Bidirectional Human-Robot Communication for Physical Human-Robot Interaction

Junxiang Wang
Carnegie Mellon University
junxiang@cmu.edu

Cindy Wang
Carnegie Mellon University

Rana Soltani Zarrin
Honda Research Institute USA

Zackory Erickson
Carnegie Mellon University

Figure 1: Example interaction with bidirectional communication. User is in a scratching scenario, where they command the robot to move to the target scratching position (marked with star) through a bidirectional verbal interaction.

Abstract

Effective physical human-robot interaction requires systems that are not only adaptable to user preferences but also transparent about their actions. This paper introduces a system for bidirectional human-robot communication in the context of physically assistive scenarios. Our method empowers users with the ability to modify a robot's planned trajectory—including its position, velocity, and force—in real time using natural language. We utilize a large language model (LLM) to interpret any trajectory modifications implied by user commands in the context of the planned motion and conversation history. Importantly, our system provides verbal feedback in response to the user, either assuring any resulting changes or posing a clarifying question. We evaluated our method in a user study with 18 older adults across three assistive tasks, comparing our bidirectional approach to a unidirectional ablation (modifications without feedback) and a baseline without modifications. Results show that participants successfully used the system to modify trajectories in real time. Moreover, the bidirectional feedback led to significantly higher ratings of interactivity, transparency, and grounding, demonstrating that the robot's verbal response is critical for creating a more intuitive user experience. Supplementary materials can be found on our project website: https://bidir-comm.github.io/

HRI '26, Edinburgh, Scotland, UK

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

CCS Concepts

• Computing methodologies → Discourse, dialog and pragmatics; **• Computer systems organization → External interfaces for robotics**.

Keywords

Human-robot communication, assistive robotics

ACM Reference Format:
Junxiang Wang, Cindy Wang, Rana Soltani Zarrin, and Zackory Erickson. 2025. Bidirectional Human-Robot Communication for Physical Human-Robot Interaction. In *Proceedings of the 2026 ACM/IEEE International Conference on Human-Robot Interaction (HRI '26)*. ACM, New York, NY, USA, 10 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

A growing body of research has demonstrated that robots can address a range of caregiving activities in autonomous physical assistance [6, 9, 24]. However, many systems lack two capabilities essential for interactive autonomy: real-time user-guided adaptation and transparent communication. Real-time adaptation allows users to adjust a robot's ongoing autonomous motion—such as tuning speed or pressure—to fit their own personal preferences and levels of comfort [8, 10, 12]. In parallel, clear communication of a robot's intent and state—such as any upcoming changes to its motions—supports higher transparency and user trust towards the robot [1, 7, 23, 35].

In this paper, we propose a system for bidirectional human-robot verbal communication in physically assistive scenarios, addressing both real-time adaptation and transparency. Our system allows users to issue verbal commands to change a robot's position, velocity, and force *in real time* as the robot autonomously executes a planned interaction for physical assistance. To establish *bidirectional communication*, our system responds to every user utterance with verbal feedback—either an assurance of the desired change or a clarifying question—thereby closing the interaction loop (an example shown in Figure 1). We design compact representations for both the planned trajectory and any modifications to it, facilitating efficient interpretation of user speech through a large language model (LLM). Our system interprets each user utterance in the context of both the planned trajectory and the conversation history (see Figure 1), and when an utterance is underspecified, it asks a clarifying question for further user input.

We evaluate our system via a within-subjects user study with older adults (n = 18) in three different physically assistive tasks. Participants successfully modified the robot's position, velocity, and force through speech with latency appropriate for real-time interaction. Beyond this capability, our bidirectional method of providing verbal feedback to user adjustments yields higher perceived interactivity and transparency than a unidirectional ablation that allows trajectory modifications but offers no feedback, underscoring the importance of bidirectional communication.

In summary, our contributions in this paper are as follows:

- We propose a bidirectional human-robot communication framework in physically assistive scenarios. This framework couples a user's trajectory commands with real-time, transparent verbal feedback from the robot, fostering a more intuitive interaction.
- We present a novel LLM-based pipeline that efficiently interprets user utterances in trajectory and conversation context. The pipeline simultaneously generates modifications to the trajectory's position, velocity, or force and verbal feedback through concise assurances or clarifying questions.
- We conduct a user study with 18 older adults and three physically assistive tasks, which demonstrates that bidirectional communication leads to higher perceived interactivity and transparency than a unidirectional ablation, while speech-based modifications to position, velocity, and force remain fast enough for real-time applications.

2 Related Works

2.1 Language-guided robot motions in HRI

The concept of influencing a robot's action through language inputs has been widely explored in different contexts, predominantly in the field of robotic manipulation. Language is often integrated in learning reward functions [18, 30, 36], selecting motion primitives on a high level [31, 38], computing latent actions [5, 13], or training general language-guided policies [20] and vision-language-action models (VLA) [11, 16]. In comparison, our method leverages the zero-shot reasoning capabilities of a general-purpose LLM. We focus on the assistive domain and utilize an LLM for real-time interpretation of utterances into trajectory modifications, without task-specific training or fine-tuning.

In physically assistive robotics, language is commonly used for issuing task-level commands, such as initiating motions in feeding assistance [4, 25] or selection of predefined motion primitives [3]. Other use cases mostly revolve around commanding a robot arm for assistive object retrieval [26, 28]. These systems are often confined to one particular assistive task, whereas our developed system can be applied to a range of physically assistive trajectories. We also make modifications to trajectories on a parameter level, directly changing all motion aspects of position, velocity, and force, without being limited to configured actions.

2.2 Human-robot dialog systems

While the previous section mainly focuses on voice interfaces being used only in the human-to-robot direction, many prior works also explore dialog systems between humans and robots, similar to our system. Speech serves as one of the most intuitive interfaces that can grant both personalization and transparency, especially for older adults and assistive scenarios [21, 27, 32].

Socially assistive robots often make use of conversations in the contexts of therapy or affective support [19, 29, 33], while these applications generally do not involve physical interactions, which is the emphasis of our work.

Dialog can also play an important role in the domain of human-robot collaboration [2, 22, 34, 37], and the information exchanged often revolves around task assignment, hence which party (human or robot) should be assigned with which step in a task. In contrast, we address the domain of assistive robotics, where humans can provide their preferences verbally when robots autonomously provide physical assistance. Additionally, we focus on assistive tasks that can be completed even with only human commands and no robot verbal feedback, and in this work, we look into the effect of providing such verbal feedback to humans—hence how bidirectional communication affects a user's perception of an interaction that may also be completed with unidirectional human-to-robot communication.

3 Methods

Our framework takes as input (1) a planned physically assistive trajectory and (2) a user utterance, and produces real-time modifications to that trajectory based on the utterance. As outlined in Figure 2, the system either applies and communicates trajectory changes when the utterance directly implies so, or poses a clarifying question seeking for more user input. In this section, we first introduce compact representations for trajectories (§3.1) and modifications to trajectories (§3.2), then discuss the LLM-based interpreter that maps an utterance to the correct modification along with a concise communication as feedback (§3.3).

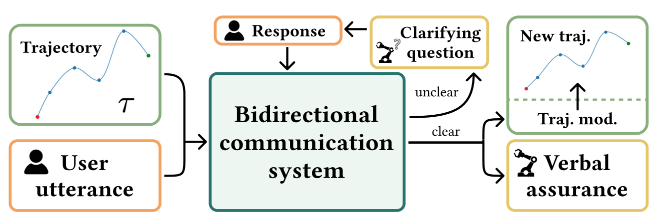

Figure 2: Flowchart of the bidirectional communication system, including two cases of verbal feedback depending on whether a user utterance is clear: (1) assuring and executing any modifications to the trajectory, or (2) posing a clarification question to request for further user input.

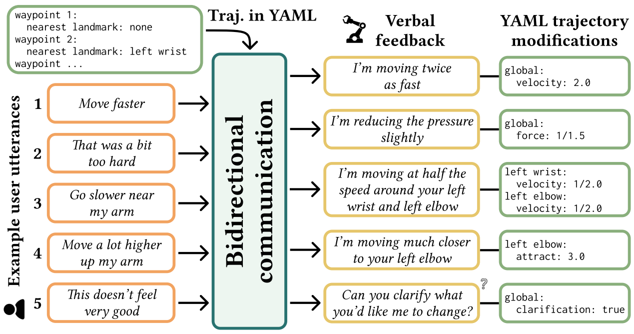

Figure 3: Example YAML trajectory and user utterances as inputs to our bidirectional communication system, along with the generated communications and trajectory modifications in YAML format.

3.1 Trajectory representation

3.1.1 Assumptions. We assume we are provided with a planned end-effector trajectory that interacts with a person, as a sequence of 3D waypoints:

$$\tau = \{w_i : w_i = (t_i, \mathbf{x}_i, v_i, f_i)\}_{i=1}^{N}, \tag{1}$$

where each waypoint w_i consists of a timestamp $t_i \in \mathbb{R}_{\geq 0}$, end-effector position $\mathbf{x}_i \in \mathbb{R}^3$, velocity magnitude $v_i \in \mathbb{R}_{\geq 0}$, and desired force magnitude $f_i \in \mathbb{R}_{\geq 0}$. Between consecutive waypoints, the robot end-effector is assumed to follow a straight-line Cartesian path, with velocity and force linearly interpolated in time. Force magnitudes f_i are intended for assistive contact with the user and are tracked by a low-level controller implementing either impedance or admittance control. We further assume access to estimated 3D positions of all human body landmarks relevant to the planned trajectory, which could be obtained from body pose estimation.

3.1.2 Trajectory representation. We build a minimal representation of the input trajectory τ that gives the LLM a high-level sketch of the planned interaction. Since the trajectory interacts with a person, we compute the nearest human-body landmark for each waypoint and use only this label to form a symbolic representation, without involving kinematic details. This representation is serialized in YAML format for structure (see top-left of Figure 3 for a snippet example). If no landmark falls within a proximity threshold, the label is left unassigned (None in the YAML), which has no geometric implication. Such a representation supports high-level identification of where interactions occur and also provides coarse physical grounding for utterance interpretation, as we discuss in §3.3.2.

3.2 Modification representation

We support trajectory modifications in position, velocity, and force, and we organize them at three scopes: *global* changes that apply to the entire trajectory (§3.2.1), *landmark* changes whose influence is defined with respect to specific body landmarks (§3.2.2), and *waypoint* changes that affect certain waypoints (§3.2.3). These scopes do not imply priority, so if a single utterance specifies changes in multiple scopes, we apply them concurrently to the trajectory. Across utterances, changes accumulate, with velocity clamped to a fixed cap (see implementation details in §4.1). We serialize these modifications in the same YAML schema used for representing trajectories.

3.2.1 Global Changes. We support changes to the velocity and force of the entire trajectory, applied uniformly to *all* waypoints. These changes correspond to broad user utterances such as "Move faster." We apply these changes in a multiplicative manner, with the scaling factor represented in the corresponding YAML fields (see examples 1 and 2 in Figure 3). Thus, a value greater than 1 indicates an increase in velocity or force, and a value between 0 and 1 indicates a decrease.

3.2.2 Landmark Changes. Our system also provides the capability for users to modify kinematic parameters relative to certain body landmarks. These changes fall into two main categories: (1) velocity and force changes around body landmarks, expressed in utterances such as "Be gentler around my knee" (see example 3 in Figure 3), and (2) position changes based on attraction/repulsion from body landmarks, expressed in utterances such as "Come closer to my elbow" and "Move away from my mouth" (see example 4 in Figure 3). We next elaborate on each of these categories.

Local velocity and force changes. These changes are represented similarly to global changes in YAML, but how they are applied to each waypoint is subject to Gaussian decay based on the waypoint's distance from the local landmark. For example, suppose $k > 1$ is the factor of velocity increase around a body landmark located at $\mathbf{p}_{\text{landmark}} \in \mathbb{R}^3$, then waypoint w's velocity v should be increased by a factor of:

$$1 + (k-1)\exp\left(\frac{-\|\mathbf{p}_{\text{landmark}} - \mathbf{x}\|^2}{2\sigma^2}\right) \tag{2}$$

where σ controls the spread of the Gaussian decay. A larger σ results in wider influence around the landmark, while a smaller σ leads to a more localized effect. We empirically set $\sigma = 7\,\text{cm}$, as pilot testing indicates this value aligns well with the localized and decaying nature of influence implied by user utterances like "around my knee." The Gaussian decay for a decrease in velocity or force is expressed similarly.

Position changes. In order to modify the position of individual waypoints in the input trajectory, we apply the notion of artificial potential fields [15], commonly used in robotics for manipulation and navigation with obstacle avoidance. Attractive and repulsive potentials can therefore be placed at body landmarks, and we compute how much to displace a waypoint w from the gradients of all potential functions, evaluated at the waypoint's position $\mathbf{x}$. We specify both attractive and repulsive potentials, as well as the intensity of each, via the `attract` field in the YAML entry for each landmark (see example 4 in Figure 3). Following the convention of multiplicative factors for velocity and force, a value greater than 1 indicates an attractive potential, and a value between 0 and 1 indicates a repulsive potential. For attractive potentials, we use the standard quadratic formulation. Suppose we have an `attract` of $k > 1$ for a landmark located at $\mathbf{p}_{\text{landmark}}$:

$$U_{\text{att}}(\mathbf{x}) = \frac{k}{2} k_p \|\mathbf{p}_{\text{landmark}} - \mathbf{x}\|^2 \tag{3}$$

The gain k_p is empirically determined to be $0.01\,\mathrm{m}^{-2}$. Our application differs from most manipulation and navigation scenarios in that there could be multiple attractive potentials (e.g. attraction to the forearm overall is represented as attractions to the elbow and the wrist) as opposed to a single goal. In order to ensure convergence and avoid waypoints already close to an attractive potential being pulled towards another goal, we weight each attractive potential by the inverse of its distance to the point of interest. The total attractive displacement for a waypoint located at $\mathbf{x}$ is therefore:

$$\Delta_{\text{att}}(\mathbf{x}) = -\sum_j \frac{w_j}{\sum_k w_k} \nabla U_{\text{att},j}(\mathbf{x}), \quad w_j = \frac{1}{\|\mathbf{p}_{\text{landmark},j} - \mathbf{x}\|} \tag{4}$$

For repulsive potentials, we use the formulation with a limit distance of effect ρ_0. Suppose we have an `attract` of $k \in (0, 1)$ for a landmark located at $\mathbf{p}_{\text{landmark}}$:

$$U_{\text{rep}}(\mathbf{x}) = \begin{cases} \frac{\eta}{2k}\left(\frac{1}{\|\mathbf{p}_{\text{landmark}} - \mathbf{x}\|} - \frac{1}{\rho_0}\right)^2 & \text{if } \|\mathbf{p}_{\text{landmark}} - \mathbf{x}\| \le \rho_0 \\ 0 & \text{if } \|\mathbf{p}_{\text{landmark}} - \mathbf{x}\| > \rho_0 \end{cases} \tag{5}$$

The gain η and the distance of effect ρ_0 are empirically determined to be $0.5\,\mathrm{m}^2$ and 10 cm, respectively. Therefore, assuming presence of multiple attractive and repulsive potentials at human-body landmarks, the net displacement of the position $\mathbf{x}$ of a waypoint w is computed as a single step in the opposite direction of the potential field gradients:

$$\Delta(\mathbf{x}) = \Delta_{\text{att}}(\mathbf{x}) - \sum \nabla U_{\text{rep}}(\mathbf{x}) \tag{6}$$

Note that this displacement is only applied once to each waypoint, unlike classical applications of potential fields, which are usually an iterative process.

3.2.3 Waypoint Changes. Lastly, our system also supports changes to the velocity and force of individual waypoints. This capability is useful for handling utterances such as "Go faster when you move away from me," which may target sections of the trajectory far away from any specific body landmark. To execute this change, the system identifies the waypoints implied by the utterance through referencing the trajectory YAML and applies the change uniformly to each one. Although the uniform application is similar to global changes, waypoint changes require generating a separate YAML entry for every modified waypoint.

3.3 Interpretation of User Utterances

We design a structured LLM prompt that translates user utterances and YAML trajectories into trajectory modifications in YAML, accompanied by a brief sentence communicating back to the user any modifications made. In this section, we first discuss some features that allow the system to adapt to various levels of desired change (§3.3.1), as well as interpreting utterances in context of both the trajectory and the conversation history (§3.3.2). Next, we focus on the generation of robot verbal feedback, which forms the concept of bidirectional communication (§3.3.3). Lastly, we provide some rationale and design choices regarding the compactness of LLM outputs (§3.3.4). The complete prompt used in our system can be found in the Appendix.

3.3.1 Granularity. The prompt is designed to handle both generic and fine-grained adjustments to the motion aspects of position, velocity, and force. For a generic adjustment without specifying further granularity (e.g. "Go slower"), we configure the change magnitude k to a default factor of 2 ($k = 2$ for increases and $k = 0.5$ for decreases), as we find this to be an empirically distinguishable magnitude for typical assistive trajectories. When the user desires more fine-grained control, (e.g. "Go slightly slower" and "Press much harder"), the prompt provides examples that allow the LLM to reason about the implied magnitude from the utterance; see examples 2 and 4 in Figure 3 for sample utterances and their corresponding YAML modifications. Finally, we bound the maximum change magnitude to a factor of 3 to avoid drastic changes.

3.3.2 Context-aware Utterance Interpretation. To create a fluid and intuitive interaction, the prompt contains two sources of context—planned trajectory and conversation history—enabling the LLM to interpret ambiguous commands without requiring excessive specificity from the user.

Trajectory context. The prompt's trajectory context is provided via the YAML representation introduced in §3.1, which specifies the nearest body landmark for each waypoint. This context allows the LLM to resolve spatial ambiguities. For instance, a user does not need to specify whether they mean their "left" or "right" elbow, as this information can be inferred from the planned motion. Trajectory context also enables correct understanding of high-level references, like a command relating to an entire "arm." Even though "arm" is not a specific body joint, the LLM can use trajectory information to deduce which landmarks (e.g., shoulder, elbow, wrist) are relevant to the user's command (see examples 3 and 4 in Figure 3).

Conversation context. The prompt also incorporates context from the conversation history, specifically the most recent verbal exchange (user utterance, YAML trajectory changes, and robot verbal feedback). This context allows the LLM to correctly interpret follow-up commands. For example, the user may say "Go faster" and a subsequent command of "A little bit more." While vague in isolation, this second utterance can be correctly interpreted by the LLM as another, smaller velocity increase. Retaining conversation history also enables reversing changes with utterances such as "Undo that" or "Forget what I just said." This conversational memory allows users to make iterative refinements naturally, as shown in Figure 1.

3.3.3 Bidirectional Communication. Bidirectional communication, a key feature of our system, is achieved by providing verbal feedback for each user utterance. We expect that the format of dialog in general, regardless of the exact content, will enhance the perception of interactivity, which is important for physically assistive scenarios. More specifically, we design the system to provide two different types of feedback to handle different utterances: assurance for any utterances that imply changes to the robot's trajectory, or proactively asking a clarifying question when the meaning of an utterance is unclear. The aim of both types of feedback is to fully communicate the robot's internal state to the user to support mutual understanding.

Assurance for change-making utterances. When a desired modification can be extracted from a user utterance, the LLM also generates one concise sentence to assure the user of the upcoming motion modification (see examples 1–4 in Figure 3). These assurances are generated without involving technical details, only communicating the magnitude of change when it is easy to interpret from a user perspective (see examples 1 and 3 in Figure 3).

Clarifying questions for unclear utterances. When an utterance is ambiguous or does not map to an adjustable parameter, even within trajectory and conversational context, the system is designed to seek clarification from the user. Such utterances could be a general expression of feeling (e.g., “This doesn’t feel good”, “I don’t like this”), or it could possibly come from incorrect speech detections, which can be common in real-world environments. In these cases, the system makes no modifications to the trajectory and instead produces a clarifying question to ask for more information. These questions are deliberately kept open, rather than suggesting specific options, to ensure that users retain control over expressing adjustments. Upon the user’s response, a second-stage prompt is then constructed to query the LLM for a new YAML block based on the user’s clarification. This prompt is designed to be much more concise than the main prompt and contains only the immediate context of the question and answer, in order to ensure a minimal response latency from the LLM. This process of posing clarifying questions is iterative, until the user’s intent is no longer ambiguous (as shown in Figure 2). We use a flag in the YAML representation of trajectory modification to indicate whether further clarification is required (see example 5 in Figure 3). These clarifying questions are phrased in plain conversational language, ensuring ease for users to interpret and act on.

3.3.4 Compactness of Response. Because LLMs generate responses autoregressively—one token at a time—the total response latency is directly influenced by the number of output tokens. To minimize latency and maximize response speed, we design the prompt to require the shortest possible response. Specifically, the output is restricted to only the YAML fields that contain modifications (change magnitude $k \neq 1$). Utterances unrelated to trajectory edits (e.g., “I’m happy today”) are also ignored entirely. Most importantly, we prioritize *landmark* changes over *waypoint*. A command like “Move faster near my wrist” could be represented by modifying a list of individual waypoints, but this would result in a verbose output with separate entries for each waypoint. Instead, we prioritize using a landmark change, which not only provides a more compact representation but also offers the smoother exponential decay described in §3.2.2. Collectively, these measures ensure the LLM’s output is concise, minimizing latency and creating a more interactive experience.

4 User Study

We conducted a *within-subjects* user study to evaluate our system’s efficacy for real-time trajectory modification and to measure the specific contribution of bidirectional verbal feedback. Specifically, we test the following two hypotheses:

H1 (Efficacy) – Users will be able to use our bidirectional communication system to effectively modify the position, velocity, and force of planned trajectories for different physically assistive tasks.

H2 (Contribution of Bidirectional Feedback) – Bidirectional verbal feedback from the robot will facilitate a more interactive and transparent experience for users, compared to a no-feedback method with the same ability to make trajectory changes.

4.1 Tasks and Implementation

We designed three different assistive tasks, where each task focuses on a different aspect of the motion (position, velocity, or force) for the user to make modifications to. All tasks were performed autonomously with a Stretch 3 robot, a mobile manipulator with a 5-DoF arm and a gripper.

(1) **Scratching (position)**: The robot held a 3D-printed scratching tool and began scratching near the participant’s left wrist. Participants were given the goal of modifying the position of scratching to an area on the upper forearm, indicated by stickers placed on the participant’s arm.
(2) **Feeding (velocity)**: The robot held a spoon and scooped from a bowl of applesauce to feed the participant a total of three times. Participants were given the goal of modifying the velocity of feeding to a degree they were comfortable with. To encourage participants to actively issue speed adjustments, the initial trajectory was intentionally designed with a slow velocity.
(3) **Bathing (force)**: The robot held a piece of dry washcloth and wiped the participant’s left forearm from the wrist to the elbow, a total of four times. Participants were given the goal of modifying the force of bathing to their liking. Given the subjective nature of force preference, the primary purpose of this task was to ensure participants felt empowered to make adjustments, rather than to converge on a specific target force.

Figure 4 shows one snapshot from each task during the actual user study, along with sample user utterances and the corresponding response from the robot. We used Microsoft Azure’s speech-to-text service to transcribe user speech. The service returns the complete utterance once the user finishes speaking. Upon receiving this utterance, the robot paused its motion and waited for the LLM to map the utterance into desired modifications—there would be a delay of approximately 1-2 seconds, an interval kept short by our compact YAML representation. We used GPT-4.1 as the LLM in our system. When the desired modification was received, the robot updated its trajectory and then restarted its motion along the new trajectory. Multiple utterances in the same interaction were treated as cumulative, so e.g. the second utterance would act on the modified trajectory from the first utterance. The maximum velocity of the robot end-effector was bound to 10 cm/s for safety.

4.2 Participants and Setting

We recruited $n = 18$ older adults from a local independent living community (7 male/11 female; age range 74–90 with $M = 82.1$ and $SD = 4.4$). Only two participants reported any experience with autonomous robots in general (levels 2 and 3 on a 5-point scale; all other participants reported no experience).

The study was conducted in an empty apartment within the community, with all tasks completed in a single session. The study design, the experiment protocol, and the consent forms received approval from the Ethics Committee of our institution.

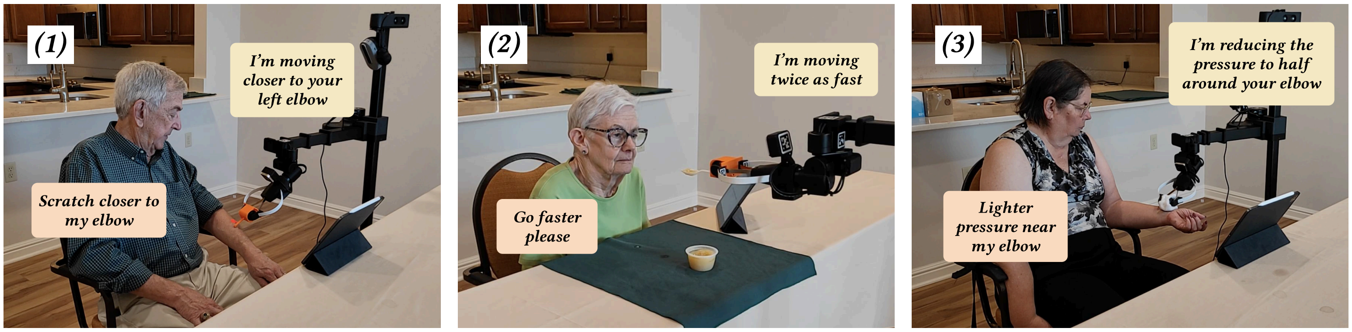

Figure 4: Snapshots from all three tasks implemented for the user study ((1) scratching, (2) feeding, and (3) bathing), with example user utterances (purple) and verbal responses from the robot (blue) for our method.

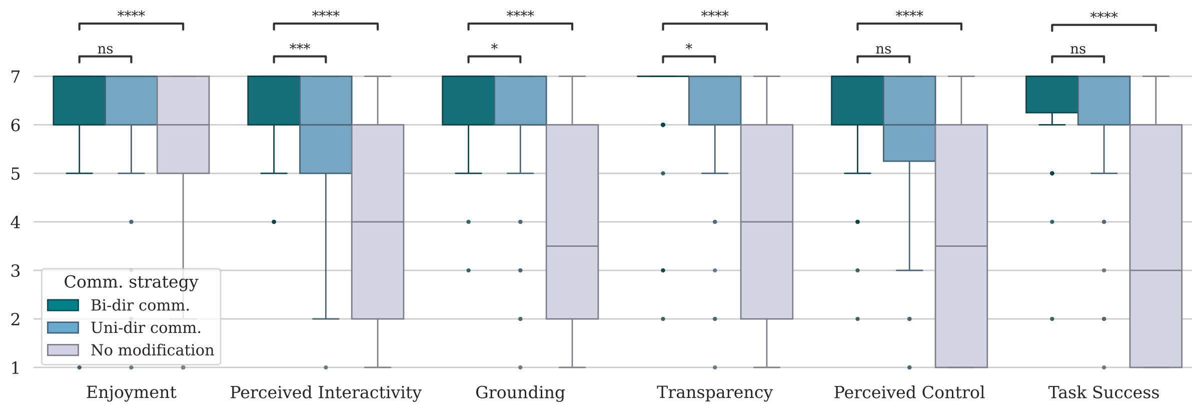

Figure 5: Box plots showing the distribution of survey responses across all participants and tasks. After fitting ordinal mixed-effects models to each question, we conduct Wald tests to assess pairwise differences between our method and both the no communication baseline and the unidirectional communication ablation. "ns" denote lack of significant difference, and asterisks denote significance levels ($p < 0.05$, $p < 0.01$, $p < 0.001$, $p < 0.0001$).

4.3 Communication Strategies and Procedure

To test our hypotheses, we designed three distinct communication strategies that participants experienced in a within-subjects manner:

- **Bidirectional communication** (Our method): This is our full proposed system. The robot would modify its trajectory according to the user's verbal command and provide verbal feedback in the form of either an assurance or a clarifying question.
- **Unidirectional communication** (Ablation): This strategy is designed to isolate the effect of the robot's verbal feedback. The robot can still modify its trajectory based on user commands, but it provides no verbal assurance or clarifying questions. If an utterance is unclear, the robot would simply pause and then resume its previous motion without change.
- **No-modification strategy** (Baseline): In this strategy, the robot would still listen for user speech and pause its motion, but it would not make any modifications to its trajectory. After the standard pause, it would always resume its original, unmodified path. This baseline is designed to measure the efficacy of trajectory changes in our method and the ablation.

Each participant completed a total of nine interactions with the robot. Each interaction (trial) consisted of performing one of the three assistive tasks (scratching, feeding, or bathing) with one of the three communication strategies above. A participant would complete all three trials for a single task—hence experiencing all communication strategies for the task—before proceeding to the next task. To mitigate ordering effects, we counterbalanced the sequence in which the three tasks were presented to each participant, as well as the order of the three communication strategies within each task.

4.4 Measures

After each trial, the participants were asked to answer a survey with the following Likert items on a 7-point scale (7 for strongly agree, 1 for strongly disagree):

L1. *(Enjoyment)* I enjoyed the interaction.
L2. *(Perceived Interactivity)* The robot felt interactive and responsive.
L3. *(Grounding)* I was confident the robot understood what I meant.

L4. *(Transparency)* I could tell exactly what changed in the robot's motion after my input.
L5. *(Perceived Control)* I felt in control of the robot's motions.
L6. *(Task success)* At the end of the trial, I was able to achieve the overall task objective.

The italicized terms inside parentheses denote the high-level concepts each item was designed to evaluate and were not shown to the participants.

We also logged the following data from each trial for quantitative analysis: the content of each user utterance, the corresponding LLM response (YAML and verbal feedback), the LLM processing latency (i.e., the motion pause duration), and the interaction timestamp for each utterance.

5 Results and Discussion

Figure 5 shows the distribution of all participants' responses to the Likert-item questions. We fit ordinal mixed-effects (proportional-odds) models with random intercepts for participant, task, and their interaction effects. Communication strategy was held as a fixed effect. Omnibus effects were assessed with a likelihood-ratio test. Pairwise differences between our method and the baseline or the ablation were evaluated with Wald tests from the fitted model, with Holm adjustment for multiple comparisons.

Figure 6 provides objective evidence for our efficacy hypothesis (H1), illustrating how users successfully modified the robot's motion to achieve the task goals. As per our task design, we visualize the two tasks that were designed with clear, objective targets. The plots show the robot's state over the normalized task progression: the top plot shows the robot's proximity to the target position for the scratching task, while the bottom plot shows the robot's velocity for the feeding task, both averaged across all participants. We do not visualize the results for the bathing task since controlling the force is guided by subjective preference rather than a desired target. In the visualized tasks, the successful modifications are evident in the clear divergence of the trajectories for the modifiable strategies (our method and the unidirectional ablation, solid lines) compared to the static, preplanned path of the no-modification baseline (dotted line). Variations in the baseline reflect the designed dynamics of the original trajectory. To complement the plotted data, the snapshots on the right of Figure 6 visualize the effect of user commands, where colored stars and arrows (red for without modification, green for with modification) visually depict the difference in robot state at a representative point in the interaction.

5.1 Modification Efficacy

As shown in the top plot for scratching in Figure 6, users were able to command the robot to scratch at the target location under the bidirectional and unidirectional strategies (as proximity goes towards zero). In contrast, under the no-modification baseline, the robot's motion was unaffected by user commands and simply followed its original, pre-planned trajectory. The accompanying snapshots on the top-right of Figure 6 provide a visual confirmation, showing the robot successfully reaching the target area (marked with star) in trial with modifications. Similarly, the bottom plot for feeding in Figure 6 shows that users were able to command substantial increases in the robot's speed under the bidirectional

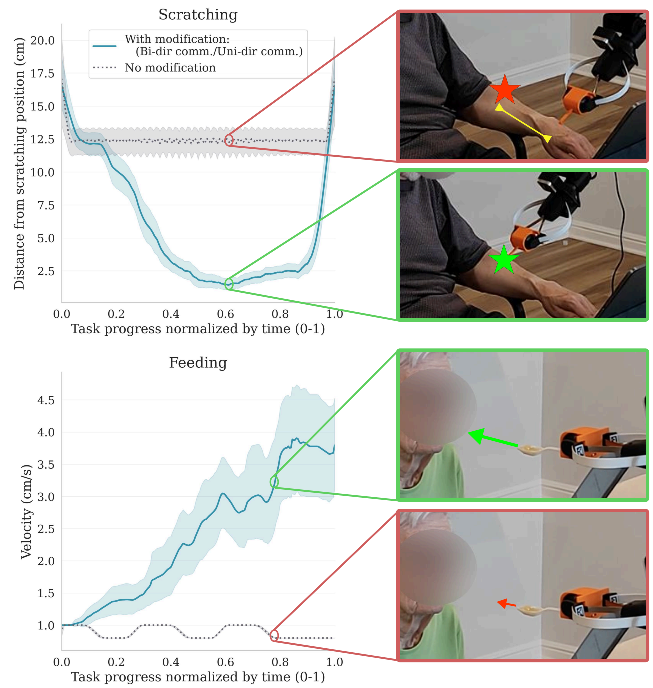

Figure 6: Efficacy of trajectory modifications, averaged over all participants. The plots show changes in position for the scratching task (top) and velocity for the feeding task (bottom) over normalized task progression. They compare the communication strategies that allow modifications (our method and the unidirectional ablation, solid lines) against the no-modification baseline (dotted line). The snapshots on the right visualize the difference in robot state, comparing a trial with modifications to one without at a representative point in the interaction.

and unidirectional strategies, increasing it by a factor of three to four compared to the cautious initial trajectory. In contrast, the velocity profile in the baseline method remained unchanged and followed the preplanned cautious trajectory, irrespective of user commands. This demonstrates the effectiveness of our underlying speech-to-modification system for changing different aspects of a robot's trajectory.

To complement the objective data, we analyzed participants' ratings of efficacy to the Likert items as shown in Figure 5. Participants reported a high degree of task success (L6) and perceived control (L5) when using our method and the unidirectional ablation. Statistical tests revealed that both modifiable strategies were rated significantly higher than the baseline for both L6 ($p < 0.0001$) and L5 ($p < 0.0001$). Importantly, the high task success ratings held across all three tasks, confirming the effectiveness of our method for modifying the intended aspect of each task: position for scratching, velocity for feeding, and force for bathing (see Figure 7), even when the latter's goal is subjective.

Moreover, there was no significant difference in task success or perceived control between our bidirectional method and the unidirectional ablation. This finding confirms that both strategies were equally effective at achieving task goals of modifying trajectories, which carries the crucial implication that any differences observed in other subjective ratings (discussed in §5.2) can be attributed directly to the presence of the robot's verbal feedback.

Finally, the system's efficiency was confirmed by its low latency. The average time from a user finishing an utterance to the robot executing the modification was measured to be 1.7 s across all trials with modifications. This rapid response time directly results from our emphasis on compact representations in LLM outputs and confirms the viability of our system for real-time interaction.

In summary, these three sources of evidence—the objective success in modifying trajectories, the high ratings of task success and user perceived control, and the low system latency—collectively support our **efficacy hypothesis (H1)**. The results confirm that our system is effective and efficient for performing real-time, physical trajectory modifications.

5.2 User Perception and Feedback

To test H2, we analyzed the participants' perception of their interactions as reported in survey ratings (shown in Figure 5). Our bidirectional method was rated significantly higher than the no-modification baseline across all measures ($p < 0.0001$).

More importantly, when compared to the unidirectional ablation, our bidirectional method was perceived as significantly more interactive (L2, $p < 0.001$). Participants also expressed higher confidence that the robot understood them (L3, $p < 0.05$) and found it easier to discern changes in the robot's motion (L4, $p < 0.05$).

This quantitative preference is strongly supported by the qualitative feedback gathered in post-study debriefings. When asked which communication strategy they preferred, all participants selected our method of bidirectional communication. Participants valued the robot's verbal assurances, with P15 noting that with assurances, "you know the robot has interpreted what you want." P18 felt that assurances made them "prepared for what's going to happen." However, P13 offered the nuance that while "reassuring" initially, such confirmations could become "annoying" in long-term use.

In contrast, participants were universally positive about the value of clarifying questions. Our bidirectional system generated a clarifying question in response to 17% of all user utterances, demonstrating its role as a key mechanism for resolving ambiguity. P13 stated that clarifying questions provide a method for them to "learn from the robot," while P6 commented mid-study that the interaction felt like a "two-way communication," presciently identifying the core of our hypothesis.

In summary, through both the survey analysis and qualitative feedback, results show that our method of bidirectional communication was strongly preferred over the unidirectional one, fostering a more interactive and transparent experience, hence supporting **H2**.

5.3 Perceptual Bias and the Need for Transparency

An interesting finding emerged when we analyzed the perceived task success (L6) in a per-task manner, as shown in Figure 7. The data reveals a stark contrast between tasks with objective versus subjective goals.

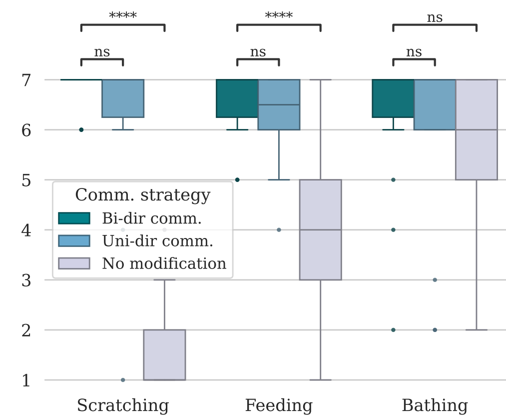

Figure 7: Box plot showing the distribution of survey responses for perceived task success (L6), separate for each task. "ns" denote lack of significant difference, and asterisks denote significance levels (∗∗∗∗ means $p < 0.0001$).

For the scratching task, which had a clear, objective target, participants correctly identified the baseline's failure to react to their utterances, rating its success very low (upper quartile of 2). This was significantly different from our method, which was rated very successful. This shows that when a change is easy to perceive, users can accurately assess the modifications to a robot's motions.

However, for the bathing task, where the "correct" force is subjective, participants reported a high degree of success (lower quartile of 5) even for the baseline method, with no statistically significant difference from our method. Such ratings suggest a powerful perceptual bias: when a change is difficult to perceive and driven by user command, participants tend to believe their command was successful, even when the robot's behavior did not change. This interpretation is further supported by qualitative feedback, where participants claimed they perceived commanded force changes, and sometimes velocity changes too, even while experiencing the no-modification baseline. Our observation finds psychological grounding in research on causal attribution [14] and the illusion of control [17], where people may infer causal relationship between their own action and a temporally subsequent event.

This finding underscores the importance of transparency in physical human-robot interactions. When a user's perception can diverge from the robot's actual behavior, the liability is on the robot to provide clear and transparent feedback. Without such grounding, the user may develop an inaccurate mental model of the system's capabilities, leading to frustration and mistrust in the long run.

6 Conclusion

We present a framework for bidirectional human-robot communication during physically assistive scenarios, where users can verbally modify a robot's trajectory in real time across the motion aspects of position, velocity, and force. Our method leverages an LLM to efficiently translate any user utterance into compactly represented trajectory modifications while simultaneously generating appropriate verbal feedback—either as an assurance of a desired change or as a clarifying question. A user study with 18 older adults demonstrates the efficacy of our method for trajectory modifications, and that the robot's bidirectional verbal feedback significantly enhances the user experience by improving perceived interactivity and transparency.

Future work. Our work has a few limitations that could open avenues for future research. First, our user study only involved a single session per participant, so how to appropriately structure bidirectional verbal response in a long term deployment scenario remains unexplored. Additionally, our system can modify a range of physically assistive planned trajectories, but how to effectively adapt the bidirectional communication framework to real-time controllers or policies is an open question. Moreover, the most intuitive bidirectional communication might not be in the form of spoken language at all times, and future research could consider integrating other methods of communication (e.g. tactile, gestures) as well.

Acknowledgments

This work is supported by Honda Research Institute USA.

References

[1] Basel Alhaji, Michael Prilla, and Andreas Rausch. 2021. Trust dynamics and verbal assurances in human robot physical collaboration. *Frontiers in Artificial Intelligence* 4 (2021), 703504.

[2] Philipp Allgeuer, Hassan Ali, and Stefan Wermter. 2024. When robots get chatty: Grounding multimodal human-robot conversation and collaboration. In *International Conference on Artificial Neural Networks*. Springer, 306–321.

[3] Ruben Alonso, Emanuele Concas, and Diego Reforgiato Recupero. 2021. An abstraction layer exploiting voice assistant technologies for effective human—Robot interaction. *Applied Sciences* 11, 19 (2021), 9165.

[4] Tapomayukh Bhattacharjee, Ethan K Gordon, Rosario Scalise, Maria E Cabrera, Anat Caspi, Maya Cakmak, and Siddhartha S Srinivasa. 2020. Is more autonomy always better? exploring preferences of users with mobility impairments in robot-assisted feeding. In *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*. 181–190.

[5] Yuchen Cui, Siddharth Karamcheti, Raj Palleti, Nidhya Shivakumar, Percy Liang, and Dorsa Sadigh. 2023. No, to the right: Online language corrections for robotic manipulation via shared autonomy. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. 93–101.

[6] Zackory Erickson, Henry M Clever, Greg Turk, C Karen Liu, and Charles C Kemp. 2018. Deep haptic model predictive control for robot-assisted dressing. In *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 4437–4444.

[7] Kerstin Fischer, Hanna Mareike Weigelin, and Leon Bodenhagen. 2018. Increasing trust in human–robot medical interactions: effects of transparency and adaptability. *Paladyn, Journal of Behavioral Robotics* 9, 1 (2018), 95–109.

[8] Sugeeth Gopinathan, Sonja K Ötting, and Jochen J Steil. 2017. A user study on personalized stiffness control and task specificity in physical human–robot interaction. *Frontiers in Robotics and AI* 4 (2017), 58.

[9] Ethan K Gordon, Rajat Kumar Jenamani, Amal Nanavati, Ziang Liu, Daniel Stabile, Xilai Dai, Tapomayukh Bhattacharjee, Tyler Schrenk, Jonathan Ko, Haya Bolotski, et al. 2024. An adaptable, safe, and portable robot-assisted feeding system. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. 74–76.

[10] Vivek Gupte, Dan R Suissa, and Yael Edan. 2023. Optometrist's Algorithm for Personalizing Robot-Human Handovers. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2366–2372.

[11] Physical Intelligence, Kevin Black, Noah Brown, James Darpinian, Karan Dhabalia, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, et al. 2025. $\pi_{0.5}$: a Vision-Language-Action Model with Open-World Generalization. *arXiv preprint arXiv:2504.16054* (2025).

[12] 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).

[13] Siddharth Karamcheti, Megha Srivastava, Percy Liang, and Dorsa Sadigh. 2022. LILA: Language-informed latent actions. In *Conference on robot learning*. PMLR, 1379–1390.

[14] Harold H Kelley and John L Michela. 1980. Attribution theory and research. *Annual review of psychology* 31, 1 (1980), 457–501.

[15] Oussama Khatib. 1986. Real-time obstacle avoidance for manipulators and mobile robots. *The International Journal of Robotics Research (IJRR)* 5, 1 (1986), 90–98.

[16] Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan Foster, Grace Lam, Pannag Sanketi, et al. 2024. OpenVLA: An open-source vision-language-action model. *arXiv preprint arXiv:2406.09246* (2024).

[17] Ellen J Langer. 1975. The illusion of control. *Journal of personality and social psychology* 32, 2 (1975), 311.

[18] Jacky Liang, Fei Xia, Wenhao Yu, Andy Zeng, Montserrat Gonzalez Arenas, Maria Attarian, Maria Bauza, Matthew Bennice, Alex Bewley, Adil Dostmohamed, et al. 2024. Learning to learn faster from human feedback with language model predictive control. *arXiv preprint arXiv:2402.11450* (2024).

[19] Maria R Lima, Maitreyee Wairagkar, Manish Gupta, Ferdinando Rodriguez y Baena, Payam Barnaghi, David J Sharp, and Ravi Vaidyanathan. 2021. Conversational affective social robots for ageing and dementia support. *IEEE Transactions on Cognitive and Developmental Systems* 14, 4 (2021), 1378–1397.

[20] Corey Lynch, Ayzaan Wahid, Jonathan Tompson, Tianli Ding, James Betker, Robert Baruch, Travis Armstrong, and Pete Florence. 2023. Interactive language: Talking to robots in real time. *IEEE Robotics and Automation Letters* (2023).

[21] Amama Mahmood, Junxiang Wang, and Chien-Ming Huang. 2024. Situated Understanding of Errors in Older Adults' Interactions with Voice Assistants: A Month-Long, In-Home Study. *arXiv preprint arXiv:2403.02421* (2024).

[22] Zhao Mandi, Shreeya Jain, and Shuran Song. 2024. Roco: Dialectic multi-robot collaboration with large language models. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 286–299.

[23] Siddharth Mehrotra, Chadha Degachi, Oleksandra Vereschak, Catholijn M Jonker, and Myrthe L Tielman. 2024. A systematic review on fostering appropriate trust in Human-AI interaction: Trends, opportunities and challenges. *ACM Journal on Responsible Computing* 1, 4 (2024), 1–45.

[24] 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).

[25] Akhil Padmanabha, Jessie Yuan, Janavi Gupta, Zulekha Karachiwalla, Carmel Majidi, Henny Admoni, and Zackory Erickson. 2024. VoicePilot: Harnessing LLMs as speech interfaces for physically assistive robots. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–18.

[26] Samuel Poirier, François Routhier, and Alexandre Campeau-Lecours. 2019. Voice control interface prototype for assistive robots for people living with upper limb disabilities. In *2019 IEEE 16th international conference on rehabilitation Robotics (ICORR)*. IEEE, 46–52.

[27] Alisha Pradhan, Amanda Lazar, and Leah Findlater. 2020. Use of intelligent voice assistants by older adults with low technology use. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 4 (2020), 1–27.

[28] Terrin Babu Pulikottil, Marco Caimmi, Maria Grazia D'Angelo, Emilia Biffi, Stefania Pellegrinelli, and Lorenzo Molinari Tosatti. 2018. A voice control system for assistive robotic arms: preliminary usability tests on patients. In *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*. IEEE, 167–172.

[29] Frank Rudzicz, Rosalie Wang, Momotaz Begum, and Alex Mihailidis. 2015. Speech interaction with personal assistive robots supporting aging at home for individuals with Alzheimer's disease. *ACM Transactions on Accessible Computing (TACCESS)* 7, 2 (2015), 1–22.

[30] Pratyusha Sharma, Balakumar Sundaralingam, Valts Blukis, Chris Paxton, Tucker Hermans, Antonio Torralba, Jacob Andreas, and Dieter Fox. 2022. Correcting robot plans with natural language feedback. *arXiv preprint arXiv:2204.05186* (2022).

[31] Lucy Xiaoyang Shi, Zheyuan Hu, Tony Z Zhao, Archit Sharma, Karl Pertsch, Jianlan Luo, Sergey Levine, and Chelsea Finn. 2024. Yell at your robot: Improving on-the-fly from language corrections. *arXiv preprint arXiv:2403.12910* (2024).

[32] Yao Song, Yanpu Yang, and Peiyao Cheng. 2022. The investigation of adoption of voice-user interface (VUI) in smart home systems among Chinese older adults. *Sensors* 22, 4 (2022), 1614.

[33] Micol Spitale, Silvia Silleresi, Franca Garzotto, and Maja J Matarić. 2023. Using socially assistive robots in speech-language therapy for children with language impairments. *International Journal of Social Robotics* 15, 9 (2023), 1525–1542.

[34] Huaxiaoyue Wang, Kushal Kedia, Juntao Ren, Rahma Abdullah, Atiksh Bhardwaj, Angela Chao, Kelly Y Chen, Nathaniel Chin, Prithwish Dan, Xinyi Fan, et al. 2024. Mosaic: A modular system for assistive and interactive cooking. *arXiv preprint arXiv:2402.18796* (2024).

[35] Junxiang Wang, Emek Barış Küçüktabak, Rana Soltani Zarrin, and Zackory Erickson. 2025. CoRI: Communication of Robot Intent for Physical Human-Robot Interaction. In *9th Annual Conference on Robot Learning*.

[36] Zhaojing Yang, Miru Jun, Jeremy Tien, Stuart J. Russell, Anca Dragan, and Erdem Biyik. 2024. Trajectory Improvement and Reward Learning from Comparative Language Feedback. In *8th Annual Conference on Robot Learning*.
[37] Albert Yu, Chengshu Li, Luca Macesanu, Arnav Balaji, Ruchira Ray, Raymond Mooney, and Roberto Martín-Martín. 2025. Mixed-Initiative Dialog for Human-Robot Collaborative Manipulation. *arXiv preprint arXiv:2508.05535* (2025).
[38] Lihan Zha, Yuchen Cui, Li-Heng Lin, Minae Kwon, Montserrat Gonzalez Arenas, Andy Zeng, Fei Xia, and Dorsa Sadigh. 2024. Distilling and retrieving generalizable knowledge for robot manipulation via language corrections. In *2024 IEEE international conference on robotics and automation (ICRA)*. IEEE, 15172–15179.

Received 30 September 2025; accepted 1 December 2025