Designing Accessible Robot Communication for Blind People

Mina Huh
University of California, Berkeley
Berkeley, California

Huihan Liu
University of Texas, Austin
Austin, Texas

Albert Yu
University of Texas, Austin
Austin, Texas

Roberto Martin-Martin
University of Texas, Austin
Austin, Texas

Yuke Zhu
University of Texas, Austin
Austin, Texas

Maya Cakmak
University of Washington
Seattle, Washington

Amy Pavel
University of California, Berkeley
Berkeley, California

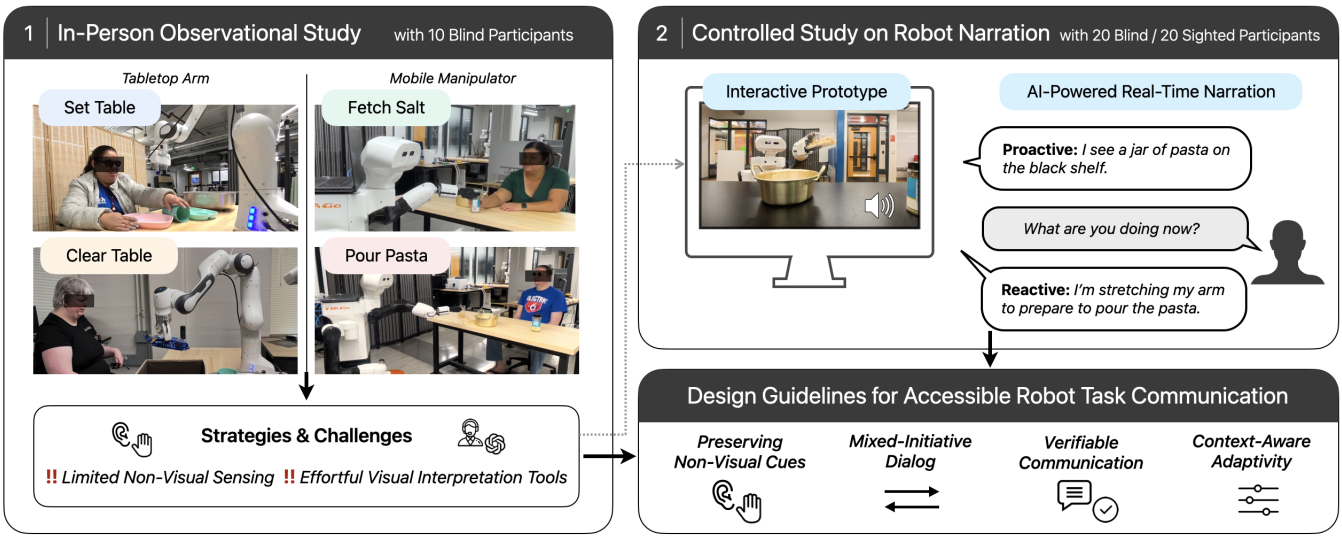

Figure 1: Blind people can not visually monitor the robot, which necessitates new design guidelines for how robots should communicate their progress when executing tasks on their behalf. Our in-person observational study with 10 blind participants reveals that current non-visual cues and visual-interpretation tools fall short. We translate these needs into an interactive robot-narration system in a controlled study with 20 blind and 20 sighted participants, informing design guidelines for accessible robot task communication.

Abstract

Robots are moving into homes with promise to reduce barriers to housework for people with disabilities and decrease effort for everyone. As robots perform tasks autonomously, people need to monitor their progress to verify task execution and intervene when necessary. While sighted users can visually observe the robot, blind users lack timely access to the robot's actions and task outcomes. Our work investigates strategies and challenges that blind people encounter when monitoring robots to inform how to make robot task communication accessible. To understand how blind people monitor robots, we conducted an in-person observational study with 10 blind participants using two robot platforms. Participants primarily used non-visual cues — listening for actions during execution and using touch to inspect the workspace afterward. However, these cues were often ambiguous, leading to missed robot errors and uncertainty about task outcomes. Some participants also turned to visual interpretation tools (*e.g.,* BeMyEyes, Meta Glasses), but these tools produced generic scene descriptions rather than descriptions of the robot's task progress. Participants therefore requested proactive, task-relevant robot narration, and the ability to ask questions. To study these needs at scale, we developed an interactive, AI-powered communication prototype that supports voice-based robot narration and question answering. Our study with 20 blind and 20 sighted participants reveals disparities in question-asking strategies and preferences across groups. We distill design guidelines for accessible robot communication that improves transparency for blind people.

The 3rd InterAI Workshop at CHI 2026, Barcelona, Spain

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

ACM Reference Format:
Mina Huh, Huihan Liu, Albert Yu, Roberto Martin-Martin, Yuke Zhu, Maya Cakmak, and Amy Pavel. 2026. Designing Accessible Robot Communication for Blind People. In *The 3rd InterAI Workshop: "Interactive AI for Human-Centered Robotics at ACM CHI 2026, April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Robots are beginning to move from labs into everyday environments. These robot systems have the potential to save time and effort for everyone and to enable greater independence for people with disabilities [35]. When people use robots to execute tasks (*e.g.*, tidying, washing dishes, cooking), they need to observe and understand the robot's actions to calibrate their expectations, adjust their instructions, and intervene as risky actions or errors occur. Substantial prior work has conveyed robot status and intent through *visual* mechanisms — e.g., motion legibility [17], trajectory shaping, gaze/attention cues [9, 34], facial expressions [14], and on-robot or environmental [11, 47, 52] displays and visualizations. Some robots offer spoken updates [25, 46], but these are often brief status markers (*e.g., "I am starting ...", "Cleaning complete"*) that complement rather than replace visual monitoring. These prior approaches support *sighted people* to understand the robot's actions and act accordingly — but, for *blind people*, robot actions may remain opaque thus denying opportunities for agency over robot behavior. As the use of robots for everyday tasks has immense potential to support blind people [13, 54], blind users should be able to assess task progress, calibrate trust in robot capabilities, detect failures, and confirm outcomes without being able to watch the robot's actions real time. Thus, we ask *how should robots communicate their actions and progress when visual monitoring is unavailable?*

In this work, we study how blind users monitor robots perfoming household tasks then distill design implications to guide accessible robot communication that supports effective monitoring and timely intervention. We first conduct an in-person observational study with 10 blind participants who interact with two robots — a tabletop arm (Panda) and a mobile manipulator (Tiago) — as the robots perform household tasks. Our findings show that blind people (1) actively leverage non-visual senses (sound, touch, smell) and (2) frequently use human- and AI-powered visual interpretation tools (*e.g.,* Meta glasses) to evaluate robot progress and success. While participants often perceived tasks as successful with high confidence, they missed robot errors and formed incorrect accounts of the robots' actions. Motivated by our observations and participant requests for task-relevant verbal robot communication, we developed an two prototype systems for robot communication and conducted a controlled study with 20 blind and 20 sighted participants. We characterize group differences in communication preferences and in how participants assess robot task progress and verbal communication quality.

Through both the observational and controlled study with 50 participants total, our findings offer actionable guidance for designing robot communication that supports non-visual monitoring rather than assuming visual access. We (1) empirically characterize how blind users currently make sense of autonomous household robots during everyday tasks, (2) quantify differences between blind and sighted users in information-seeking and in how they perceive narration quality, and (3) distill design guidelines that help robot builders communicate progress, risks, and failures in ways that are usable without vision. These guidelines inform robot communication that is accessible by design, helping future household robots serve a broader range of users.

2 Related Work

2.1 Assistive and Accessible Robots for Daily Living

Prior work has explored *assistive robots* that support daily living by addressing specific access needs for people with disabilities, including mobility [12, 26, 48], self-care [16], and social well-being [39, 44, 55]. For blind and low vision people in particular, researchers have developed robot systems for tasks in which the primary barrier is limited visual access: non-visual navigation [10, 27, 28, 30, 32, 40] and object localization [24, 38]. These systems guide the person to an object or location with explicit *physical guidance* (*e.g.,* the robot moves to a location as the person holds onto it) or *verbal guidance* (*e.g.,* the robot states the remaining distance to a location) [32]. Such assistive robots demonstrate how robots can improve the independence and safety of disabled people, but they only address constrained, accessibility-specific goals. As robots increasingly move from research labs into public and private spaces for general-purpose use [45], it becomes essential that all robots — not only assistive robots — are *accessible* to diverse users. However, such robots currently pose potential accessibility barriers to blind and low vision users. For example, a blind person instructing a robot to tidy their room can not visually observe if the robot takes an unsafe or inefficient route (*e.g.*, over crumpled clothes), or glance at the physical environment to notice changes (*e.g.*, a shirt placed in the wrong place). But, prior work on assistive robots that guide blind users via physical or verbal directions [10, 27, 28, 30, 32, 40] does not yet explore how an autonomous robot might communicate *its own actions* and *task progress* when acting on behalf of the user. As a step toward accessible robots for daily living, we study how blind users understand robots actions during household tasks, and derive communication considerations to promote *non-visual situational awareness* during task execution.

2.2 Communication and Transparency in HRI

A long history of prior work in Human Robot Interaction (HRI) has explored robot communication and transparency to support users understanding robot's current and future actions [53] as well as the reasoning and confidence behind those actions [18, 50]. Prior work in explainable robotics has examined *what* information users seek [50], *when* explanations are desired [51], and *how* robots can convey more information visually with motion cues [17], eye gaze [9], visual projection [15], and deictic gestures [20, 23]. Other work has augmented robots with sound cues [37] and rich language-based communication [36, 53] to explain robot failures [53] and offer explainations when they detect user confusion [36]. While natural language-based explainations may benefit blind and low vision users, these works presume a sighted observer who can visually ground references and verify outcomes, and whose information needs may differ from users who cannot visually monitor the robot. In contrast, our work investigates blind users' information needs

for robots performing autonomous tasks to derive communication considerations that support non-visual transparency.

3 In-Person Observational Study

To understand how blind people make sense of robotic behaviors, task progress, and outcomes, we conducted an in-person observational study with 10 blind participants. We used a Wizard-of-Oz [41] setup, with a researcher teleoperating the robot to ensure controlled failure injection and participant. Our study investigates the following research questions:

RQ1. *What strategies do blind people use to monitor robots' task progress and outcomes?*

RQ2. *What challenges do blind people encounter when using their current approaches?*

3.1 Method

Participants. We recruited 10 blind participants (P1–P10, Table 2) through a local chapter of the National Federation of the Blind (NFB) and word of mouth. We asked participants to bring any visual assistive technologies they use in daily life (*e.g.,* smart glasses, visual interpretation smartphone applications) so they could use them during the study as desired. The study was approved by our institution's IRB, and participants received $70 for the 2-hour session.

Robots. We used two widely used robot platforms with complementary form factors and capabilities: TIAGo, a mobile manipulator with a wheeled base and dual arms [43], and the Franka Emika Panda, a table-top robotic arm [42]. We selected these platforms to cover common household task demands, and given that robot morphology and noise can shape how people perceive robot behavior [29]. During the study, a researcher teleoperated both robots using a 3Dconnexion SpaceMouse, and participants were not informed of the teleoperation.

Tasks & Procedure. The study took place in a lab space staged with household objects (*e.g.,* a table setting, sink, trash items, pantry items). We began with a brief description of the space and each robot's appearance, size, and capabilities. Following prior work [13], we then allowed participants a brief tactile exploration of the robots while they were stationary. Participants then supervised two robots completing four real-world household tasks (Table 3): (1) setting the table with matching plates and cups, (2) cleaning up the table by sorting dishes and trash, (3) searching for and bringing salt, and (4) bringing pasta and pouring it into a pot. The tabletop Franka performed (1) and (2) while the mobile Tiago performed (3) and (4). The robot provided a brief verbal update at the start and end of each task to convey minimal status updates. In each task, we introduced two scripted failures spanning diverse error types (Table 3) following prior taxonomies of failures in HRI [22], to examine how participants detect, interpret, and respond to common breakdowns.

Participants could use any non-visual cues and any assistive technologies they typically use to verify the robots' task completion. When they wanted to touch the robot or workspace to verify task state, we paused the robot and allowed touch only when it was idle and away from the area to prevent unexpected collisions. Participants were asked to think aloud while completing each task. After each task, they briefly summarized what they believed had happened and completed a short survey assessing perceived task success, understanding of the robot's behavior and outcomes, comfort relying on the robot, and cognitive workload using selected NASA-TLX ratings [21]. After all tasks, we conducted a semi-structured interview to understand participants' strategies and challenges. We video-recorded all sessions.

Figure 2: Blind participants used a variety of strategies to perceive the task and robot status, including technological (meta glasses, AI apps, and remote human assistance) and sensory (smell and touch).

3.2 Results

Non-visual sensing to assess task progresss and outcomes. In all four tasks, all 10 participants reported using sound cues to track when the robot began moving and when made contact with objects (*e.g.,* grasping and placing) (Table 1). Participants reported differences in the usefulness of sound cues between the two robots. For movement, 4 participants (P1, P4, P7, P10) said spatial audio made it easier to infer where the mobile robot was heading, while the tabletop robot's movement direction was less apparent from sound alone. P1 and P4 reported that the distance of the mobile robot still remained hard to tell based on the sound alone such that they felt uncomfortable when they heard the robot moving nearby. For grasping, participants reported that the grasping sounds were more perceptible with the tabletop robot (nearby) than for the mobile robot (often farther away).

Participants also detected failures using sound. All 10 participants noticed pasta spilling in `task4_pour_pasta` because the pasta sounded different when falling into the pot versus onto the table, and 8 participants noticed a cup knocked over in `task1_set_table`. However, other errors were difficult to infer from audio, and sound could also be misleading. P9 misidentified the pepper bottle the robot brought, explaining it *"sounded like salt."* We also observed 3 participants misinterpreting environmental or robot sounds. For example, P9 heard a building fan and assumed the robot was scanning objects. P1 and P5 also misinterpreted the sound of the robot's arm retracting as the robot opening a pasta box with scissors (P1) or picking up spilled pasta (P5).

Participants varied in their ability to extract information from audio cues. P2 (a musician) explained that cups and plates make distinctive sounds and that he could tell which of the two the robot was placing. In contrast, P10 said, *"I have no clue what objects the robot is placing each time. I have to wait until I can touch it after."* P8 added: *"Cups going into the sink make a clear sound, but when some small trash goes into the trash bin, it's not really making a noise."*

Table 1: Approaches blind participants used to monitor and verify robot task execution.

Task	Monitoring and verification approaches (# participants)				
	Auditory cues	Tactile inspection	Smell/taste cues	Human assistance	AI assistance
Set table	10	5	0	2	4
Clear table	10	8	0	1	2
Fetch salt	10	2	4	3	3
Pour pasta	10	7	0	1	2

Takeaway: Blind participants used audio cues to monitor task progress, but audio cues were often absent, ambiguous, or confounded by environmental & robot motion sounds.

Blind participants also used touch to verify task outcomes after the robot finished (Table 1). Particiapnts touched objects to check the layout of the place settings in task1_set_table, confirm sorting in task2_clear_table, identify the bottle in task3_fetch_salt, and locate spilled pasta in task4_pour_pasta. However, touch-based verification was often incomplete. In task1_set_table, P1 missed a cup while searching by hand and became confused about where it went. In task2_clear_table, eight participants swept the table with their hands to confirm nothing was missed, yet 5 of them still overlooked the remaining napkin on the edge of the table.

In task3_fetch_salt, after the robot brought a pepper bottle to the table, P6 tried to locate it by touch but accidentally pushed it to the floor. In task4_pour_pasta, two participants attempted to collect spilled pasta but still missed small pieces left on the table.

Participants also described touch as undesirable in realistic household contexts. In task2_clear_table, 8 participants checked the trash bin and sink to confirm correct sorting, but highlighted this typically would not be feasible, as P10 put it, *"Who would want to put their hand into the trash to check what the robot did? I expect the robot to work for me, not me working for it."* P4 noted safety concerns *"If there was actually a boiling pot of water here, trying to touch around it would be really dangerous."* Beyond touch, some participants used other senses to verify outcomes. In task3_fetch_salt, four participants opened the bottle to smell (and in some cases taste) the spice. Still, P9 emphasized the limitations of these strategies: *"Touching or smelling is slower than hearing. I can only check when the task is complete because I can't interfere while the robot is still moving."*

Takeaway: Blind participants used touch and smell to assess task outcomes, but these approaches are slow, high-effort, and can be undesirable or unsafe in everday contexts.

AI and human-powered assistance for visual interpretation. Seven participants used AI-powered visual interpretation tools such as Be My AI [1] powered by ChatGPT, Seeing AI [4], and Gemini Live [5] to capture photos or stream video and receive descriptions (Figure 2). Four also used Meta Ray-Ban smart glasses [6], which reduced the need to hold and aim a phone. Participants often combined AI with other strategies. For example, using touch first and then confirming with AI (P4), or cross-checking multiple AI tools to increase confidence (P2, P4, P7, P9). In task3_fetch_salt, P7 listened to Gemini and then used Be My AI to double-check that the object the robot brought was salt.

Framing the robot and target objects was difficult when using phone-based AI tools. In task3_fetch_salt, P6 repeatedly tried to capture a label for Be My AI, but the label stayed out of frame. He eventually switched to a barcode reader, which was also slow and effortful. Participants also noted that AI descriptions were often at the wrong level of granularity or required careful prompting. For instance, P4 and P7 asked what the AI saw in task2_clear_table and received broad descriptions (*e.g., a robotic arm on a table with kitchen items*), which necessitated follow-up questions. P7 explained, *"I should ask the right question, otherwise the descriptions are not useful."* Participants further reported inaccuracies and hallucinations as models often relied on a single snapshot or narrow view and thus failed to capture motion or context. P1 noted, *"My Meta Glases said the robot is stationary, when I could clearly hear it moving."* AI tools were also harder to use when the robot was far away, partially occluded, or turned away from the camera.

Four participants used human-powered remote visual interpretation services including Be My Eyes [2] and Aira [3] (Figure 2). These tools allowed participants to share live video of the task over a call and ask questions. Participants proactively provided task context up front to get more relevant descriptions from the agent. In task2_clear_table, P3 used the service to check if the robot had cleared all items on the table and explained *"When [the visual interpreter] said there's nothing on the table, I'm not sure if it's true or whether my camera was not showing it all."* Participants highlighted that the services were often expensive (*e.g.*, $100 for 50 minutes of Aira), unreliable with poor connectivity, and could reduce independence and privacy. P9 noted *"I want the robot to describe things, not them [visual interpreters] so that I can feel more independent."* Thus, 9 participants reported they preferred AI tools over human assistance.

Takeaway: AI tools enable independent verification, but require effort to ask questions and frame the image, then often provide inaccurate or overly generic descriptions.

Task Outcomes.

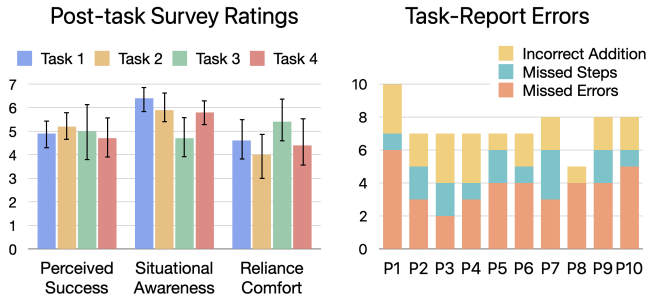

Figure 3: (Left) Average post-task survey ratings for perceived success, situational awareness, and reliance comfort. (Right) Frequency of three types of inaccuracies (missed errors, missed steps, incorrect addition) in participants' task reports.

Post-task survey ratings indicated that participants percieved the robots as moderately successful, while rating their situational awareness as relatively high across all tasks (Figure 3, left). The lower reported situational awareness for `task3_fetch_salt` may be due to the difficulty of verifying the object through hearing or touch. Despite high self-ratings of situational awareness, there were systematic inaccuracies in participant reports of the robot's taks executions as participants missed robot errors, omitted task steps, and reports of events that did not occur (Figure 3, right). Participants had an average of 7.5 inaccuracies (SD=1.27). 51% were missed errors, 30% were incorrect additions, and 20% were missed steps. Incorrect additions were often related to wrong inferences of robot or environment sounds. For example, P6 and P7 interpreted the arm retraction noise as the robot recovering spilled pasta or a knocked cup and perceived the task as successful. P1 mistook fan noise for robot scanning and anticipated it to move towards it while the robot was idle.

These findings show that blind participants often overstated how well they tracked the robot's actions and outcomes. This gap motivates robot communication that enables accurate, non-visual monitoring and verification. When asked about desired support to monitor robots' work, all participants requested rich robot narration, along with the ability to ask questions on demand to confirm details and resolve uncertainty.

4 Controlled Study with an Interactive Robot-Communication Prototype

We next examined a core design choice in robot communication: how much initiative the robot should take versus how much control the user should retain. Specifically, we compared an *reactive* mode (the robot speaks only when asked) with a *mixed-initiative* mode (the robot provides proactive task updates while also answering questions) to test whether combining proactive updates with user-driven queries better supports monitoring than relying on questions alone. Prior work has characterized how sighted users seek task-relevant information from robots [50] and how robots should communicate about failures [31], but it remains unclear how these needs and preferences change when users cannot rely on visual verification. We therefore conducted a controlled online study using an interactive narration prototype (Figure 4) to characterize group differences in information seeking and to evaluate how communication mode shapes monitoring experience and task understanding. We investigate:

- **RQ1.** *How do blind and sighted users differ in information-seeking strategies while monitoring robot tasks?*
- **RQ2.** *How does robot initiative in communication shape monitoring experience and preferences across groups?*
- **RQ3.** *How does communication mode affect objective task understanding during monitoring?*

4.1 Method

Participants. We recruited 20 blind (P11-P30) and 20 sighted (S1-S20) through mailing lists and word of mouth. All blind participants described themselves as fully blind with little to no light perception and used a screen reader to access our web-based prototype.

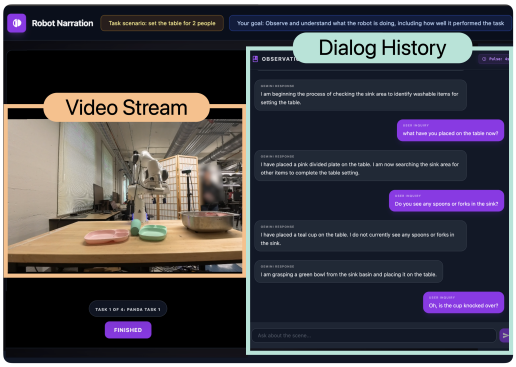

Figure 4: We developed a web-based communication prototype that supports voice interactions while the task video is streaming.

None had participated in our in-person study. The study was IRB-approved, and participants received $30 for a 1-hour session.

Robot Communication Prototype. We developed an AI-powered web-based communication system that supports voice interactions with robots while the task video is streaming. We implemented communication with an automated, grounded pipeline (rather than human-authored narration) to avoid describer-style confounds and to evaluate strategies that could plausibly run on future robots.

Our system supports two modes: reactive, where the system speaks only in response to user questions, and mixed-initiative, where it provides periodic status updates while also answering questions on demand. To generate grounded, time-aligned responses, we collected 3 sources of logs from robots' task execution: (1) the robot's task steps, (2) robot-captured video, and (3) robot internal status (*e.g.,* joint movements) following Wang et al. [53]. During the study, the system uses the video timestamp to align the participant-facing video with the corresponding robot traces (hidden from participants) and feeds this aligned context to a streaming multimodal model to generate spoken responses in real time.

Tasks & Procedure. We used the same 2 robot platforms and 4 household tasks as in §3, and recorded each task from a fixed user viewpoint to approximate how a co-located observer would monitor the robot. Participants watched these recordings during the study and used the prototype to receive robot communication and ask questions in real time. We randomized the order of robot/task videos and communication conditions across participants. After each task, participants completed a comprehension quiz to assess task understanding.

4.2 Results

Group Differences in Information Seeking. Blind participants asked significantly more questions than sighted participants in both conditions ($p < 0.01$). In the reactive condition, blind participants often repeated status-check questions in the absence of proactive narration, such as P33 asking *"what are you doing now?"* four times. Several blind participants also asked about sounds they heard in the video, such as *"What is this sound?"* (P24, P36). However, our system could not answer these questions as audio was not included in the model's input context, highlighting an opportunity for future

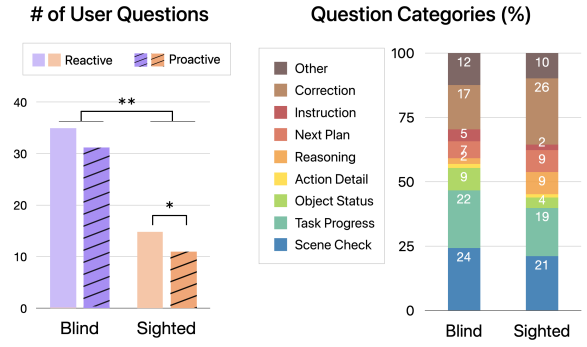

Figure 5: (Left) Mean number of user questions from Blind vs. Sighted participants under *Reactive* vs. *Mixed-initiative* narration. (Right) Distribution of question categories in each group (%). Asterisks denote significance tested with Wilcoxon: * $p < .05$, ** $p < .01$.

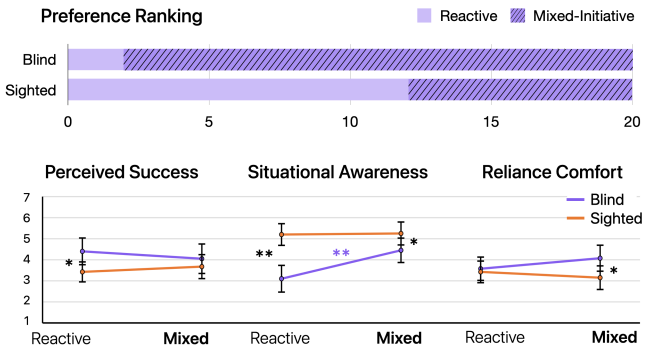

Figure 6: (Top) Preference ranking for *Reactive* vs. *Mixed-Initiative* narration in each group. (Bottom) Group × condition interaction plots (mean with 95% CI error bars). Asterisks denote significance tested with Wilcoxon. * $p < .05$, ** $p < .01$.

systems to incorporate robot microphones for grounding communication. Sighted participants asked significantly fewer questions when proactive narration was available, and they tended to ask more questions for the mobile manipulator tasks, where the robot frequently moved away from the user's viewpoint or faced away while reaching, making key actions harder to observe. Blind participants more frequently asked for scene context (*"What do you see in the kitchen?* – P15, *"Where on the counter did you put the shaker?"* – P28) and object state (*"Are you done setting the table?* – P13, *"Does the jar have a lid on top of it?" – P23*). Sighted participants also asked scene and object-status questions when the mobile manipulator moved away or faced away while reaching, as S3 asked *"Describe the object that you're trying to reach. Is there any label on it?"* Sighted participants asked a higher proportion of questions about the robot's next action (*"What will you do with the spilled pasta?" – S4*) and reasoning (*"Why are you putting the pink cup in the trash bin?"* – S13). Sighted participants also offered more *feedback and corrections* than blind participants (e.g., "*The green cup is oriented incorrectly. Please rotate it.*" – S20), and they also corrected the narration when it mismatched what they observed (*"The jar is full, not half full [as the robot described.]"* – S4).

Experience and Preferences by Communication Mode. The majority of blind participants preferred the mixed-initiative mode that offers proactive narration, while more than half of the sighted participants preferred the reactive mode. Blind participants explained that proactive updates reduced the burden of needing to "know what to ask." As P19 noted, *"When they only answer my question, it relies on me asking good questions. When I don't even know what they are doing, I can't ask good questions."* Similarly, P13 said, *"Without a narration [reactive mode], I felt like I was just kind of asking questions into a void."* Under the mixed-initiative mode, blind participants reported higher situational awareness, noticing more robot errors and perceiving a significantly lower task success rate than in the reactive mode. In contrast, sighted participants preferred the reactive mode. They described proactive narration as duplicating visual information, often repetitive, and sometimes disruptive to interaction. S12 *"heard the same messages read out loud over and over again and didn't know when to interrupt to ask a question,"* and S3 found it *"frustrating to hear the robot repeating its action every 5 seconds."*

Task Understanding and Perceived Communication Quality. Figure 7 summarizes participants' comprehension accuracy and their ratings of communication quality.

Blind participants scored significantly lower than sighted participants on the post-task quizzes ($p < .001$), indicating a gap in objective task understanding. While blind participants reported higher situational awareness in the mixed-initiative mode than the reactive mode (Figure 6), this increase did not translate to higher quiz accuracy. Sighted participants also provided more detailed critiques of robot performance, including safety-relevant issues. For example, S17 noted that *"The robot did not behave safely, it pushed the pot of presumably-boiling water."* Figure 7 also shows clear group differences in perceived communication usefulness and quality. Sighted participants rated the robots' answers as less useful and lower quality, frequently explaining timing and hallucinations as issues. S15 remarked, *"The narrations weren't in sync with robot behaviors,"* and S14 expressed frustration about incorrect narrations: *"Why did you say that you would pick up the teal cup when you were picking up the plate?"*

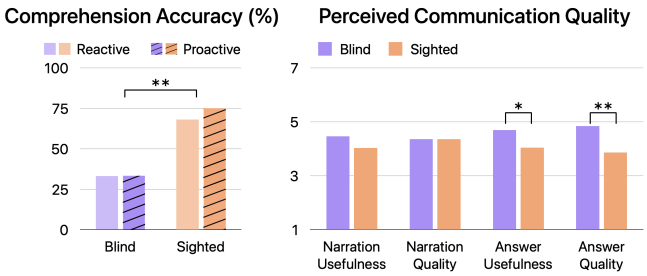

Figure 7: (Left) Comprehension accuracy (%). Sighted participants achieved higher accuracy than blind participants in both conditions. (Right) Mean Likert ratings (1–7) for narration and answer usefulness and quality by group. Blind participants rated answer usefulness and quality higher than sighted participants. Asterisks denote significance tested with Wilcoxon. * $p < .05$, ** $p < .01$.

5 Design Guidelines

We distill design guidelines for accessible robot task communication from our two studies, grounded in blind participants' existing monitoring strategies and information needs. We also draw on general accessibility perspectives that emphasize designing around users' abilities and practiced routines (*e.g.*, ability-based design [56]) and adjacent guidance for non-visual description [19] that emphasizes complementing rather than masking environmental cues.

DG1. Make non-visual sensing reliable and interpretable.
DG2. Provide proactive narration that complements question answering.
DG3. Communicate risks and failures clearly to support correction and recovery.
DG4. Design for appropriate trust in robot narration.
DG5. Adapt communication to context and provide user control.
DG6. Design and evaluate with blind users under ecologically valid conditions.

6 Conclusion

We address a key gap in HRI: supporting blind users' monitoring and verification of robot task execution in everyday tasks. We conducted (1) an in-person study with 10 blind participants to characterize non-visual monitoring strategies and breakdowns, and (2) a controlled study with 20 blind and 20 sighted participants to explore how to design accessible robot narration. We contribute empirically grounded design guidelines for accessible robot task communication. We hope our findings inform robot systems that are safer, more transparent, and more usable for all.

Acknowledgments

We appreciate all participants for their time and valuable feedback. We thank Rutav Shah for his help with the robot hardware.

References

[1] Last visited: 2025. https://www.bemyeyes.com/bme-ai/
[2] Last visited: 2025. https://www.bemyeyes.com/
[3] Last visited: 2025. https://aira.io/
[4] Last visited: 2026. https://www.seeingai.com/
[5] Last visited: 2026. https://gemini.google/overview/gemini-live/
[6] Last visited: 2026. https://www.meta.com/ai-glasses/
[7] Last visited: 2026. https://www.w3.org/WAI/standards-guidelines/
[8] Last visited: 2026. https://firebase.google.com/docs/database
[9] Henny Admoni and Brian Scassellati. 2017. Social eye gaze in human-robot interaction: a review. *Journal of Human-Robot Interaction* 6, 1 (2017), 25–63.
[10] Shiri Azenkot, Catherine Feng, and Maya Cakmak. 2016. Enabling building service robots to guide blind people a participatory design approach. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 3–10.
[11] Kim Baraka, Stephanie Rosenthal, and Manuela Veloso. 2016. Enhancing human understanding of a mobile robot's state and actions using expressive lights. In *2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, 652–657.
[12] Tapomayukh Bhattacharjee, Maria E Cabrera, Anat Caspi, Maya Cakmak, and Siddhartha S Srinivasa. 2019. A community-centered design framework for robot-assisted feeding systems. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*. 482–494.
[13] Mayara Bonani, Raquel Oliveira, Filipa Correia, André Rodrigues, Tiago Guerreiro, and Ana Paiva. 2018. What my eyes can't see, a robot can show me: Exploring the collaboration between blind people and robots. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. 15–27.
[14] Cynthia Breazeal and Brian Scassellati. 1999. How to build robots that make friends and influence people. In *Proceedings 1999 IEEE/RSJ international conference on intelligent robots and systems. Human and environment friendly robots with high intelligence and emotional quotients (cat. No. 99CH36289)*, Vol. 2. IEEE, 858–863.
[15] Ravi Teja Chadalavada, Henrik Andreasson, Robert Krug, and Achim J Lilienthal. 2015. That's on my mind! robot to human intention communication through on-board projection on shared floor space. In *2015 European Conference on Mobile Robots (ECMR)*. IEEE, 1–6.
[16] Tiffany L Chen, Matei Ciocarlie, Steve Cousins, Phillip M Grice, Kelsey Hawkins, Kaijen Hsiao, Charles C Kemp, Chih-Hung King, Daniel A Lazewatsky, Adam E Leeper, et al. 2013. Robots for humanity: using assistive robotics to empower people with disabilities. *IEEE Robotics & Automation Magazine* 20, 1 (2013), 30–39.
[17] Anca D Dragan, Kenton CT Lee, and Siddhartha S Srinivasa. 2013. Legibility and predictability of robot motion. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 301–308.
[18] Connor Esterwood and Lionel P Robert. 2021. Do you still trust me? human-robot trust repair strategies. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE, 183–188.
[19] Louise Fryer. 2016. *An introduction to audio description: A practical guide*. Routledge.
[20] Atmaraaj Gopal, Arihiro Yorita, Naoyuki Kubota, and Matthias Rätsch. 2025. Nmm-hri: Natural multimodal human-robot interaction with voice and deictic posture via large language model. (2025).
[21] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.
[22] Shanee Honig and Tal Oron-Gilad. 2018. Understanding and resolving failures in human-robot interaction: Literature review and model development. *Frontiers in psychology* 9 (2018), 861.
[23] Chien-Ming Huang and Bilge Mutlu. 2013. Modeling and Evaluating Narrative Gestures for Humanlike Robots.. In *Robotics: Science and Systems*, Vol. 2.
[24] Felix Huppert, Gerold Hoelzl, and Matthias Kranz. 2021. GuideCopter-A precise drone-based haptic guidance interface for blind or visually impaired people. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–14.
[25] iRobot. [n. d.]. Roomba Robot. https://www.irobot.com/en_US/roomba.html
[26] Rajat Kumar Jenamani, Tom Silver, Ben Dodson, Shiqin Tong, Anthony Song, Yuting Yang, Ziang Liu, Benjamin Howe, Aimee Whitneck, and Tapomayukh Bhattacharjee. 2025. FEAST: A Flexible Mealtime-Assistance System Towards In-the-Wild Personalization. *arXiv preprint arXiv:2506.14968* (2025).
[27] Rie Kamikubo, Seita Kayukawa, Yuka Kaniwa, Allan Wang, Hernisa Kacorri, Hironobu Takagi, and Chieko Asakawa. 2025. Beyond Omakase: Designing Shared Control for Navigation Robots with Blind People. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–17.
[28] Seita Kayukawa, Daisuke Sato, Masayuki Murata, Tatsuya Ishihara, Akihiro Kosugi, Hironobu Takagi, Shigeo Morishima, and Chieko Asakawa. 2022. How users, facility managers, and bystanders perceive and accept a navigation robot for visually impaired people in public buildings. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 546–553.
[29] Laura Kunold, Nikolai Bock, and Astrid Rosenthal-von der Pütten. 2023. Not all robots are evaluated equally: the impact of morphological features on robots' assessment through capability attributions. *ACM Transactions on Human-Robot Interaction* 12, 1 (2023), 1–31.
[30] Masaki Kuribayashi, Kohei Uehara, Allan Wang, Shigeo Morishima, and Chieko Asakawa. 2025. Wanderguide: Indoor map-less robotic guide for exploration by blind people. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–21.
[31] Gregory LeMasurier, Alvika Gautam, Zhao Han, Jacob W Crandall, and Holly A Yanco. 2024. Reactive or proactive? how robots should explain failures. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. 413–422.
[32] Shuijing Liu, Aamir Hasan, Kaiwen Hong, Runxuan Wang, Peixin Chang, Zachary Mizrachi, Justin Lin, D Livingston McPherson, Wendy A Rogers, and Katherine Driggs-Campbell. 2024. Dragon: A dialogue-based robot for assistive navigation with visual language grounding. *IEEE Robotics and Automation Letters* 9, 4 (2024), 3712–3719.
[33] Meta Platforms, Inc. Last visited: 2026. React: A JavaScript library for building user interfaces. https://react.dev/
[34] Bilge Mutlu, Toshiyuki Shiwa, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2009. Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*. 61–68.
[35] Amal Nanavati, Vinitha Ranganeni, and Maya Cakmak. 2023. Physically assistive robots: A systematic review of mobile and manipulator robots that physically assist people with disabilities. *Annual Review of Control, Robotics, and Autonomous Systems* 7 (2023).
[36] Andreas Naoum, Parag Khanna, Elmira Yadollahi, Mårten Björkman, and Christian Smith. 2025. Adapting robot's explanation for failures based on observed human behavior in human-robot collaboration. *arXiv preprint arXiv:2504.09717* (2025).
[37] Nnamdi Nwagwu, Adeline Schneider, Ibrahim Syed, Brian J Zhang, and Naomi T Fitter. 2024. The benefits of sound resound: an in-person replication of the ability of character-like robot sound to improve perceived social warmth. (2024).

[38] Adil Rahman, Md Aashikur Rahman Azim, and Seongkook Heo. 2023. Take my hand: Automated hand-based spatial guidance for the visually impaired. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.

[39] Rebecca Ramnauth, Dražen Brščić, and Brian Scassellati. 2025. A Robot-Assisted Approach to Small Talk Training for Adults with ASD. *arXiv preprint arXiv:2505.23508* (2025).

[40] Vinitha Ranganeni, Mike Sinclair, Eyal Ofek, Amos Miller, Jonathan Campbell, Andrey Kolobov, and Edward Cutrell. 2023. Exploring levels of control for a navigation assistant for blind travelers. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. 4–12.

[41] Laurel D Riek. 2012. Wizard of oz studies in hri: a systematic review and new reporting guidelines. *Journal of human-robot interaction* 1, 1 (2012), 119–136.

[42] RoboDK. 2026. Franka Emika Panda Robot. https://robodk.com/robot/Franka/Emika-Panda. Accessed: 2026-01-07.

[43] PAL Robotics. 2026. TIAGo – Mobile Manipulator Robot for Research. https://pal-robotics.com/robot/tiago/. Accessed: 2026-01-07.

[44] Brian Scassellati, Henny Admoni, and Maja Matarić. 2012. Robots for use in autism research. *Annual review of biomedical engineering* 14, 1 (2012), 275–294.

[45] Eike Schneiders, Anne Marie Kanstrup, Jesper Kjeldskov, and Mikael B Skov. 2021. Domestic robots and the dream of automation: Understanding human interaction and intervention. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.

[46] Julie Shah and Cynthia Breazeal. 2010. An empirical analysis of team coordination behaviors and action planning with application to human–robot teaming. *Human factors* 52, 2 (2010), 234–245.

[47] Moondeep C Shrestha, Ayano Kobayashi, Tomoya Onishi, Erika Uno, Hayato Yanagawa, Yuta Yokoyama, Mitsuhiro Kamezaki, Alexander Schmitz, and Shigeki Sugano. 2016. Intent communication in navigation through the use of light and screen indicators. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 523–524.

[48] Breelyn Kane Styler, Lesong Jia, Henny Admoni, Reid Simmons, Rory Cooper, Na Du, and Dan Ding. 2025. Evaluating Feedback Modality Preferences of Power Wheelchair Users During Manual Robotic Arm Control. In *2025 International Conference On Rehabilitation Robotics (ICORR)*. IEEE, 620–627.

[49] Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805* (2023).

[50] Lennart Wachowiak, Andrew Coles, Gerard Canal, and Oya Celiktutan. 2025. What Questions Should Robots Be Able to Answer? A Dataset of User Questions for Explainable Robotics. *arXiv preprint arXiv:2510.16435* (2025).

[51] Lennart Wachowiak, Andrew Fenn, Haris Kamran, Andrew Coles, Oya Celiktutan, and Gerard Canal. 2024. When do people want an explanation from a robot?. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. 752–761.

[52] Michael Walker, Hooman Hedayati, Jennifer Lee, and Daniel Szafir. 2018. Communicating robot motion intent with augmented reality. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 316–324.

[53] Zihan Wang, Brian Liang, Varad Dhat, Zander Brumbaugh, Nick Walker, Ranjay Krishna, and Maya Cakmak. 2024. I can tell what i am doing: Toward real-world natural language grounding of robot experiences. *arXiv preprint arXiv:2411.12960* (2024).

[54] Yize Wei, Nathan Rocher, Chitralekha Gupta, Mia Huong Nguyen, Roger Zimmermann, Wei Tsang Ooi, Christophe Jouffrais, and Suranga Nanayakkara. 2025. Human Robot Interaction for Blind and Low Vision People: A Systematic Literature Review. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–19.

[55] Sally Whelan, Kathy Murphy, Eva Barrett, Cheryl Krusche, Adam Santorelli, and Dympna Casey. 2018. Factors affecting the acceptability of social robots by older adults including people with dementia or cognitive impairment: a literature review. *International Journal of Social Robotics* 10, 5 (2018), 643–668.

[56] Jacob O Wobbrock, Shaun K Kane, Krzysztof Z Gajos, Susumu Harada, and Jon Froehlich. 2011. Ability-based design: Concept, principles and examples. *ACM Transactions on Accessible Computing (TACCESS)* 3, 3 (2011), 1–27.

A In-Person Observational Study

A.1 Participants

We recruited 10 blind participants (P1–P10 in Table 2) through a local chapter of the National Federation of the Blind (NFB) and word of mouth. Eligibility included being able to travel to the study location in person. We ensured accessible commute support with local access door-to-door services.

PID	Gender	Age	Vision	Onset
P1	Female	38	Totally Blind	Congenital
P2	Male	53	Totally Blind	Acquired
P3	Female	38	Totally Blind	Congenital
P4	Female	40	Totally Blind	Congenital
P5	Male	47	Totally Blind	Congenital
P6	Male	46	Totally Blind	Congenital
P7	Female	44	Totally Blind	Congenital
P8	Female	37	Totally Blind	Acquired
P9	Female	30	Totally Blind	Congenital
P10	Male	41	Totally Blind	Acquired

Table 2: Demographics of blind participants in the in-person observational study

A.2 4 Household Tasks

We designed 4 everyday household tasks: two with a tabletop manipulator and two with a mobile manipulator. Each was presented with a short scenario to provide participants with a concrete goal and expectations for the outcome. During the study, a researcher teleoperated the robot following an execution script for each task. We also introduced 2 planned errors at specific moments for each task, to explore how blind participants identify and react to these errors.

Task 1: Set the Table

Scenario. Your friend is coming over, and you want to set the table with matching plates and cups. You ask the robot to help place the items neatly on the table. As the task runs, pay attention to what the robot seems to be doing and whether the final setup matches your expectations.

Steps.

(1) Two dishes (pink, teal) and two cups (pink, teal) are on the sink on the table; nothing else is on the table.
(2) The robot picks up a pink plate and places it on the table.
(3) The robot picks up a teal cup and places it next to the pink plate.
(4) The robot picks up a teal plate and places it on the table (it contacts the teal cup during placement).
(5) The robot picks up a pink cup and places it next to the teal plate.

Injected errors.

- Misinterpretation error: The robot places two place settings with mismatched colors (the teal cup is placed next to the pink plate).
- Control error: While placing the second plate, the robot knocks over the nearby teal cup and does not pick it up.

Task 2: Clear the Table

Scenario. After your friend leaves, the table is cluttered with cups, bowls, and snacks. You ask the robot to clear the table so you can use it for work. There is a sink and a trash bin on the table. The robot should put reusable items into the sink and trash into the bin. Notice what the robot moves, what it leaves behind, and how confident you feel about the outcome.

Steps.

(1) Two plastic cups (pink, teal), one purple bowl, one empty Oreo package, one crumpled paper towel, and one leftover bread are laid out on the table.
(2) The robot picks up the pink cup and places it into the trash bin.
(3) The robot picks up the teal cup and places it into the sink.
(4) The robot picks up the bread and places it into the trash bin.

Robot	Task	Task description	Error type	Error description
Panda	**Set table**	From a sink-side staging area, the robot places two plates and two cups onto the table, arranging one cup next to each plate. The intended outcome is two color-matched plate–cup settings.	Misinterpretation	**Color mismatch + layout:** The robot forms two place settings with mismatched colors (*e.g.,* pink plate is placed next to teal cup).
			Control error	**Cup collision + risky grasp:** While placing the second plate, the robot bumps a nearby cup; the grasp appears marginal (held at the edge), raising safety concerns.
Panda	**Clear table**	Given mixed items on a cluttered table, the robot clears items by placing reusable dishware into a sink and disposing of trash into a bin. One small item remains on the table at the end.	Planning error	**Mis-sorted reusable:** The robot places a reusable cup into the trash bin instead of in the sink.
			Sensing error	**Missed paper towel:** The robot fails to notice a crumpled paper towel (outside its field of view) and leaves it on the table.
Tiago	**Fetch salt**	The robot navigates to a counter with seasonings, selects an item, returns to the user at the kitchen table, and places the item near the user's workspace.	Sensing error	**Wrong item retrieved:** The robot retrieves the wrong seasoning (lemon pepper instead of salt), likely due to label orientation.
			Control error	**Missed grasp + retries:** The robot must try multiple times to grasp the object since initial attempts fail, delaying the task.
Tiago	**Pour pasta**	The robot retrieves a pasta jar from a counter, returns to the user, tilts the jar to pour pasta into a pot, and sets the jar down afterward.	Control error	**Spillage during pour:** Some pasta spills onto the table or floor during pouring.
			Sensing error	**Residual pasta left:** The robot stops pouring before the jar is empty, leaving pasta remaining inside.

Table 3: Summary of four teleoperated household-robot tasks and injected issues.

(5) The robot picks up the plastic bowl and places it into the trash bin.
(6) The robot picks up the Oreo package and places it into the trash bin.
(7) The crumpled paper towel remains on the table (out of the robot's field of view).

Injected errors.

- Planning error: The robot places the reusable pink cup into the trash bin (instead of the sink).
- Sensing error: The robot misses the crumpled paper towel and leaves it on the table.

Task 3: Fetch Salt

Scenario. You are sitting at the kitchen table while preparing pasta, and realize you need salt for boiling pasta. The robot will search for the salt, navigate back to you, and put it on the table. Notice what the robot is doing and what object it brings to the table.
Steps.

(1) A pot is on the table in front of the participant, and the robot starts near the participant's table.
(2) The robot navigates to the other side of the kitchen toward a shelf containing cereal boxes, fruit, a pasta jar, and salt and pepper. The labels on the salt and pepper are facing away from the robot.
(3) The robot attempts to grasp the pepper and makes multiple grasp attempts before succeeding.
(4) The robot grasps the pepper, navigates back to the participant's table, and places it on the table next to the pot.

Injected errors.

- Control error: The robot tries multiple times to grasp the object due to the approach angle and distance.
- Sensing/Planning error: The robot brings the wrong object (lemon pepper instead of salt).

Task 4: Pour Pasta

Scenario. You are preparing other ingredients and ask the robot to help by preparing pasta. The robot will search for a pasta jar, bring it to the table, and pour pasta into a pot. Pay attention to how the robot moves through space and how you understand what happens during pouring.
Steps.

(1) The robot navigates back to the kitchen shelf and picks up the pasta jar. The pasta jar has no lid.

(2) The robot grips the pasta jar upright and navigates back to the user.
(3) The robot adjusts its left arm joint to find a pouring angle.
(4) The robot pours pasta into the pot. Some of the pasta spills onto the table and onto the floor.
(5) The robot places the pasta jar on the table. There's remaining pasta in the jar.

Injected errors.

- Control error: The robot spills some pasta onto the table and/or floor.
- Early termination / sensing error: Some pasta remains in the jar after pouring.

Safety Measures. To support participant safety and to reliably introduce planned robot errors, we ran the sessions using a Wizard-of-Oz setup [41]. A researcher continuously supervised the robot and held the emergency stop button. We maintained a safe distance between the robot and participants throughout the study, and whenever participants wanted to touch the robot or inspect the workspace during execution, we paused the robot first to prevent collisions. We used a mix of real kitchenware (*e.g.,* cups, plates, sink) and mock props (*e.g.,* a trash bin). Before starting, we described the objects and the surrounding workspace and invited blind participants to tactually explore them (*e.g.,* materials, object locations, table boundaries) to build familiarity with the study setup. To improve manipulation reliability, we made minor task modifications (*e.g.,* using a pasta jar without a lid). To reduce spill risk and simplify cleanup during the pasta-pouring task, we used an empty pot with no boiling water.

B Online Study

B.1 Participants

We recruited 20 blind participants (P11–P30) through a local chapter of the National Federation of the Blind (NFB) and word of mouth, and 20 sighted participants (S1–S20) through an institutional mailing list (Table 4). To minimize potential bias from prior exposure to the same task scenarios and planned errors, we did not recruit participants in our earlier in-person study.

(a) Blind participants (P11–P30)

PID	Gender	Age	Vision	Onset
P11	Female	57	Totally blind	Congenital
P12	Female	34	Totally blind	Congenital
P13	Female	40	Totally blind	Congenital
P14	Female	19	Totally blind	Acquired
P15	Female	76	Totally blind	Congenital
P16	Male	28	Totally blind	Congenital
P17	Non-binary	30	Totally blind	Congenital
P18	Female	37	Totally blind	Congenital
P19	Male	24	Totally blind	Congenital
P20	Female	52	Light perception	Acquired
P21	Male	45	Totally blind	Congenital
P22	Female	41	Totally blind	Acquired
P23	Female	43	Totally blind	Acquired
P24	Male	53	Totally blind	Acquired
P25	Male	60	Totally blind	Congenital
P26	Male	49	Light perception	Acquired
P27	Male	47	Totally blind	Congenital
P28	Male	62	Light perception	Acquired
P29	Male	25	Light perception	Acquired
P30	Male	29	Totally blind	Acquired

(b) Sighted participants (S1–S20)

PID	Gender	Age	Vision	Onset
S1	Male	28	Sighted	N/A
S2	Non-binary	27	Sighted	N/A
S3	Male	23	Sighted	N/A
S4	Male	27	Sighted	N/A
S5	Male	22	Sighted	N/A
S6	Female	21	Sighted	N/A
S7	Female	26	Sighted	N/A
S8	Female	24	Sighted	N/A
S9	Male	24	Sighted	N/A
S10	Male	28	Sighted	N/A
S11	Female	25	Sighted	N/A
S12	Female	20	Sighted	N/A
S13	Male	26	Sighted	N/A
S14	Female	20	Sighted	N/A
S15	Female	30	Sighted	N/A
S16	Female	25	Sighted	N/A
S17	Male	31	Sighted	N/A
S18	Female	31	Sighted	N/A
S19	Female	20	Sighted	N/A
S20	Female	18	Sighted	N/A

Table 4: Participant demographics for the video-based study.

B.2 Robot Communication Prototype

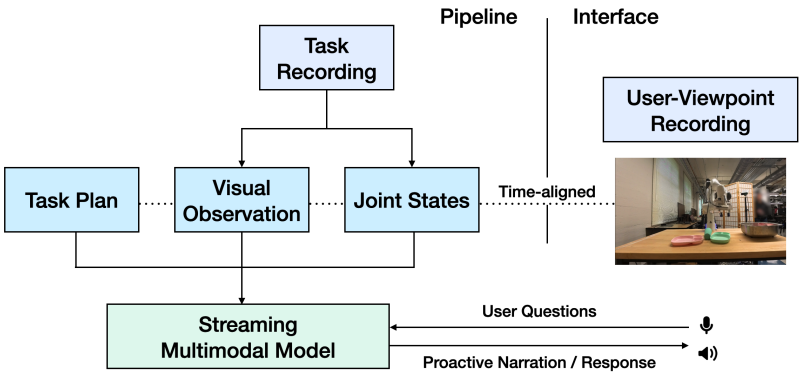

Figure 8: From the robot-side task recording, we extract the robot's visual observations and joint states. Together with a scripted task plan, we input these time-indexed data as context to a streaming multimodal model. On the user side, participants watch a user-viewpoint recording that is time-alinged with the robot-side signals to preserve a realistic sense of live progress. The model generates proactive narration and answers users' questions in real time via spoken responses.

Figure 8 summarizes the system used to generate conversational robot narration in our study. Following Wang et al. [53], we extract the robot's visual observations and joint states from the robot's task recording and provide these data with a scripted task plan into a streaming multimodal model. Participants view a separate user-viewpoint recording that is time-aligned with these robot logs so that narration is grounded in the same underlying moment of the task as the video they are watching. We use Gemini Live [5] for end-to-end speech interaction and do not rely on separate automatic speech recognition or text-to-speech.

The system supports two interaction styles during playback: *reactive* and *mixed-initiative*. In the reactive condition, the model responds only to participants' spoken questions, using the current video time to retrieve the corresponding robot logs. Because each task is handled in a single model session, participants can ask questions about earlier events already discussed, but the system does not answer questions about future events beyond the current playback time. In the mixed-initiative condition, the system additionally triggers brief narration updates at a fixed pulse (every 5 seconds). For each proactive update, we set the reference time to $t = \texttt{video.currentTime} + 2.5\,\text{s}$ to compensate for generation latency, so that the spoken narration aligns with what participants are seeing when it is actually heard. Proactive narration is interruptible – if a participant asks a question while a proactive update is being generated or spoken, the system preempts narration and prioritizes answering the user query.

The system was implemented as a web application (React [33] & TypeScript) integrated with a streaming multimodal model (Gemini Live API [49]) to process and generate speech in real time. For blind participants, we designed the interface to be screen-reader accessible following WCAG guidelines [7] and iteratively tested it on three of the most popular screen readers: NVDA, JAWS, and VoiceOver. We deployed the system online and logged interaction traces and survey responses with Firebase realtime database [8].

B.3 Task Design

Task Understanding Quizzes. To assess participants' understanding of each robot task, we added a post-task quiz after each video. For every task, the quiz contained two required questions: (1) a task-critical question about whether the robot achieved the intended outcome, and (2) a secondary question about an additional task-relevant detail (*e.g.,* a noteworthy event during execution or a specific state detail) that was not strictly necessary to complete the task but helped characterize what participants noticed and remembered. Participants answered both questions in free-text fields in the post-task survey.

Task 1: Set the table.

- Q1: Which objects are on the table at the end of the task?
- Q2: What happened when the robot placed the teal plate (e.g., did anything go wrong)?

Task 2: Clear the Table.

- Q1: Which objects are in the trash bin at the end of the task?
- Q2: Which objects are in the sink at the end of the task?

Task 3: Fetch Salt.

- Q1: What items were on the shelf during the search (e.g., when the robot scanned the shelf)?
- Q2: What did the robot bring back to the participant?

Task 4: Pour Pasta.

- Q1: Where is the pasta jar after the robot pours pasta into the pot?
- Q2: Is there any pasta remaining in the jar after pouring? (Yes/No with an optional brief explanation.)

B.4 Results

Categorization of user questions.

Label	Definition	Example
Scene Check	Ask what is currently visible in the scene.	*What is on the table right now?*
Task Progress	Ask what step the robot is on, or how far along the task is.	*Are you done pouring the pasta?*
Object Status	Ask about the state/location of a specific object.	*Where did the teal plate end up?*
Action Detail	Ask for fine-grained details of an action.	*How did you carry things like a pasta box?*
Reasoning	Ask why the robot chose an action or behaved a certain way.	*Why did you choose the short pasta instead of the long pasta?*
Next Plan	Ask what the robot will do next.	*What will you do next?*
Instruction	Provide new instruction to the robot.	*Go to the cabinet.*
Correction	Point out the robot's error or mismatch in narration.	*No, that's not going into the pot.*
Other	Anything that does not fit the above categories.	*Can you repeat that last part?*

Table 5: Question taxonomy used for annotating participant questions in Figure 5).

Across all sessions, participants asked 3,646 questions in total (Blind: 2,632; Sighted: 1,014). To analyze question content, we developed a taxonomy of question categories and annotated each question with a single label (Table 5 provides definitions and representative examples). First, one researcher randomly sampled 100 questions from each group and proposed an initial set of categories; a second researcher independently reviewed the taxonomy and resolved disagreements through discussion until reaching consensus. We then filtered out non-English questions and very short utterances (fewer than three words). Using the label definitions and few-shot examples for each category, we labeled the remaining questions with `gpt-5-mini`.

Model Inaccuracies Observed in the Online Study. Although the narration pipeline was generally effective for describing observable actions and responding to straightforward questions, we observed several recurring inaccuracies in the online study. First, the model sometimes provided incorrect information about spatial properties (*e.g.,* distance, height, or overall size). Second, when asked questions with assumptions (*e.g., "Did you already pick up the cup?"*), the model hallucinated and provided a plausible-but-incorrect response rather than stating uncertainty. In some cases the model appropriately *abstained* when the answer was not supported, suggesting that errors were not uniform. Below we include representative excerpts illustrating these patterns:

Human: "How high is that?"
Robot: "That shelf is approximately 3 feet high."

Robot: "The person is approximately 10 feet from the kitchen shelf."

Human: "Is it a big counter?"
Robot: "I cannot determine the overall size of the counter from my current perception. I am continuing my search for the salt."

Human: "Do you see anyone in the video? What is that person doing?"
Robot: "The person placed the salt jar on the black counter."

Human: "Did you already pick up the cup that is knocked over?"
Robot: "Yes. I picked it up and put it upright."

While mitigating these inaccuracies is not a primary contribution of this work, they highlight important opportunities for future robot narration systems to be more timely, faithful, and transparent in uncertainty, in both proactive narration and question answering.