Effective Task Training Strategies for Instructional Robots

Allison Sauppé and Bilge Mutlu
Department of Computer Sciences, University of Wisconsin–Madison
1210 West Dayton Street, Madison, WI 53706 USA
asauppe@cs.wisc.edu, bilge@cs.wisc.edu

***Abstract*—From teaching in labs to training for assembly, a key role that robots are expected to play is to instruct their users in completing physical tasks. While task instruction requires a wide range of capabilities, such as effective use of verbal and nonverbal language, a fundamental requirement for an instructional robot is to provide its students with task instructions in a way that maximizes their understanding of and performance in the task. In this paper, we present an autonomous instructional robot system and investigate how different instructional strategies affect user performance and experience. We collected data on human instructor-trainee interactions in a pipe-assembly task. Our analysis identified two key instructional strategies: (1) *grouping* instructions together and (2) *summarizing* the outcome of subsequent instructions. We implemented these strategies into a humanlike robot that autonomously instructed its users in the same pipe-assembly task. To achieve autonomous instruction, we also developed a repair mechanism that enabled the robot to correct mistakes and misunderstandings. An evaluation of the instructional strategies in a human-robot interaction study showed that employing the grouping strategy resulted in faster task completion and increased rapport with the robot, although it also increased the number of task breakdowns. Our model of instructional strategies and study findings offer strong implications for the design of instructional robots.**

I. Introduction

As robots enter instructional roles such as teaching in classrooms, training for assembly on a shop floor, and teaching medical students surgical procedures, they will need to effectively present task instructions, providing clarifications and corrections when needed, to improve task outcomes and user experience. Robots' success in instruction will depend on their effectiveness first in their use of language, including linguistic and nonverbal cues [2, 5, 14, 22], and second in their presentation of task information, including what information they disclose at a given moment, how they present task information, and how they correct misunderstandings. This paper focuses on the latter problem of effectively presenting task information and explores how robots might adopt the strategies that human instructors use to present task information and what strategies might be most effective.

Human instructors carefully plan instructions to maximize their students' ability to integrate the material, such as first choosing a subgoal to address in a task and plan future instructions to address the chosen subgoal to help contextualize the instructions [4, 10]. To aid participants in completing the step, instructions are iteratively refined until they are atomic. Instructors might also engage the student in the instruction, encouraging "learning by doing" to enable the student to achieve a deeper understanding of the instructions by performing them [1]. These discourse strategies might inform how a robot should order instructions and engage participants.

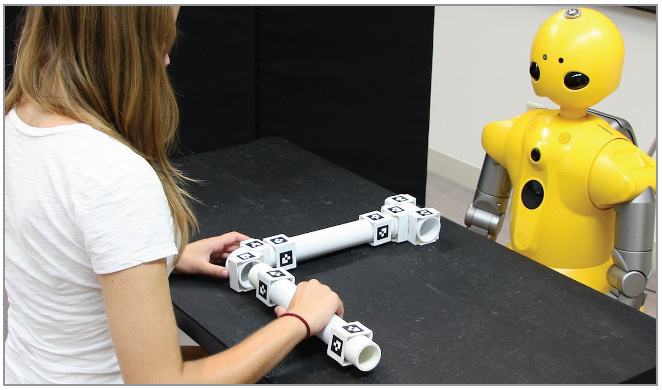


Fig. 1. The robot autonomously guiding a participant in assembling pipes.

In addition to an effective method of delivery, task-based instruction requires instructors to monitor student understanding and progress and to provide feedback and corrections. As the instructor and student progress in the task, they may encounter *breakdowns*—misunderstandings or miscommunication concerning the task goals—that can impede task progress. Instructors need to *repair* these breakdowns by resolving such differences in understanding. Failure to repair breakdowns might lead to compounded breakdowns later in the interaction, further hindering progress. This repair is often context-specific in that it requires knowledge of prior actions and current expectations in order to succeed. Additionally, humans use a variety of techniques to repair breakdowns [12] and adapt their use of these techniques to the context of the interaction [20].

In this paper, we build a better understanding of these instructional and repair strategies by collecting and analyzing data from human instructor-trainee pairs on task instruction. We then implement models of these strategies on an autonomous robot system that guides users through a pipe-assembly task, mimicking real-world assembly tasks in which robots are expected to participate (Figure 1). This system enables the robot to use each of the teaching strategies employed by human instructors to provide students with task instructions and to autonomously handle repair when breakdowns arise. Using this system, we conducted an exploratory human-robot interaction study to assess the tradeoffs between different instructional strategies in measures such as the number of repairs conducted, task completion time, and user experience with the robot. In summary, our work makes the following contributions:

1) A better understanding of human-human instruction.
2) Models for planning instructions and repairing breakdowns and their implementation in a robot system.
3) The validation of our models and their implementation in an instructional scenario and an understanding of the effectiveness of different instructional strategies.
4) The demonstration of an integrated process for designing effective robot behaviors that involves modeling human behaviors, implementing the resulting model in robots, and evaluating implemented behaviors in a user study.

II. Background

In order to enable robots to successfully fulfill instructional roles, it is necessary to understand what instructional strategies would be best for robots to follow. We draw inspiration from how humans give task instruction to model and implement teaching strategies that maximize task outcomes and student experience in human-robot instruction. This section reviews prior work on strategies that humans use in presenting task information and on the development of instructional robots.

A. Instruction in Human-Human Interaction

Effectively communicating a series of instructions is a complex task that has been studied at a number of levels, including how human instructors develop and communicate instructions for their students. Prior work has suggested that instructors follow a discourse planning process based on iterative refinement, where the instructor first picks a subgoal to complete and then further decomposes the subgoal into atomic actions [4, 10]. Instructions are then ordered based on logical segmentations of steps to help students contextualize the task [11]. These models provide important insights into how instructors break task goals into a set of instructions.

Successfully directing a student in a task also relies on feedback from the student. Despite the best efforts of instructors, there will inevitably be instances of *breakdowns*—misunderstandings or miscommunication concerning task goals—that can either impede ongoing progress or lead to breakdowns in the future [29]. To correct breakdowns, humans engage in *repair*, a process that allows participants to correct misunderstandings and helps ensure that all participants have a similar understanding of the relayed information [12, 29]. The process of engaging in repair is often context-sensitive [21]. For example, when a topic is being discussed in a classroom, the instructor frequently initiates repair to clarify students' statements. However, when the classroom is engaged in a task, students are more likely to initiate repair with their peers.

B. Instruction in Human-Robot Interaction

Prior research in robotics has explored how robots might function in instructional settings, such as daycare facilities and classrooms [15, 25, 24], and aid in task instruction, such as offering assistance in a hand washing task [13] and giving directions in a cooking task [27]. Among these studies, work on task instruction has focused on how robots might adapt task instructions to user needs and instructional goals. For instance, Torrey et al. [26] explored how adapting the comprehensiveness of the robot's instructions to its user's expertise might affect task outcomes and user experience. They found that more comprehensive instructions resulted in fewer mistakes among novices, while experts rated the robot as more effective, more authoritative, and less patronizing when it provided brief descriptions. Foster et al. [6] studied the effects of the order in which the robot provided task goals along with instructions on student recall of task steps, showing that providing task goals prior to issuing task steps resulted in fewer requests for repetition by the student later in the task.

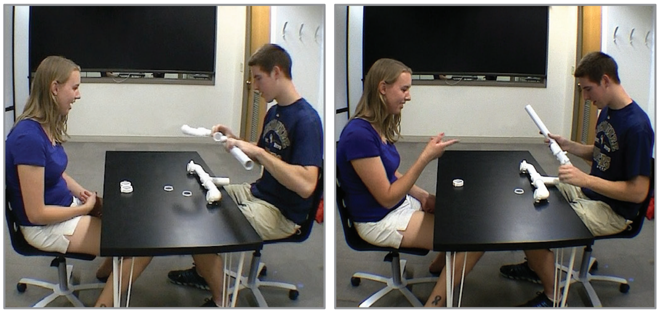


Fig. 2. The *instructor* (participant on the left) directing the *student* (participant on the right) in assembling a predetermined pipe configuration.

Just as repair is necessary in human instruction, robots must also be capable of identifying breakdowns and offering repair for effective human-robot instruction. Prior work has explored a variety of techniques to alleviate the need for repair, such as taking into account the speaker's perspective [28] or mitigating the negative impact of breakdowns through framing [18]. While these studies point to instructional and repair strategies as key elements of the design of instructional robots, enabling robots to use strategies that maximize task outcomes and student experience requires a better understanding and models of effective task instruction. The following section details our work on developing such models.

III. Modeling

To better understand human teaching strategies, we collected video data of human-human interactions during an instructional pipe-assembly task that resembled assembly tasks in which robots might guide humans, such as furniture assembly. Below, we discuss our data collection process, analysis, and the instruction models we constructed from the data.

A. Data Collection

We collected video data from eight instructor-trainee dyads during a pipe-assembly task. In each of these interactions, one participant (the instructor) first learned how to connect a set of pipes into a particular formation from a pre-recorded video. Instructors were given as much time as necessary to re-watch the video and were provided use of the pipes during training. Upon learning the instructions, the instructor trained the second participant (the trainee) on how to correctly assemble the pipes without the aid of the video (Figure 2).

Eight males and eight females aged 18 to 44 ($M = 23.75$, $SD = 8.56$) were recruited from the local community. Each interaction was recorded by a video camera equipped with a wide-angle lens to capture the participants and the task space. The instructional portion of the task, excluding the time the first participant spent learning how to construct the pipes, ranged from 3:57 to 6:44 minutes ($M = 5:11$, $SD = 2:19$).

B. Analysis

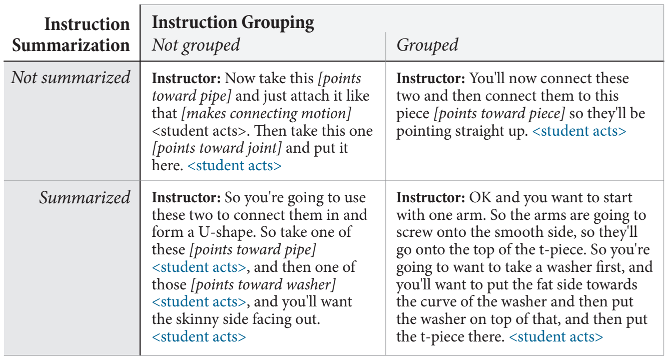
The analysis of the videos involved coding for significant events, including the number of instructions given during a single turn, whether subsequent instructions were summarized, and how repair was initiated and given. To ensure reliability of the coding, a second coder analyzed the videos. The inter-rater reliability showed substantial agreement between the primary and secondary coders (79% agreement, Cohen's κ = .74) [17].

The analysis of our data helped us to better understand different strategies instructors use to deliver instructions and confirmed examples for our understanding of repair gained from the literature. In our data, we observed instructors organizing their instructions along two major factors: how many instructions they gave at once, and whether or not they gave a high-level summary of what the next few instructions would accomplish. We coded our videos with these two factors. Our analysis showed that, considering all instructions given across all dyads, 72% of instructions involved descriptions of individual steps, while 28% were grouped with one or more other instructions. Twenty-one percent of all instructions were prefaced with a summary of the instructions that followed, with the remaining 79% of instructions not including a summary.

Our analysis also showed that instructors always initiated the repair verbally, regardless of whether they became aware of the breakdown verbally, such as a question by the trainee, or visually, such as noticing that the task space was not configured correctly. We found that 65% of these repairs were *trainee-initiated*, while 35% of repairs were *instructor-initiated*.

Trainee-initiated repair—also called *requests*—always involved verbal statements that clarified or confirmed instructor expectations when the trainee either did not understand or misunderstood an instruction. These statements ranged from brief queries (e.g., "What?") to more detailed requests, such as "Where should the pipe go?" Consistent with prior work that associated confusion with not understanding and clarification with misunderstanding [8, 12, 16], we classified requests into the categories *confusion*, *confirmation*, and *clarification*.

Where trainee-initiated repair was directed towards better understanding expectations, instructor-initiated repair clarified or corrected the trainee's perceptions of the task. Instructors initiated repair under one of two circumstances: *mistake detection* and *hesitancy*. When instructors noticed the trainee performing an action that the instructor knew not to be consistent with the goals of that instruction, such as picking up the wrong piece, they verbally corrected the trainee. When instructors noticed that the trainee was hesitating to take action, which was indicated by an average delay of 9.84 seconds in following an instruction, they asked if the trainee needed help.



Instruction Summarization	**Instruction Grouping** *Not grouped*	*Grouped*
Not summarized	**Instructor:** Now take this *[points toward pipe]* and just attach it like that *[makes connecting motion]* <student acts>. Then take this one *[points toward joint]* and put it here. <student acts>	**Instructor:** You'll now connect these two and then connect them to this piece *[points toward piece]* so they'll be pointing straight up. <student acts>
Summarized	**Instructor:** So you're going to use these two to connect them in and form a U-shape. So take one of these *[points toward pipe]* <student acts>, and then one of those *[points toward washer]* <student acts>, and you'll want the skinny side facing out. <student acts>	**Instructor:** OK and you want to start with one arm. So the arms are going to screw onto the smooth side, so they'll go onto the top of the t-piece. So you're going to want to take a washer first, and you'll want to put the fat side towards the curve of the washer and then put the washer on top of that, and then put the t-piece there. <student acts>

Fig. 3. Examples of how the two factors found in our modeling, *instruction grouping* and *instruction summarization*, can be jointly used.

C. Model

Our analysis informed the development of a model with two components: *instructional strategies* and *repair*.

1) Instructional Strategies: As noted in our analysis, instructor strategies for organizing instructions involved two factors: grouping and summarization. In *grouping*, instructors vary the number of instructions given from $1 \ldots i$ before the student completes the instructions. Instructors may provide one instruction at a time and allow the student to carry it out before providing the next instruction or offer grouped instructions by conveying i instructions, given that $i > 1$, prior to the student fulfilling the instructions. When instructors provide *instruction summarization*, they preface their instructions with a high-level summary of the goal of the subsequent k instructions. For example, when the next four steps will result in a set of pipes forming a U-shape, the instructor may say "Now, we'll be taking a few pipes and connecting them into a U-shape" prior to giving the first step. While we categorized instructional strategies into the grouping and summarization factors, our analysis demonstrated that all four possible combinations of these factors were exhibited, as illustrated in Figure 3.

2) Repair: Regardless of the instructional strategy utilized, we observed instructors engage in three forms of repair: *requests*, *hesitancy*, and *mistake detection*. Below, we describe these behaviors and present model components for determining whether repair is needed and, if so, how it might be performed.

Requests: All trainee requests, including questions and statements, were considered as requests for repair. To enable the model to determine the appropriate response, we classified requests into semantic categories using semantic-language modeling. For example, "Which piece do I need?" and "What piece should I get?" were recognized as the same question.

Hesitancy: Depending on the task, indicators such as time elapsed since the last interaction or time elapsed since the workspace was last changed can signal hesitancy in performing instructions. For the pipe-assembly task, we chose to use the time elapsed since the workspace was last changed as a conservative predictor of hesitancy-based breakdowns, as using time elapsed since the last interaction could result in incorrectly inferring hesitancy while the trainee is still working. Based on our observations of how long human instructors waited before offering repair, we considered 10 seconds of no change to the workspace to indicate a hesitancy-based breakdown.

Mistake Detection: While requests and hesitancy-based breakdowns are triggered by the student's action or inaction, mistake detection requires checking the student's work. In our proposed model, we chose a simulation-theoretic approach to direct the robot's behavior in relation to the participant. This approach posits that humans represent the mental states of others by adopting their partner's perspective to better understand the partner's beliefs and goals [7, 9]. This approach has been used in designing robot behaviors and control architectures to allow robots to consider their human partner's perspective [3, 19]. In the context of an instructional task, the instructor has a mental model of an action that they wish to convey to the trainee. Following instruction, the instructor can assess gaps in the trainee's understanding or performance by comparing the trainee's actions to their mental model of the intended action and noting the differences that occur.

Following the simulation-theoretic approach, we defined a set of instruction goals $P = \{p_1, \ldots, p_n\}$ for the robot regarding the result of the participant's action or inaction given the current instruction. Depending on the task, P may vary at each step of the instruction, as some instruction goals may no longer be applicable, while others may become applicable. As the participant engages in the task, the robot will evaluate whether the current state of the workspace is identical to the set of instruction goals P^*. If any of the individual task goals p_k do not match p_k^*, then there is a need for repair.

How repair is carried out depends on which task goal p_k has been violated. As we observed in our analysis of the human-human interactions, the instructor repaired only the part of the instruction that was currently incorrect. Additionally, there is an inherent ordering to the set P that is informed by the participant's perception of the task. The participant's ordering of P is informed by *elaboration theory*, which states that people order their instructions based on what they perceive as being the most important and then reveal lower levels of detail as necessary [20]. By imposing an ordering of decreasing importance on the set P based on these principles for a given task, we can ensure that each p_k takes precedence over any p_{k+n} for $n > 0$. If multiple p_k are violated, then the task goal with the lowest k is addressed first. An example of this ordering can be seen if a participant has picked up the wrong piece and attached it in the wrong location. The instructor first repairs the type of piece needed and then the location of that piece.

Although we discuss the model for detecting mistakes in terms of task steps and goals, this model can also be extended to understanding and repairing verbal mistakes. For example, if the participant mishears a question and responds in a way that is inconsistent with the answers expected, then repair is needed. The appropriate answers of the intended question can be formalized as p_k, and any answer that does not fulfill p_k can be considered as a cause for repair.

IV. System

To create an autonomous system that implements our models, we contextualized our task in the same scenario used for modeling human-human interactions. Using our findings from the previous stage, we designed our system to enable the processing of both verbal and visual information to check the participant's workspace and to detect and repair breakdowns.

A. Hardware

We implemented our model on a Wakamaru humanoid robot (Figure 1). Our model uses information provided by both video and audio captured at 12 frames per second using a Microsoft Kinect stereo camera and microphone-array sensor. The camera and microphone were suspended three feet above the participant's workspace, as shown in Figure 5. This camera setup provided a visible range of the workspace of 43 inches by 24 inches. A second stereo camera was placed behind the robot to track the participant's body and face.

B. Architecture

The architecture for our model involved four modules: *vision*, *listening*, *dialogue*, and *control*. The vision and listening modules capture and process their respective input channels. The control module uses input from these modules to decide the need for repair and relays the status of the workspace to the dialogue module if feedback from the robot is needed.

The pipe-assembly task used in our implementation involves multiple copies of five types of pieces: three types of pipes (short, medium, and long) and two types of joints (elbow and t-joints). All pieces were marked with augmented reality (AR) tags to allow detection by the workspace camera. The orientation of each tag was used to identify object type, location, and rotation. The location and orientation of tags on pipes and joints were consistent across each type of object, and tag locations on each object were known to the system.

1) Vision Module: The vision module was designed to achieve two goals: to detect the status of the participant's workspace and to process information on the participant's location. Sensing necessary for achieving each of these goals is managed by a separate camera.

At each frame, the vision module processes the frame to discover which pipes are connected, creating a graph of pipe connections, C. There are three main instructions to building C: finding the AR-tag glyphs in the frame, associating those glyphs with pieces, and detecting which pieces are connected based on a set of heuristics. The description of these instructions are omitted due to space limitations.

At the completion of the participant's turn, C is checked against the correct workspace configuration, C^*. If the two graphs are isomorphic—identical in structure—then the participant has successfully completed the instruction. If the graphs are not isomorphic, then the robot will discover an inconsistency between the two graphs during the isomorphism check. The lowest p_k^* which is violated is then passed to the control module. In those cases where the system needs to check multiple instructions at once, the graph C is built incrementally by systematically eliminating possibly extraneous pieces and then comparing against C^*.

The second goal of the vision module—detecting the participant's location—is checked at every frame. When the

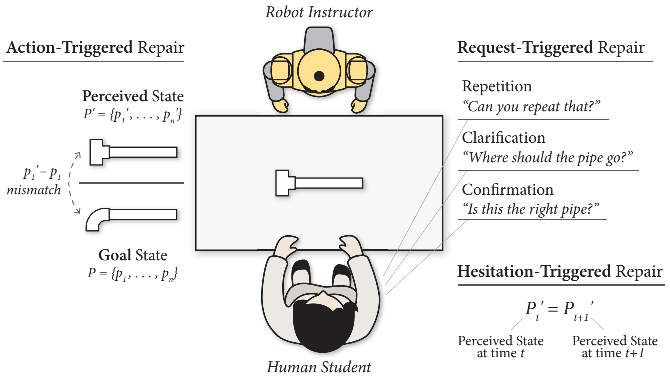


Fig. 4. Examples of the three types of repair. In *action-triggered repair*, the student's configuration of pieces does not match what the robot knows to be the correct configuration. *Request-triggered repair* is initiated when the student directs a question or statement to the robot that requires the robot to respond appropriately. In *hesitation-triggered repair*, the workspace remains unchanged for more than 10 seconds, prompting the robot to offer assistance.

participant is within 1 ft. of the workspace, the robot repositions its head so that it is gazing at the table, monitoring the workspace. When the participant is further away (e.g., standing back to check their work, retrieving the piece), the robot raises its head and gazes toward the participant's face. However, if the participant or the robot is talking, or if the robot is checking the workspace in response to a prompt from the user, the robot looks toward the participant or where on the workspace changes have been made, respectively.

2) Listening Module: The listening module detects and categorizes requests from the participant into semantic meanings using the capabilities of the Microsoft Kinect sensor and speech-recognition API. We provided the API with a grammar that included speech acts from our data on human-human instruction that we marked as one of the following semantic meanings:

- *Request for repetition*: (e.g., "What did you say?" "Can you repeat the instructions?")
- *Check for correctness*: (e.g., "Is this the right piece?" "I'm done attaching the pipe.")
- *Check for options*: (e.g., "Which pipe do I need?" "Where does it go?")

Utterances that did not belong to one of these categories, such as confirmation of an instruction, were ignored by the system.

We use a dialogue manager to coordinate responses to each type of query. Each recognized utterance has an associated semantic meaning that indicates the purpose of the utterance. For example, the phrase "What did you say?" is assigned the semantic meaning of "recognition request." These semantic meanings allow the control module to understand the type of utterance processed and to reply to the utterance appropriately given the current state of the participant's workspace. To process requests that refer to the workspace, the system first checks the state of the workspace through the vision module. For example, asking "Did I do this right?" requires the robot to determine whether the current workspace is correct.

3) Control Module: Decisions on the robot's next action are determined by the control module. It uses input from the vision and dialogue modules and, following a simulation-theoretic approach, makes decisions by comparing this input to actions that the robot expects in response to its instructions. According to our model, we define a set P that describes which possible expectations can be violated by the participant. Consistent with elaboration theory, ordering of task expectations are based on observations from our study of human instructor-trainee interactions, which resulted in the following categories:

- *Timely Action* (p_0): The participant acted in a timely fashion.
- *Correct Piece* (p_1): The participant used the correct piece.
- *Correct Placement* (p_2): The participant placed the piece in the correct location relative to the current workspace.
- *Correct Rotation* (p_3): The participant rotated the piece correctly relative to the current workspace.

The first expectation ensures that the participant does not hesitate for too long, which might indicate confusion, when adding the next piece. Based on our previous analysis, we considered a 10-second delay in changing the workspace after the last instruction to indicate hesitancy. The remaining expectations ensure that the participant chooses the correct piece to add, adds the piece in the correct location, and rotates the piece correctly. Figure 4 illustrates p_1, p_2, and p_3.

4) Dialogue Module: After evaluating input from the vision and listening modules, the control module passes three pieces of information to the dialogue module: current instruction, the semantics associated with the speaker's last utterance (if any), and the control module's evaluation of the workspace (if any).

Given this information, the dialogue module initiates the appropriate verbal response, choosing from among predefined dialogue acts based on which task instruction the participant is completing, the current layout of the workspace, and the type of question the participant asked. Not all responses depend on all three pieces of information; for example, requests for repetition of the last instruction are independent of how the workspace is currently configured, and responses to hesitancy are independent of the current workspace and interaction with the participant. However, a request to check if an instruction has been correctly completed requires knowledge of both the instructions completed and the current layout of the workspace.

V. Evaluation

To evaluate the effectiveness of the strategies that we identified from our analysis in human-robot instruction, we conducted a study that followed the same task setup as our modeling study. Due to a lack of sufficient theory that would predict the effects of these instructional strategies on trainee performance and experience, we chose not to pose any hypotheses and performed an exploratory evaluation.

A. Study Design

To assess the effectiveness of and tradeoffs between various teaching strategies, we designed a between-participants-design study to compare four different models of teaching strategies that fell along two factors: *grouping* and *summarization*. Grouping defines how many instructions are issued during the instructor's turn. For the purposes of our study, grouping has two levels: *no grouping*, where a single instruction is given during the round, and *grouping*, where a set of two or more instructions are given at once. Summarization defines whether or not the instructor gives a summary of the objective of the next few instructions. In our study, we created two levels of this

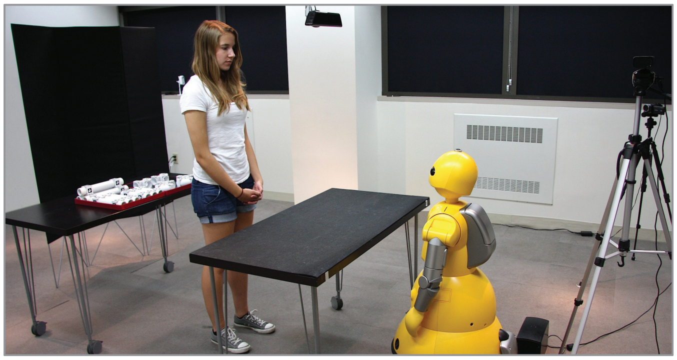


Fig. 5. The setup used in our experimental evaluation. After the robot gave an instruction, the participant retrieved the necessary pieces from behind them and assembled the pieces on the workspace in front of the robot. A camera above the workspace captured the configuration of the pieces.

factor: *no summarization*, where the instructor does not give summaries, and *summarization*, where the instructor offers summaries. We observed the instructor-trainee pairs in our modeling study to exhibit all four combinations of these two factors and created four conditions for our study: (1) *no grouping, no summarization*, (2) *grouping, no summarization*, (3) *no grouping, summarization*, and (4) *grouping, summarization*.

The architecture detailed in the previous section was used in all conditions. Differences between conditions were controlled in the control module that managed decisions on how to structure instructions. Additionally, the dialogue module responded to requests in the grouping level that did not exist in the no grouping level (e.g., repeating multiple instructions).

B. Task

All participants were autonomously guided through assembling a set of pipes by the robot in the setup shown in Figure 5. Participants were given two bins—one for pipes and one for joints—that contained only the pieces necessary for completing the task, mimicking the setup in which different types of parts might be kept at a workshop. Following an introduction, the robot directed the participant in the assembly task by issuing instructions according to the condition to which the participant was assigned, varying the number of instructions provided and whether or not high-level summaries of future instructions were provided. The robot also provided repair as necessary. Following completion of the task, the robot thanked the participant. Completing the task took between 3:57 and 9:20 minutes ($M = 6:44$, $SD = 1:23$).

In the *no grouping, no summarization* and *no grouping, summarization* conditions, the robot provided one instruction at a time, while the grouping condition involved two to four instructions at a time. Additionally, in the *no grouping, summarization* and the *grouping, summarization* conditions, the robot provided a high-level summary of the next few steps prior to giving instructions, while it provided no summary in the other conditions. Following instructions, the participant retrieved the pieces to complete the steps and assembled the pieces on the table. If the participant requested repetition or clarification, the robot answered. When the participant asked the robot to check the workspace, it confirmed correct actions or provided repair according to our model. If no repair was needed, it congratulated the participant on completing the task and proceeded to the next instruction or set of instructions.

The resulting pipe-structure included a total of 15 connected pipes and joints. While the resulting structure was a tree that had no cycles, it had no predefined “root” piece, making the computational complexity of checking for isomorphism against the correct structure an NP-hard problem. We significantly reduced the runtime of this operation by exploiting domain knowledge in our data structure in the form of an incidence matrix of connected joints versus pipes. Once all the pipes were connected, checking for graph isomorphism required approximately 10K permutations of the incidence matrix—far fewer than the hundreds of trillions of checks required without knowledge of the incidence matrix.

C. Procedure

Following informed consent, participants were guided into the experiment room. The experimenter explained the task and introduced the participant to the pieces used in the task. After the experimenter exited the room, the robot started the interaction by explaining that it would provide step-by-step instructions for assembling the pipes. The robot then provided instructions until the participant completed the entire structure. At the end of the task, the robot thanked the participant. The participant then completed a questionnaire and received $5.

D. Participants

A total of 32 native English speakers between the ages of 18 and 34 ($M = 23$, $SD = 4.9$) were recruited from the local community. These participants had backgrounds in a range of occupations and majors. All conditions were gender balanced.

E. Measures & Analysis

We used two objective measures to evaluate participant performance in the task: number of breakdowns and task time. Number of breakdowns was defined as the number of times the participant made a mistake in fulfilling an instruction or asked for repetition or clarification of the instruction. We also measured task completion time, expecting a lower number of repairs to indicate a faster task time. These measures were coded from video recordings of the trials. To ensure reliability of the measures, a second experimenter coded for repairs. The inter-rater reliability showed substantial agreement (87% agreement, Cohen's κ = .83) [17].

We also used subjective measures that collected data on the participant's impressions of the robot, including likability, naturalness, and competency, the participant's experience with the task, and their rapport with the robot. Participants rated each item in our scales using a seven-point rating scale. A confirmatory factor analysis showed high reliability for all scales, including the likability (10 items, Cronbach's $\alpha = .846$), naturalness (6 items, Cronbach's $\alpha = .842$), and competency of the robot (8 items, Cronbach's $\alpha = .896$) and participant experience (8 items, Cronbach's $\alpha = .886$) and rapport with the robot (6 items, Cronbach's $\alpha = .809$).

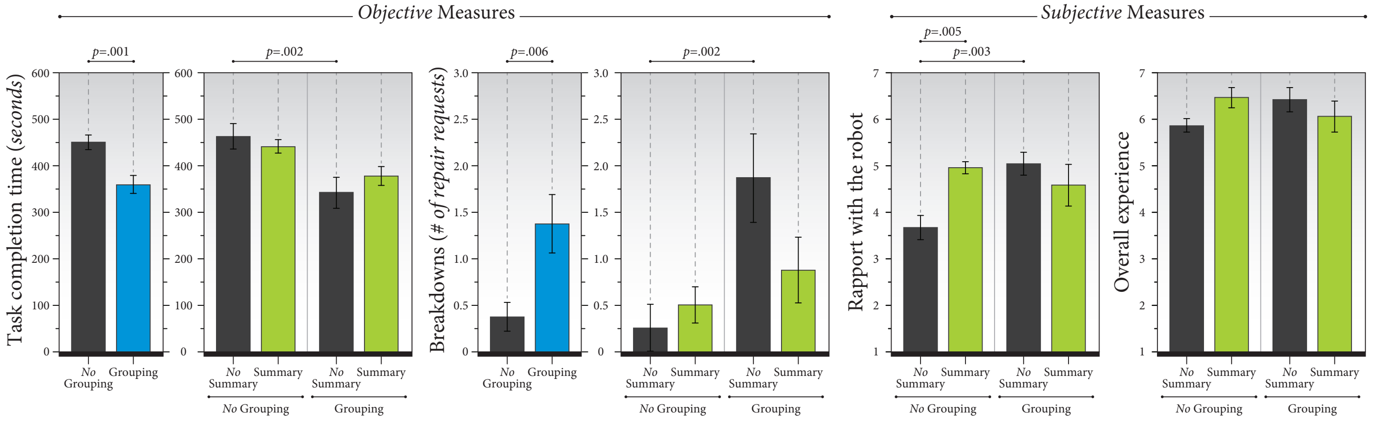


Fig. 6. Results from our evaluation. Significant and marginal results were found for total task time, number of breakdowns encountered, participants' perceived rapport with the robot, and their overall experience with the task.

Our analysis of data from these measures involved a two-way analysis of variance (ANOVA), including grouping, summarization, and the interaction between them as fixed-effect factors. For main and interaction effects, we used α levels of .050 and .10 for significant and marginal effects, respectively. We conducted four contrast tests to understand the effects of each factor in the absence or presence of the other factor using a Bonferroni-adjusted α level of .0125 (.05/4) for significance.

F. Results

We primarily report marginal and significant effects of the instructional strategies used by the robot on objective and subjective measures and summarize them in Figure 6.

To ensure that possible errors in the robot's autonomous behavior did not negatively affect participant evaluations, we examined video recordings of the study for mistakes by the system. Our criteria for removing data included 1) whether or not the robot offered incorrect instruction or repair and 2) whether or not the robot failed more than once to understand a single speech act by the participant. Our examination found no instances of system error regarding the configuration of the pipes in the instructions it gave or the repair it offered, indicating no instances of an incorrect instruction or repair. While the robot failed to understand 21% of the participants at least once during their entire interaction, no single speech-act was misunderstood more than once, as participants either more clearly reiterated or rephrased their statement.

To evaluate the effectiveness of the instructional strategies, we measured the number of breakdowns that occurred during the task and the time taken to complete the task. The analysis of this data showed that grouping instructions significantly reduced task completion time, $F(1,28) = 13.35$, $p = .001$, $\eta^2 = .313$, while significantly increasing the number of breakdowns, $F(1,28) = 8.87$, $p = .006$, $\eta^2 = .213$. Summarization had no overall effect on task time, $F(1,28) = 0.07$, $p = .793$, $\eta^2 = .002$, or the number of breakdowns, $F(1,28) = 1.25$, $p = .274$, $\eta^2 = .030$. The analysis also showed a marginal interaction effect between grouping and summarization over the number of breakdowns, $F(1,28) = 3.47$, $p = .073$, $\eta^2 = .083$, but no interaction effects were found over total task time, $F(1,28) = 1.29$, $p = .266$, $\eta^2 = .030$. Contrast tests across conditions showed that, when the robot did not provide a summary, grouping instructions significantly reduced task completion time, $F(1,28) = 11.47$, $p = .002$, $\eta^2 = 269$, but resulted in a significant increase in the number of breakdowns, $F(1,28) = 11.71$, $p = .002$, $\eta^2 = .282$.

The subjective measures captured the participants' perceptions of the robot, including likability, naturalness, and competency, their rapport with the robot, and their overall experience with the task. The analysis showed an interaction effect between grouping and summarization over the participants' rapport with the robot, $F(1,28) = 8.76$, $p = .006$, $\eta^2 = .211$. When the robot provided no summary, grouping instructions improved participant rapport with the robot, $F(1,28) = 10.81$, $p = .003$, $\eta^2 = .260$. When the instructions were not grouped, summarization also improved rapport with the robot, $F(1,28) = 9.54$, $p = .005$, $\eta^2 = .230$. Consistent with the results on participant rapport, we also found a marginal interaction effect between grouping and summarization over participants' ratings of their overall experience with the task, $F(1,28) = 3.68$, $p = .065$, $\eta^2 = .115$.

VI. Discussion

The data from our objective and subjective results provided a number of findings to guide the design of instructional robots, the implications of which we highlight below.

Our objective results showed that grouping instructions resulted in a tradeoff between task completion time and the number of breakdowns that the participants encountered. We found that participants completed the task significantly faster when the robot grouped its instructions than when the robot provided instructions one-by-one. We observed that when participants received multiple instructions, they retrieved all parts necessary to complete these instructions from the bins at once, proceeded with assembling multiple pieces in a sequence, and sought confirmation of the correctness of the whole sequence from the robot, completing the overall assembly significantly faster. When participants received instructions one-by-one, they instead retrieved pieces one-by-one and proceeded to the next instruction only when the robot confirmed the successful completion of an assembly, which resulted in overall longer task completion times. Contrary to the improvement in task completion times, participants encountered significantly more breakdowns when the robot grouped its instructions than when the robot provided individual instructions. We speculate that grouped instructions required participants to retain a greater amount of information, which might have impaired their understanding or recall of the instructions, resulting in mistakes in the assembly that had to be repaired by the robot.

Further analysis into breakdowns that occurred with grouped instructions showed that 60% of breakdowns occurred in the first set of instructions, which contained four instructions, 25% occurred in the second, third, and fifth set of instructions, which all contained three instructions, and 15% occurred in the fourth set of instructions, which contained two instructions. This distribution of breakdowns indicates an increase in the number of breakdowns as the number of grouped instructions increases, which might indicate a greater cognitive load placed on the participant by the introduction of more pieces into an instruction [23]. Additionally, participants may have demonstrated selective attention when the robot provided grouped instructions, causing them to miss information [23]. Our data on the number of breakdowns provided limited support for this explanation; in carrying out grouped instructions, participants encountered fewer breakdowns when the robot provided a summary of subsequent steps ($M = 0.88$, $SD = 0.99$) than when no summary was provided ($M = 1.88$, $SD = 1.36$), although this effect was not significant at α level .0125. The summary provided by the robot might have consolidated the participants' understanding of the grouped instructions. However, some of the breakdowns that occurred early in the interaction may have been caused by the participant acclimating to the task or the task involving a greater variety of pieces to choose from at the beginning.

Our analysis of the subjective measures showed a significant interaction effect between grouping and summarizing on participant rapport with the robot. We found that participants reported higher rapport with the robot when it grouped instructions with no summary than when the robot used neither grouping nor summarization. This improvement might be due to the quicker, less monotonous experience that the robot offered when it delivered instructions all at once and spent no time on summarizing them. The results also showed that participants reported higher rapport with the robot when the robot provided a summary of subsequent steps along with individual instructions than when it neither grouped its instructions nor provided a summary. Consistent with the interaction effect on participant rapport with the robot, we also found a marginal interaction effect between grouping and summarizing on their overall experience with the task, although the contrast tests did not show significant differences at α level .0125. We speculate that, when the robot provided a summary of what was ahead in the task, as a summary involved information on upcoming steps, participants might have felt more informed and perceived the robot as more invested, although this information did not improve task performance.

Design Implications: These results have a number of implications for the design of instructional robots. Our results suggest that, despite resulting in more mistakes, grouping significantly improves task completion times, making it ideal for settings in which faster task completion are critical and mistakes are not costly. Furthermore, coupling summarization with grouping alleviates some of the mistakes caused by providing multiple instructions at once. However, there are many scenarios where providing instructions one-by-one might be preferable. For example, with more complex tasks or students who might have trouble keeping up with the robot's instructions (e.g., novices), providing instructions one-by-one might help the student complete the task with fewer breakdowns. Additionally, in situations where mistakes could be dangerous or costly, individual instruction might reduce the chance of these mistakes occurring. In these scenarios, including summaries of upcoming instructions might also improve student rapport with the robot.

Limitations: The work presented here has three key limitations. First, although our model considers two structural components of instruction-giving, there may be other components we did not observe in our modeling study and thus did not include in our model. Analyses of human interactions in a more diverse set of instructional scenarios may enable the development of richer models of instruction. Second, while our repair model offered repair when prompted, the system did not proactively offer repair due to the difficulty of accurately discerning when mistakes occurred. The structure of the task and available methods for perception made it difficult to continuously update a model of the workspace and determine whether it was being modified, as participants obstructed the camera's view when modifications were occurring. Third, our evaluation focused on testing only the immediate effects of the proposed instructional strategies on student performance and perceptions. We plan to extend our work to explore a more diverse set of instructional scenarios, instructions that are distributed over time, and long-term effects of the proposed strategies on task-based instruction.

VII. Conclusion

As robots move into roles that involve providing users with task guidance, such as teaching in labs and assisting in assembly, they need to employ strategies for effective instruction. In this paper, we described two key instructional strategies—grouping and summarization—based on observations of human instructor-trainee interactions in a pipe-assembly task. We implemented these strategies on a robot that autonomously guided its users in this task and evaluated their effectiveness in improving trainee task performance and experience in human-robot instruction. Our results showed that, when the robot grouped instructions, participants completed the task faster but encountered more breakdowns. We also found that summarizing instructions increased participant rapport with the robot. Our findings show that grouping instructions results in a tradeoff between task time and breakdowns and that summarization has some benefits under certain conditions, suggesting that robots selectively use these strategies based on the goals of the instruction.

Acknowledgments

We thank Brandi Hefty, Jilana Boston, Ross Luo, Chien-Ming Huang, and Catherine Steffel for their contributions to and National Science Foundation award 1149970 and Mitsubishi Heavy Industries, Ltd. for their support for this work.

References

[1] L. Alfieri, P.J. Brooks, N.J. Aldrich, and H.R. Tenenbaum. Does discovery-based instruction enhance learning? *Journal of Educational Psychology*, 103(1):1–18, 2011.

[2] S. Andrist, E. Spannan, and B. Mutlu. Rhetorical robots: making robots more effective speakers using linguistic cues of expertise. In *Proc. HRI'13*, pages 341–348, 2013.

[3] E. Bicho, W. Erlhagen, L. Louro, and E. Costa e Silva. Neuro-cognitive mechanisms of decision making in joint action: A human–robot interaction study. *Human Movement Science*, 30(5):846–868, 2011.

[4] N. Blaylock, J. Allen, and G. Ferguson. Managing communicative intentions with collaborative problem solving. In Kuppevelt J.C.J. and R.W. Smith, editors, *Current and New Directions in Discourse and Dialogue*, pages 63–84. Springer, 2003.

[5] J.-D. Boucher, U. Pattacini, A. Lelong, G. Bailly, F. Elisei, S. Fagel, P.F. Dominey, and J. Ventre-Dominey. I reach faster when I see you look: gaze effects in human–human and human–robot face-to-face cooperation. *Frontiers in Neurorobotics*, 6, 2012.

[6] M.E. Foster, M. Giuliani, A. Isard, C. Matheson, J. Oberlander, and A. Knoll. Evaluating description and reference strategies in a cooperative human-robot dialogue system. In *Proc. IJCAI'09*, pages 1818–1823, 2009.

[7] V. Gallese and A. Goldman. Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2(12):493–501, 1998.

[8] B. Gonsior, D. Wollherr, and M. Buss. Towards a dialog strategy for handling miscommunication in human-robot dialog. In *Proc. RO-MAN'10*, 2010.

[9] J. Gray, C. Breazeal, M. Berlin, A. Brooks, and J. Lieberman. Action parsing and goal inference using self as simulator. In *Proc. RO-MAN'05*, 2005.

[10] B.J. Grosz and S. Kraus. Collaborative plans for complex group action. *Artificial Intelligence*, 86(2):269–357, 1996.

[11] B.J. Grosz and C.L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3): 175–204, 1986.

[12] G. Hirst, S. McRoy, P. Heeman, P. Edmonds, and D. Horton. Repairing conversational misunderstandings and non-understandings. *Speech Communication*, 15(3): 213–229, 1994.

[13] J. Hoey, P. Poupart, C. Boutilier, and A. Mihailidis. POMDP models for assistive technology. In *Proc. AAAI 2005 Fall Symposium*, 2005.

[14] C.-M. Huang and B. Mutlu. Robot behavior toolkit: generating effective social behaviors for robots. In *Proc. HRI'12*, pages 25–32, 2012.

[15] T. Kanda, R. Sato, N. Saiwaki, and H. Ishiguro. A two-month field trial in an elementary school for long-term human–robot interaction. *IEEE Transactions on Robotics*, 23(5):962–971, 2007.

[16] T. Koulouri and S. Lauria. Exploring miscommunication and collaborative behaviour in HRI. In *Proc. SIGDIAL'09*, 2009.

[17] J.R. Landis and G.G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977.

[18] M.K. Lee, S. Kiesler, J. Forlizzi, S. Srinivasa, and P. Rybski. Gracefully mitigating breakdowns in robotic services. In *Proc. HRI'10*, 2010.

[19] M.N. Nicolescu and M.J. Mataric. Linking perception and action in a control architecture for human-robot domains. In *Proc. HICSS'03*, 2003.

[20] C.M. Reigeluth, M.D. Merrill, B.G. Wilson, and R.T. Spiller. The elaboration theory of instruction: A model for sequencing and synthesizing instruction. *Instructional Science*, 9(3):195–219, 1980.

[21] P. Seedhouse. The relationship between context and the organization of repair in the l2 classroom. *International Review of Applied Linguistics in Language Teaching*, 37 (1):59–80, 1999.

[22] M. Staudte and M.W. Crocker. Visual attention in spoken human-robot interaction. In *Proc. HRI'09*, pages 77–84, 2009.

[23] J. Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2):257–285, 1988.

[24] F. Tanaka and J.R. Movellan. Behavior analysis of children's touch on a small humanoid robot: Long-term observation at a daily classroom over three months. In *Proc. RO-MAN'06*, 2006.

[25] R. Tanaka and T. Kimura. The use of robots in early education: a scenario based on ethical consideration. In *Proc. RO-MAN'09*, 2009.

[26] C. Torrey, A. Powers, M. Marge, S.R. Fussell, and S. Kiesler. Effects of adaptive robot dialogue on information exchange and social relations. In *Proc. HRI'06*, pages 126–133, 2006.

[27] C. Torrey, A. Powers, S.R. Fussell, and S. Kiesler. Exploring adaptive dialogue based on a robot's awareness of human gaze and task progress. In *Proc. HRI'07*, 2007.

[28] J.G. Trafton, N.L. Cassimatis, M.D. Bugajska, D.P. Brock, F.E. Mintz, and A.C. Schultz. Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 35(4):460–470, 2005.

[29] C.J. Zahn. A reexamination of conversational repair. *Communications Monographs*, 51(1):56–66, 1984.