later sessions, whereas in earlier sessions they preferred using more confirmation. This shift supports the hypothesis of gradual adaptation, as operators adjusted their speaking strategies in response to users' engagement over time. Further details are described in Section 4.2.2.

An exception to this trend was Q5 (Guidance Informativeness), which received the highest score in the first session but gradually declined in the following sessions. One possible reason is that participants became more familiar with the guidance content over time, which may have reduced their perception of informativeness in later sessions.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	**Avg.**
1st Guidance	3.93	3.43	4.10	4.13	**4.17**	3.93	4.07	4.43	4.17	4.10	4.13	4.05
2nd Guidance	3.80	3.97	4.23	4.00	3.97	3.97	4.17	4.47	**4.40**	4.10	4.07	4.10
3rd Guidance	3.83	4.00	4.30	4.17	3.73	4.10	4.20	4.47	4.37	4.13	4.30	4.15
4th Guidance	**4.27**	**4.23**	**4.33**	**4.23**	4.07	**4.23**	**4.30**	**4.53**	**4.40**	**4.47**	**4.27**	4.30

Bold values indicate the highest score among the four turns of guidance, while underlined values indicate the lowest score.

Table 7: Average scores of user questionnaire item per guidance session.

User Utterance Categories	
Meaning	**Representative Utterance**
Formal Affirmation	Yes.
Casual Agreement	Yeah.
Hesitant Response	Hmm. Yeah.
Question Asking	What are penguins' natural enemies?
Operator Utterance Categories	
Meaning	**Representative Utterance**
Fact Statement	It is said that they can dive more than 200 meters.
Exclamation	Amazing, right?
Confirmation	Yeah, that's right.
Prompting	Do you have anything you're curious about?
Transition	Now, let's move to the next animal.
Short Response	Yes.
Explanation	Yes, that's right. There are many types of penguins.
Apology	Sorry, I don't know that much about it.
Description	Basically, they feed on fish, but some aggressive orcas also target seals, dolphins, and even whales.

Table 8: Cluster meanings generated by GPT-4o and representative utterances (translated from the original Japanese).

4.2 Linguistic Analysis

We conducted two types of linguistic analysis on transcribed dialogue data to investigate how interaction patterns relate to perceived hospitality. We first clustered utterances to identify distinct utterance categories, and analyzed how the frequency of these categories correlated with users' hospitality ratings. We then examined temporal shifts in operators' language use by comparing utterance categories across early and late sessions.

4.2.1 Utterance Clustering Using Sentence Embeddings

To explore the typical utterance categories used during mobile robot guidance, we conducted separate clustering analyses on user and operator utterances from manually transcribed dialogue data. Each utterance was first embedded using the Japanese Sentence-BERT model[4]. We then applied K-means clustering using Euclidean distance to the Sentence-BERT embeddings of the utterances. The number of clusters was selected in the range of 2 to 10, and the optimal value was determined using the silhouette score (Shahapure and Nicholas, 2020). For each cluster, we extracted five utterances closest to the centroid as representative examples. We provided GPT-4o with the representative utterances from all clusters at once and instructed it to generate a concise description for each cluster.

The results of utterance clustering and their descriptions are summarized in Table 8. User utterances were clustered into four categories (formal affirmation, casual agreement, hesitant response, and question asking), and operator utterances were clustered into nine categories (fact statement, exclamation, confirmation, prompting, transition, short response, explanation, apology, and description). Note that the transition category, which involves prompting users to move to the next exhibit, is specific to the context of mobile robot guidance.

To examine how specific utterance clusters relate to perceived hospitality, we conducted a cluster proportion analysis. Specifically, we divided the dataset into two groups based on the median value of satisfaction ratings: one group with high satisfaction (60 dialogues) and another with low satisfaction (60 dialogues). For each group, we calculated the proportion of utterances in each category separately for users and operators. We then compared these proportions between the two groups using two-proportion z-tests. The Benjamini–Hochberg procedure (FDR-BH) with $\alpha = 0.05$ was applied to control the false discovery rate across multiple comparisons.

The results in Table 9 revealed that, compared to the low satisfaction group, users in the high satisfaction group exhibited significantly lower pro-

[4]https://huggingface.co/sonoisa/sentence-bert-base-ja-mean-tokens-v2

Proportion of User Utterance Categories			
Meaning	**High Sat.**	**Low Sat.**	**Sig.**
Formal Affirmation	6.4%	**9.9%**	**
Casual Agreement	**29.9%**	24.5%	**
Hesitant Response	**38.4%**	36.3%	
Question Asking	25.3%	**29.4%**	**
Proportion of Operator Utterance Categories			
Meaning	**High Sat.**	**Low Sat.**	**Sig.**
Fact Statement	10.0%	**12.0%**	
Exclamation	**5.4%**	4.6%	
Confirmation	**14.5%**	13.6%	
Prompting	**9.0%**	7.6%	
Transition	7.7%	**9.5%**	
Short Response	**7.4%**	5.4%	*
Explanation	**12.7%**	13.6%	
Apology	**15.7%**	14.8%	
Description	17.6%	**18.8%**	

* $p < .05$, ** $p < .01$ (FDR-BH corrected)

Table 9: Proportions of utterance categories for user and operator utterances across high and low satisfaction groups. Bold values indicate the higher proportion between the two groups for each category.

portions of formal affirmation and question asking, but a higher proportion of casual agreement. Interestingly, the low satisfaction group showed more frequent question asking. This may suggest that participants in these sessions felt less informed, or that their questions were not sufficiently addressed, leading to reduced satisfaction.

On the operator side, dialogues rated with high satisfaction contained a significantly higher proportion of short response utterances. Upon closer examination of the surrounding context, we found that these short responses frequently occurred in reaction to users' casual questions, such as "There's a penguin in the back, right?" or "It's insanely heavy. Huh? Twenty tons?" While user questions were associated with lower satisfaction, these examples suggest that when operators responded in a casual and empathetic manner using short responses (e.g., "Yeah" or "That's right"), it may help offset the negative impact of user questioning. Such responses can make users feel heard and understood, contributing to a more hospitable experience.

These findings suggest that more casual and empathetic communication styles are associated with higher perceptions of hospitality in mobile robot guidance. In contrast, none of the other utterance categories showed significant differences between groups, perhaps because they are perceived as routine components of the guidance task rather than elements that directly shape users' social or emotional impressions.

4.2.2 Differences in Operator Utterance Categories Across Early and Late Sessions

As reported in Section 4.1.2, participants' perceived hospitality ratings tended to increase over the course of the sessions, with this trend particularly evident during the operator's later stages. To better understand the linguistic changes accompanying this trend, we analyzed how operators' use of utterance categories varied across different session stages.

First, we divided the sessions into four groups by combining operator and user session stages. Two levels were defined for each role: "Operator-early" and "Operator-late" for operators, and "User-12" (first and second sessions) and "User-34" (third and fourth sessions) for users. Here, "Operator-early" refers to the first half of sessions conducted by each operator, and "Operator-late" to the latter half. This resulted in four combinations, each containing 30 dialogue sessions. These four groups were then utilized for cross-group comparisons of utterance category proportions.

To identify significant distributional shifts in utterance category usage, we conducted pairwise Fisher's exact tests across all group combinations. This allowed us to determine which utterance categories became significantly more or less prevalent between different session stages. The False Discovery Rate was controlled using the FDR-BH procedure (with $\alpha = 0.05$).

Table 10 summarizes significant differences in utterance category proportions across session groups. Figure 4 shows the significant shifts in utterance categories across different combinations of user stage (Session 12 vs. Session 34) and operator stage (Early vs. Late), as derived from the data in Table 10.

In Fig. 4, three trends can be observed: (1) Compared to the operator's early-stage sessions (Early_12 and Early_34), short responses were used more frequently during the late stage (Late_12). (2) The use of exclamatory utterances decreased when operators guided users in later sessions (Early_34 and Late_34) compared to their initial sessions (Early_12), suggesting that operators became less reliant on emotionally expressive responses as they became more familiar with

Comparison	Category	Proportions
Early_12 vs. Early_34	Exclamation	**6.4**% vs. 3.9%
Early_12 vs. Late_12	Short Response	4.8% vs. **7.5**%
	Explanation	10.9% vs. **14.5**%
Early_12 vs. Late_34	Exclamation	**6.4**% vs. 4.1%
	Confirmation	**16.6**% vs. 11.7%
Early_34 vs. Late_12	Short Response	4.4% vs. **7.5**%

Table 10: Statistically significant differences in utterance category proportions between pairs of user–operator groups. Each row shows a group comparison where a specific utterance category exhibited a significant shift in usage distribution.

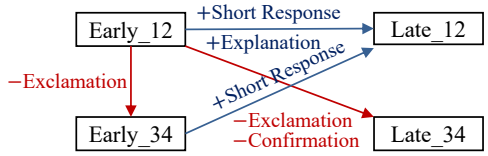

Figure 4: Utterance category transition map showing significant shifts across combinations of user (session 12 vs. session 34) and operator (early vs. late) stages.

the users. (3) When guiding familiar users (Session 34), operators in the late stage (Late_34) tended to use fewer confirmation utterances than in their early stage (Early_12), as their experience increased.

We presume that as operators gained experience and became more familiar with users, they tended to use more short responses and fewer exclamatory and confirmatory utterances. This shift reflects a transition toward more natural and efficient communication, which likely enhanced user comfort and interaction fluency, contributing to improved perceptions of hospitality.

4.3 Multimodal Behavioral Pattern Analysis

To explore how interaction behaviors unfolded during a session, we analyzed the session data from a time-series perspective, aiming to identify multimodal behavioral patterns and their relationship with perceived hospitality.

Following the approach by Zhou et al. (2024), we divided each five-minute session into consecutive five-second intervals. For each interval, we extracted three binary features: whether the user spoke, whether the operator spoke, and whether the operator performed a nonverbal behavior through the WebUI (including control of actions and facial expressions). Each interval was categorized into one of eight (2^3) possible speech

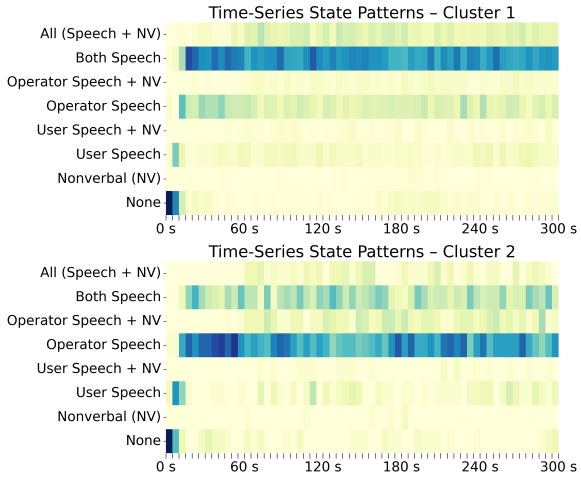

Figure 5: Multimodal behavioral time-series patterns. "NV" refers to the operator's nonverbal behavior, which includes both physical actions and expressions of emotion.

User Questionnaire Item	CL1	CL2	Sig.
Q1. Guidance Satisfaction	**4.1**	3.5	*
Q2. Active Participation	**4.0**	3.6	
Q3. Interest in Exhibit	**4.3**	3.9	
Q4. Familiarity with Robot	**4.3**	3.7	*
Q5. Guidance Informativeness	**4.2**	3.4	*
Q6. Guidance Clarity	**4.2**	3.7	*
Q7. Distance Appropriateness	**4.2**	4.1	
Q8. Shared Attention on Exhibit	**4.5**	4.4	
Q9. Speech–Action Consistency	**4.4**	4.2	
Q10. Speech Appropriateness	**4.2**	4.1	
Q11. Action Appropriateness	4.2	4.2	

* $p < .05$, ** $p < .01$ (FDR-BH corrected)

Table 11: Comparison of user questionnaire item scores between Cluster 1 (CL1) and Cluster 2 (CL2). Bold values indicate the higher mean score for each item.

and nonverbal behavior states based on the combination of these features. By concatenating the sequence of states across all intervals, we generated a time-series representation for each session. We then clustered the time-series data for all 120 sessions. K-modes clustering (Chaturvedi et al., 2001) was applied with the number of clusters ranging from 2 to 10. The optimal number was determined by selecting the number of clusters that minimized the average intra-cluster distance.

The clustering analysis resulted in two distinct groups: Cluster 1 (n = 90) and Cluster 2 (n = 30). The corresponding speech and nonverbal time-series heatmaps are shown in Fig. 5, which illustrate the distribution of states across 60 time

intervals for each cluster. For example, the state "None" indicates that neither the user nor the operator spoke during the interval, and the operator did not perform any nonverbal behavior, while "All (Speech+NV)" indicates that both the user and the operator spoke, and the operator also performed nonverbal behavior. Overall, Cluster 1 exhibited a higher frequency of speech from both users and operators, suggesting a more dynamic and interactive communication style. In contrast, Cluster 2 was characterized by operator-dominated speech with minimal user participation, leading to a more one-sided interaction pattern.

We further compared user questionnaire scores between the two clusters. A Mann-Whitney U test was conducted for each item, and the FDR-BH (with $\alpha = 0.05$) was applied across all 11 items to account for multiple comparisons. The results are summarized in Table 11. Significant differences were found in four questionnaire items (Q1 and Q4–Q6), all of which had higher scores in Cluster 1. These findings suggest that sessions in Cluster 1 involved more balanced and interactive communication between the user and the operator compared to the more one-sided interactions observed in Cluster 2, which likely contributed to a higher perception of hospitality.

5 Conclusion and Future Work

In this study, we explored factors contributing to perceived hospitality in mobile robot guidance through a WOZ experiment with a teleoperated Teleco robot. We analyzed 120 dialogue sessions from both linguistic and multimodal behavioral perspectives. The linguistic analysis revealed that short responses, which reflected a casual and empathetic communication style, were associated with higher hospitality ratings. We also observed a general increase in participant satisfaction over successive sessions, accompanied by shifts in operators' linguistic usage. Furthermore, the multimodal behavior analysis identified distinct behavioral patterns, demonstrating that fostering mutual engagement between users and operators, rather than relying on one-sided interactions, played a critical role in enhancing perceptions of hospitality.

This research has several limitations that should be addressed in future work. First, the current categorization of utterances may lack sufficient granularity, as question types were not clearly differentiated between yes/no and open-ended questions. A more detailed categorization of utterance types can be used for analysis. Second, additional analysis methods should be applied, such as sentiment analysis (Wankhade et al., 2022) and turn-taking analysis (Ghilzai and Baloch, 2015), to provide a more comprehensive understanding of the interactive nature in mobile robot guidance. Third, as the current study relied primarily on utterances, a wider range of multimodal data should be incorporated into the analysis, including user–robot spatial positioning, movement patterns during guidance, and fluctuations in participants' ECG signals, to obtain deeper insights into factors influencing perceived hospitality. Fourth, since our experiments were conducted in a controlled environment simulating an aquarium-like setting, future experiments should be extended to real-world settings, such as an actual aquarium, to further validate our findings. Finally, as the WOZ approach may limit the exploration of the robot's true capabilities in mobile guidance, we plan to gradually incorporate automatic functions (e.g., utterance generation, gesture and facial expression generation, and autonomous movement control) into future experiments. We also aim to embed the identified influential factors into the autonomous mobile robot systems to evaluate their practical effectiveness in enhancing hospitality during real-world interactions.

Acknowledgments

This work was supported by JST Moonshot R&D Grant Number JPMJMS2011.

References

Arash Ajoudani, Andrea Maria Zanchettin, Serena Ivaldi, Alin Albu-Schäffer, Kazuhiro Kosuge, and Oussama Khatib. 2018. Progress and prospects of the human–robot collaboration. *Autonomous Robots*, 42:957–975.

Wolfram Burgard, Armin B Cremers, Dieter Fox, Dirk Hähnel, Gerhard Lakemeyer, Dirk Schulz, Walter Steiner, and Sebastian Thrun. 1999. The museum tour-guide robot RHINO. In *Proceedings of Autonome Mobile Systeme 1998*, pages 245–254.

Anil Chaturvedi, Paul E Green, and J Douglas Caroll. 2001. K-modes clustering. *Journal of Classification*, 18:35–55.

Abhinav Dahiya, Alexander M Aroyo, Kerstin Dautenhahn, and Stephen L Smith. 2023. A survey of multi-

agent human–robot interaction systems. *Robotics and Autonomous Systems*, 161:104335–104353.

Shazia Akbar Ghilzai and Mahvish Baloch. 2015. Conversational analysis of turn taking behavior and gender differences in multimodal conversation. *European Academic Research*, 3(9):10100–10116.

Takamasa Iio, Satoru Satake, Takayuki Kanda, Kotaro Hayashi, Florent Ferreri, and Norihiro Hagita. 2020. Human-like guide robot that proactively explains exhibits. *International Journal of Social Robotics*, 12:549–566.

Yoshihisa Kondo, Hiroyuki Yomo, Shogo Nishimura, Akira Utsumi, and Takahiro Miyashita. 2023. A practical implementation of multi-radio Wi-Fi for teleoperated mobile robots. In *Proceedings of the 2023 IEEE International Conference on Omni-layer Intelligent Systems (COINS)*, pages 1–6.

Mikio Nakano, Yuji Hasegawa, Kazuhiro Nakadai, Takahiro Nakamura, Johane Takeuchi, Toyotaka Torii, Hiroshi Tsujino, Naoyuki Kanda, and Hiroshi G Okuno. 2005. A two-layer model for behavior and dialogue planning in conversational service robots. In *Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3329–3335.

Amit Kumar Pandey and Rodolphe Gelin. 2018. Pepper: The first machine of its kind. *IEEE Robotics & Automation Magazine*, 25(3):40–48.

Nicole Robinson, Brendan Tidd, Dylan Campbell, Dana Kulić, and Peter Corke. 2023. Robotic vision for human-robot interaction and collaboration: A survey and systematic review. *ACM Transactions on Human-Robot Interaction*, 12(1):1–66.

Francisco Rubio, Francisco Valero, and Carlos Llopis-Albert. 2019. A review of mobile robots: Concepts, methods, theoretical framework, and applications. *International Journal of Advanced Robotic Systems*, 16(2):1–22.

Ketan Rajshekhar Shahapure and Charles Nicholas. 2020. Cluster quality analysis using silhouette score. In *Proceedings of the 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 747–748.

Abdel-Nasser Sharkawy. 2021. A survey on applications of human-robot interaction. *Sensors & Transducers*, 251(4):19–27.

Thomas B Sheridan. 2016. Human–robot interaction: Status and challenges. *Human Factors*, 58(4):525–532.

Sebastian Thrun, Michael Beetz, Maren Bennewitz, Wolfram Burgard, Armin B Cremers, Frank Dellaert, Dieter Fox, Dirk Haehnel, Chuck Rosenberg, Nicholas Roy, et al. 2000. Probabilistic algorithms and the interactive museum tour-guide robot MINERVA. *The International Journal of Robotics Research*, 19(11):972–999.

Aarni Tuomi, Iis P Tussyadiah, and Paul Hanna. 2021. Spicing up hospitality service encounters: The case of Pepper. *International Journal of Contemporary Hospitality Management*, 33(11):3906–3925.

Biel Piero E Alvarado Vásquez and Fernando Matía. 2020. A tour-guide robot: Moving towards interaction with humans. *Engineering Applications of Artificial Intelligence*, 88:103356–103373.

Mayur Wankhade, Annavarapu Chandra Sekhara Rao, and Chaitanya Kulkarni. 2022. A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55(7):5731–5780.

Zhaojun Yang, Gina-Anne Levow, and Helen Meng. 2012. Predicting user satisfaction in spoken dialog system evaluation with collaborative filtering. *IEEE Journal of Selected Topics in Signal Processing*, 6(8):971–981.

Akishige Yuguchi, Seiya Kawano, Koichiro Yoshino, Carlos Toshinori Ishi, Yasutomo Kawanishi, Yutaka Nakamura, Takashi Minato, Yasuki Saito, and Michihiko Minoh. 2022. Butsukusa: A conversational mobile robot describing its own observations and internal states. In *Proceedings of the 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 1114–1118.

Xulin Zhou, Takuma Ichikawa, and Ryuichiro Higashinaka. 2024. Collecting and analyzing dialogues in a tagline co-writing task. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 3507–3517.