Anticipatory Robot Control for Efficient Human-Robot Collaboration

Chien-Ming Huang*
Department of Computer Sciences
University of Wisconsin–Madison
cmhuang@cs.wisc.edu

Bilge Mutlu
Department of Computer Sciences
University of Wisconsin–Madison
bilge@cs.wisc.edu

Abstract—**Efficient collaboration requires collaborators to monitor the behaviors of their partners, make inferences about their task intent, and plan their own actions accordingly. To work seamlessly and efficiently with their human counterparts, robots must similarly rely on predictions of their users' intent in planning their actions. In this paper, we present an *anticipatory control* method that enables robots to proactively perform task actions based on anticipated actions of their human partners. We implemented this method into a robot system that monitored its user's gaze, predicted his or her task intent based on observed gaze patterns, and performed anticipatory task actions according to its predictions. Results from a human-robot interaction experiment showed that anticipatory control enabled the robot to respond to user requests and complete the task faster—2.5 seconds on average and up to 3.4 seconds—compared to a robot using a *reactive control* method that did not anticipate user intent. Our findings highlight the promise of performing anticipatory actions for achieving efficient human-robot teamwork.**

Index Terms—**Action observation, gaze, intent prediction, anticipatory action, human-robot collaboration**

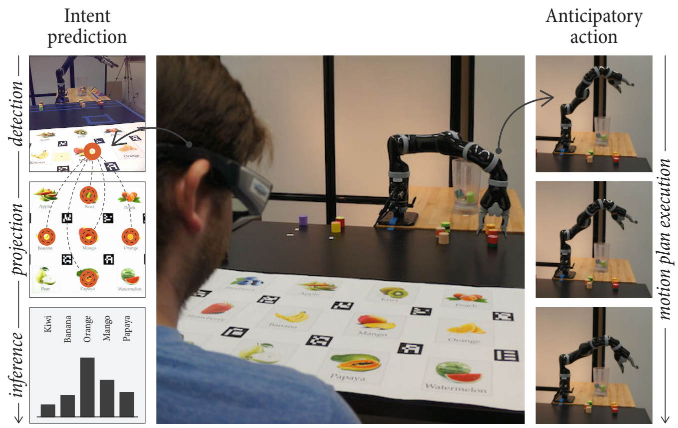

Fig. 1. We propose an "anticipatory control" method that enables robots to proactively plan and execute actions based on an anticipation of a human partner's task intent as inferred from their gaze patterns.

I. Introduction

Efficient teamwork requires seamless and tight coordination among collaborators. In order to achieve such coordination, collaborators must not only be aware of each other's actions, but they must also *anticipate* the actions of their partners and proactively plan their own actions [1]. This anticipatory planning is achieved by observing the behaviors of others, anticipating future actions based on these observations, and preparing one's own actions according to anticipated actions [2], [3]. Gaze behavior is a critical source of information about task intent [4], a predictor of motor actions [5], [6], and a facilitator in a range of important social functions from enabling shared attention [7] to performing joint tasks [8].

Prior research in human-robot interaction has demonstrated how monitoring user behaviors can help robots anticipate user actions [9] and how anticipatory robot actions can enhance the safety of the collaboration [10] and improve task efficiency by reducing user idle time [11], [12]. These studies illustrate the promise that anticipation and proactive actions hold for improving human-robot collaboration. This paper explores a specific mechanism for anticipatory action that enables a robot to monitor the covert gaze patterns of its user, infer user task intent based on these patterns, and engage in proactive task actions in order to achieve a more seamless collaboration.

This work makes three specific contributions to research in human-robot interaction: (1) an "anticipatory control" *method* for robots to proactively plan and perform goal-directed actions based on an anticipation of the task intent of a human collaborator; (2) a *system* implementation of this method into an autonomous robot that integrated real-time tracking of gaze, prediction of task intent based on a trained model, and on-the-fly planning of robot motions; and (3) *data* on the effects of anticipatory robot action on human-robot collaboration as well as *insights* into design and technical challenges involved in realizing anticipatory control. These contributions inform the development of robot systems for settings such as manufacturing plants that require highly coordinated teamwork.

In the remainder of this paper, we first review prior work on action observation and intent prediction as well as anticipatory robot actions (Section II). We then present our anticipatory control method and its implementation into an autonomous robot system in Section III and describe the design of and findings from a human-robot interaction study that evaluated the system in Section IV. Finally, we conclude with a discussion of the findings and limitations of our work (Section V).

* The author is currently with the Department of Computer Science, Yale University and can be reached at chien-ming.huang@yale.edu.

II. Background

Collaboration requires that the parties involved employ a set of cognitive and communicative mechanisms to coordinate their actions toward a shared goal. Robots that are designed to collaborate with people must similarly utilize these mechanisms to coordinate their actions with their human counterparts. The paragraphs below provide a brief review of research literature on relevant cognitive and communicative mechanisms that support human collaboration and of prior explorations of how such mechanisms may facilitate human-robot teamwork.

A. Action Observation and Intent Prediction

A key facilitator of collaboration is *action observation*, a process in which collaborators monitor the actions of their partners in order to understand their goals and predict what they will do next [3], [13]. In this process, individuals map the observed actions of others onto their own motor representation of the same actions [14], [15], which enables them to proactively prepare their own goal-directed actions [1]. Gaze serves as a particularly critical source of information to signal task intent. Interaction partners expect that an area being gazed toward in the task space is the next space to be acted upon [4]. Awareness of partner's gaze facilitates task coordination [16] and improves efficiency in collaboration [17].

Relevant prior work on the relationship between gaze and intent includes computational models that aim to predict task intent from gaze cues. For example, the future actions of a driver operating a motor vehicle can be predicted from the driver's gaze cues using sparse Bayesian learning [18]. Gaze can also predict a performer's task state while making a sandwich using a dynamic Bayesian network [19]. In a collaborative sandwich-making task, the requester's task intent can be inferred from their gaze patterns using an SVM-based classifier [20].

B. Anticipatory Robot Action

Prior research in human-robot interaction has explored how robots may predict the intent and anticipate the actions of their users in order to serve as effective collaborators. This work includes the development of novel methods for goal inference from observed human actions by mapping the observed actions to a robot's action repertoire [21] and by coupling these observations with object affordances [9]. Researchers have also proposed novel computational representations that enable robots to anticipate collaborative actions in the presence of uncertainty in sensing and ambiguity in task states, demonstrating robust anticipation through an integration of all available sensor information with a knowledge of the task [12].

Previous work also includes the development of several robot systems that utilize anticipation of user actions to improve human-robot collaboration. For instance, a robot system designed to engage in co-located collaborations with humans observed the motions of its human partners to predict workspace occupancy and planned its motion accordingly in order to minimize interference with them [10]. Another robot system observed the reaching motion of its human counterparts, predicted the intended reach target, and used this prediction to selectively reach toward to a different target [22]. Anticipation of actions enabled a virtual robot to adapt to its user's workflow in a simulated assembly scenario, improving the fluidity of collaboration [11]. Finally, a robot system that was developed to provide shoppers in a shopping mall with information was able to approach shoppers effectively by anticipating their walking behavior based on walking trajectories and velocities [23].

Additionally, prior work includes studies that link gaze and task intent, including the development of a robot system that predicted the intent of its users based on their motions and used these predictions to determine where it should look in the environment [24]. This linking not only directed the robot's attention toward the task-relevant parts of the environment but it also signaled shared attention to human partners. Previous research has also studied how people could utilize the gaze cues of a robot to understand its intent and how this understanding might facilitate efficient cooperation [25].

While research in human-robot interaction highlights the promise of predicting user intent and performing anticipatory actions for facilitating human-robot collaboration, how robots may draw on the gaze patterns of their users to predict and act according to user task intent and what specific effects anticipatory robot actions may have on human-robot teamwork remain unexplored. In the next section, we describe a novel "anticipatory control" method that seeks to close this gap.

III. Enabling Anticipatory Control

We propose an *anticipatory control* method that involves monitoring user actions, predicting user task intent, and proactively controlling robot actions according to predicted user intent as an alternative to *reactive control* methods that utilize direct, explicit user input. In this section, we present the implementation of this method as a real-time autonomous robot system following a sense-plan-act paradigm. To provide context for the development and implementation of our proposed method, we devised a task in which a robot works as a "server" preparing smoothies for a human "customer" that represents interactions common in day-to-day collaborations.

The proposed method integrated six components: (1) gaze tracking, (2) speech recognition, (3) intent prediction, (4) anticipatory motion planning, (5) speech synthesis, and (6) robotic manipulation. Figure 2 illustrates how these components are integrated by the implemented system, and the sections below provide detail on their functioning and implementation.

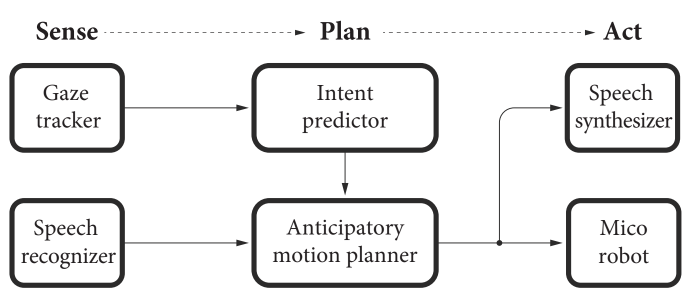

Fig. 2. Components of our anticipatory robot system.

A. Gaze Tracking

The gaze-tracking component captured gaze fixations from a pair of SMI Eye-Tracking Glasses V.1[1] worn by the user. It then performed a *projective transformation* using the Jacobi method to map gaze fixations in the camera-view space to locations in the physical task space. These points were subsequently used to infer what task-relevant items were being looked toward. The mapping between the camera-view space and physical space and the association between locations in the physical space and environmental items were realized by locating a set of predefined Aruco markers.[2]

B. Speech Recognition & Synthesis

Microsoft Speech API 5.4[3] was used to build a speech-recognition component to recognize user utterances and a speech-synthesis component to realize the robot's speech. A flexible recognition grammar was specified to minimize speech recognition errors and to accommodate different verbalizations of user requests, such as "I would like to have mango," "Could I have papaya," or simply "Peach." The robot's speech included greetings, confirmations of user requests, such as "You ordered mango," task instructions, such as "Next one," and a "Thank you" remark uttered at the end of the interaction.

C. Robotic Manipulation

A six-degree-of-freedom Kinova MICO robot arm[4] was used as the manipulator to pick up the requested items and place them at a target location, which in the context of our task involved placing smoothie ingredients into a blender. The arm was controlled using the `MoveIt!` platform[5] and was given a clear representation of the environment for motion planning.

D. Intent Prediction

The intent-prediction component built on our existing framework for predicting user task intent based on gaze patterns using a support vector machine (SVM) [20]. In the development of this framework, we devised a collaborative sandwich-making scenario in which a "server" added ingredients requested by a "customer" and aimed to predict which ingredient the customer would choose next based on his or her gaze patterns. To train the SVM, we collected data from 276 episodes of human interactions following this scenario and used four features—number of glances, duration of the first glance, total duration, and most recently glanced item—as predictors of the intended ingredient. An offline cross-validation analysis showed that the trained SVM predicted user intent based on gaze patterns approximately 1.8 seconds prior to verbal requests with reasonable accuracy (76%).

To create a real-time intent-prediction component, we used the entire 276-episode dataset to train an SVM classifier that predicted user task intent based on gaze features extracted from the history of the items toward which the user has looked. Using the four features described above as input, the classifier provided the ID number of the ingredient that the system predicted to be the item that the user would request next and a score for the confidence of the classifier in its prediction.

E. Anticipatory Motion Planning

Using the `MoveIt!` platform, the anticipatory motion planner utilized the prediction and confidence value that the intent-prediction component provided to proactively plan and execute motion toward the predicted item (Algorithm 1). If the confidence of the prediction was higher than *planThreshold*, set to 0.36, the motion planner planned a motion toward the predicted item. If the confidence was higher than *execThreshold*, set to 0.43, it executed only a part of the planned motion based on its current confidence (see the description of the *splitPlan* method below). Taken from our prior work [20], these thresholds indicate that if the confidence of a prediction is higher than 0.36, the prediction could be correct, and that if it exceeds 0.43, the prediction was unlikely to be incorrect.

Instead of using the current prediction and confidence, denoted, respectively, as *currPred* and *currProb*, directly from the intent-prediction component, the anticipatory motion planner maintained a history of the 15 latest predictions, including the current prediction. The gaze-tracking component provided readings at approximately 30 Hz, and thus the length of the prediction history was chosen to be approximately 500 milliseconds. The prediction history was then used to calculate a weighted prediction, p_i', that discounted past predictions using the exponential decay function defined in Equation 1.

$$p_i' = p_i \times (1 - decayRate)^i \tag{1}$$

In this function, p_i denotes the probability of the ith prediction in the history. The *decayRate*, set to 0.25, indicates the rate at which the weight of the prediction decayed, and the resulting prediction (*weightedPred*, i.e., p_i') is the prediction with the highest weight summed over the prediction history.

The anticipatory motion planner maintained a plan library that stored a set of candidate motion plans from which

Algorithm 1 Anticipatory Robot Control

```
Require: currPred, currProb
while true do
    predHistory ← UPDATEPREDHISTORY(currPred, currProb)
    weightedPred, weightedProb ← GETWEIGHTEDPRED(predHistory)
    if weightedProb ≥ planThreshold then
        motionPlan ← RETRIEVEPLAN(weightedPred)
        if (motionPlan = ∅) or (weightedPred ≠ currMotionTarget) then
            MAKEPLAN(weightedPred)
        end if
    end if
    if weightedProb ≥ execThreshold then
        motionPlan ← RETRIEVEPLAN( )
        subPlan1, subPlan2 ← SPLITPLAN(motionPlan)
        REQUESTEXEC(subPlan1)
        UPDATEPLANLIBRARY(weightedPred, subPlan2)
    end if
end while
```

[1] http://www.eyetracking-glasses.com
[2] http://www.uco.es/investiga/grupos/ava/node/26
[3] https://msdn.microsoft.com/en-us/library/ee125077(v=vs.85).aspx
[4] http://www.kinovarobotics.com/service-robotics/products/robot-arms/
[5] http://moveit.ros.org

it chose when the robot had made a prediction of the user's request. The *currMotionTarget* variable denotes the motion target associated with the most recent plan. The *makePlan* function utilized the RRT-Connect algorithm [26] (the `RRTConnectkConfigDefault` planner in `MoveIt!`) to create a motion plan toward the *weightedPred* item. The *splitPlan* function took a motion plan and split it into two sequential sub-plans proportionally based on the confidence of the prediction, denoted as *weightedProb*. Higher confidence values moved the robot closer and closer to the predicted item. Although this iterative planning could bring the robot to a position in which it could grasp the ingredient, we chose to delay the grasp until the user made a verbal request in order to more easily recover from errors.

The implementation of the anticipatory-motion-planning component involved three threads: a *planning* thread that implemented Algorithm 1, an *execution* thread that executed motion plans, and a *speech* thread that processed user requests. The planning thread put a motion request into a plan queue using the *requestExec* function. The execution thread regularly checked the queue of plans and executed them. When processing a verbal request, the speech thread checked if the robot's current motion target—if it had one—matched the user's request. If it did, the robot carried out the rest of the motion plan in order to complete the request. Otherwise, it stopped the current motion and made a new plan, directing motion toward the item requested by the user. We note that anticipatory control was used for determining and reaching toward requested items and not for transporting grasped items to the target location.

F. System Limitations

Our anticipatory robot system had three main sources of error: tracking, projection, and prediction. Tracking errors resulted directly from the eye-tracking system. Even with a state-of-the-art eye-tracking system that was calibrated for each user following the manufacturer-recommended calibration procedure, some amount of tracking error was unavoidable. A second source of error arose from the projection process of gaze fixations provided by the eye tracker to the workspace. Mismatched tracking rates between the eye tracker and the tracker used for Aruco markers led to incorrect inferences regarding which items were gaze targets. Finally, the intent-prediction component provided erroneous predictions partly due to errors that cascaded through the tracking and projection processes and partly due to the limitations of the trained model.

IV. Evaluation

In this section, we describe the design of and findings from a human-robot interaction experiment that evaluated the effectiveness of the proposed anticipatory control method in supporting team performance and user experience.

A. Hypothesis

Our central hypothesis is that *anticipatory* control, as implemented in the robot system described in Section III, would enable the robot to more effectively respond to user requests, thus resulting in improvements in team performance and user perceptions of the robot, compared to other, more *reactive* forms of control.

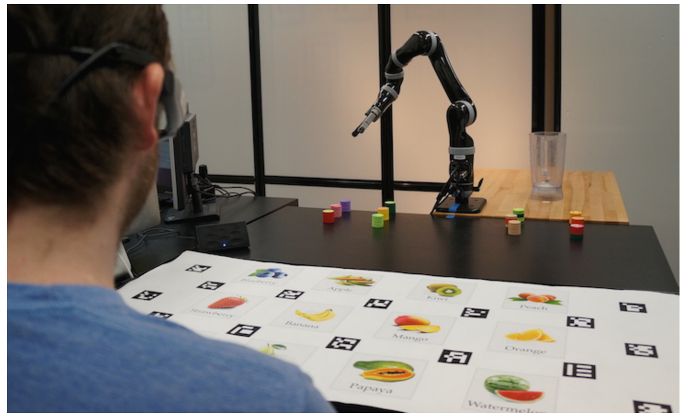

Fig. 3. The setup of the human-robot interaction experiment. Between the robot and the user were a menu for the user from which to select ingredients and a workspace for the robot to prepare the order.

B. Experimental Task, Design, & Conditions

To test our hypothesis, we devised an experimental task in which human participants, acting as "customers," ordered two fruit smoothies from a robot system that served as a "café worker." During the task, participants sat across from the robot with a menu of 12 different fruit choices placed in front of them (Figure 3). Participants were asked to choose a total of five fruits from the menu for each order and to request one fruit at a time using verbal requests.

Two experimental conditions—*anticipatory* and *reactive*—were implemented on the robot system for evaluation. In the *anticipatory* condition, the robot predicted the user's choices and proactively planned and executed its motions based on its prediction, as described in Section III. In the *reactive* condition, the robot responded only to the user's verbal requests.

The experiment followed a within-participants design. The only independent variable was whether or not the robot anticipated user choices before acting on them. Each participant interacted with the robot in both conditions, and the order of conditions was counterbalanced across trials. We designed the experimental task to involve practices people commonly follow in daily interactions that one would expect at a café in order to minimize learning effects and the need for extensive training.

C. Procedure

Upon receiving informed consent, the experimenter provided the participant with an explanation of the task and described how they could interact with the robot. The participant was fitted with head-worn eye-tracking glasses. The experimenter then performed a calibration procedure for eye tracking followed by a verification procedure for gaze projection. In this verification procedure, the experimenter asked the participant to look toward four different ingredients on the menu, one at a time, and to name the ingredient toward which they were looking in order to determine the accuracy of the gaze projection after the eye tracker was calibrated. The participant then followed the robot's instructions to complete a drink order and filled out a questionnaire to evaluate their experience with and perceptions of the robot. This procedure was then repeated for the other condition. After interacting with the robot in both conditions, the experimenter collected demographic information and interviewed the participants for additional comments on differences they may have observed in the robot's behaviors between the two conditions.

D. Measures

We expected the performance of the anticipatory robot system to be affected by the potential errors accumulated throughout the pipeline of tracking the participant's eyes, inferring gaze targets, and predicting participant intent. To gain a more detailed understanding of the effects of these errors on team performance, we employed two system measures: *projection* accuracy and *prediction* accuracy.

Projection accuracy (%): The number of matches between gazed and reported items divided by the total number of items (i.e., four per participant), measured during the gaze-projection-verification procedure.

Prediction accuracy (%): The number of matches between system predictions and user requests divided by the total number of user requests (i.e., five per interaction episode), measured during the experimental task.

To assess the effectiveness of anticipatory and reactive control methods in supporting human-robot collaboration, we utilized a number of objective and subjective measures. Objective measures included *response time* and *time to grasp*.

Response time (milliseconds): The duration between when the participant verbally placed a request and when the robot started moving toward the requested item. For the anticipatory system, this measure captured the time it took to initiate a planned motion if the robot's prediction matched the user's request. Otherwise, it additionally captured the time needed to stop the current motion toward an incorrect prediction and the time to plan and initiate motion toward the correct target. For the reactive system, the measure only captured the time needed to plan and initiate motion toward the requested item as soon as the request was recognized.

Time to grasp (seconds): The duration between when the participant verbally requested an item to when the robot grasped the requested item. This measure was also considered as an approximation of *task time*, as the procedure to transport the grasped item to the target location to complete user requests was the same for both conditions.

In addition to the objective measures described above, we used a questionnaire to assess participants' subjective perceptions of the robot's anticipatory behaviors, particularly its perceived *awareness* and *intentionality*. The awareness scale, consisting of four items (Cronbach's α = 0.74), aimed to measure how aware participants thought the robot was of their intended choices. The intentionality scale, consisting of four items (Cronbach's α = 0.83), aimed to capture participant perceptions of how mindful, conscious, intentional, and intelligent the robot appeared.

Finally, a single item, "The robot only moved to pick up an item after I verbally issued a request," served as a manipulation check, examining whether or not users were able to discern the difference between the anticipatory and reactive systems.

E. Participants

Twenty-six participants were recruited from the local community. Two participants were excluded from the data analysis due to failures in eye tracking or in online motion planning. The resulting 24 participants (16 females, 8 males) were aged between 18 and 32 (*M* = 22.21, *SD* = 4.15). Four participants reported having interacted with a similar robot arm prior to their participation in the current study. The study took 30 minutes, and participants were paid $5 USD.

F. Results

The paragraphs below report on results from our system, objective, and subjective measures. We describe findings from the system measures first in order to provide context for the objective and subjective measures, as they were affected by the potential errors accumulated through the eye-tracking, gaze-projection, and intent-prediction phases.

System measures — The overall *projection accuracy* for our anticipatory system was 81.25%. Incorrectly inferred items were usually immediate neighbors (i.e., above, below, to the left, and to the right) of the intended targets. This accuracy rose to 91.67% if neighbors were considered as correct.

Out of 120 predictions, the anticipatory system made 53 incorrect predictions, yielding 55.83% *prediction accuracy*. However, eight of these incorrect predictions were due to not being able to make any prediction, because the users did not look toward any items on the menu prior to making requests. Additionally, in another 18 trials, participants did not look toward the requested item but rather looked toward other items, resulting in incorrect predictions. Possible explanations for these behaviors are that participants decided on their next ingredient during the previous request, that the eye tracker failed to accurately capture gaze direction, or that gaze projection was erroneous. Our system reached 59.82% prediction accuracy in cases where a prediction was made and 77.5% accuracy if the user had glanced at the intended item. Baseline accuracy (chance) varied between 8.33% (1/12) and 12.5% (1/8).

To analyze the data from the objective and subjective measures, we used one-way repeated-measures analysis of variance (ANOVA) following a linear mixed-models procedure in which *control method*, either anticipatory or reactive, was set as a fixed effect, and *participant* was set as a random effect, as suggested by Seltman [27]. Table I, Table II, and Figure 4 summarize results from this analysis.

Manipulation check — We found significant differences in participant perceptions of when the robot moved toward the requested item across the two conditions (Table II), indicating that participants were able to discern the differences resulting from our experimental manipulation.

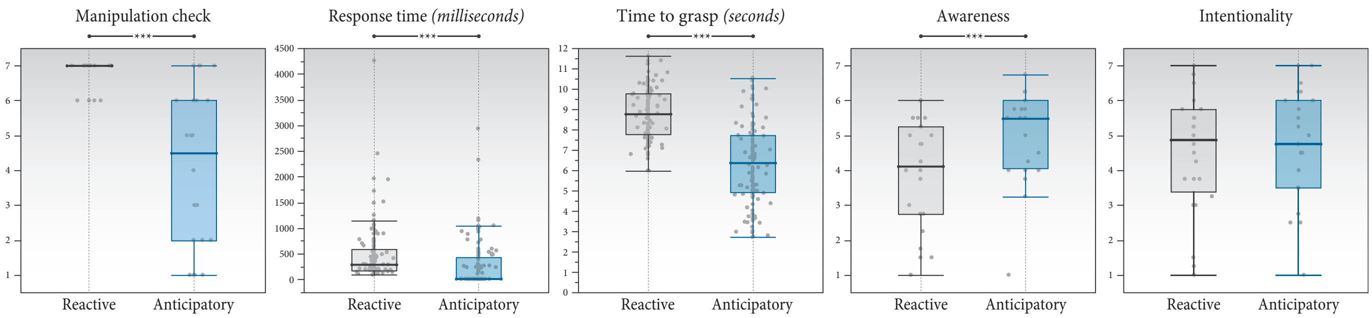

Fig. 4. Tukey boxplots of data from the manipulation check, objective measures, and subjective measures. The extents of the box represent the the first and third quartiles. The line inside the box represents the second quartile (the median). The difference between the first and third quartiles is the interquartile range (IQR). The ends of the whiskers represent the first quartile minus 1.5 times IQR and the third quartile plus 1.5 times IQR. (***) denotes $p < .001$.

Objective measures — Table I, Table II, and Figure 4 provide results from our objective measures, including *response time* and *time to grasp*. We found that anticipatory control enabled the robot to more efficiently respond to and complete participant requests than did reactive control. The average duration for finding and initializing a valid motion plan toward the target item, which corresponded to the response time of the reactive system, was 482.71 ms, indicating a reasonably responsive system in the context of our task. Anticipatory control based on predicted participant intent reduced the response time by 226.3 ms. If the predictions were correct, the response time on average for the anticipatory system was 51.03 ms.

The anticipatory system proactively moved toward the predicted item of choice based on its confidence in the prediction. This proactive execution reduced time to grasp by 2.51 seconds. When predictions were correct (55.83% of the time), the anticipatory system would have partially completed its movement toward the requested item by the time it received the participant's verbal request, resulting in a 3.4-second advantage. When predictions were incorrect but involved items neighboring the requested item (78.33% of the time), anticipatory control still benefited time to grasp (3-second advantage), as the system would have moved toward the vicinity of the correct item, providing it with a time advantage in moving toward the correct item. We note again that *time to grasp* is an approximation of *task time*.

TABLE I
DESCRIPTIVE STATISTICS OF OBJECTIVE MEASURES FROM THE ANTICIPATORY CONTROL CONDITION BROKEN DOWN INTO CORRECT AND INCORRECT PREDICTIONS AS WELL AS NEIGHBORING-ITEM PREDICTIONS THAT ARE CONSIDERED CORRECT AND INCORRECT.

Control method	Prediction	Response time *(ms)*	Time to grasp *(s)*
Reactive		482.71(SD=551.33)	8.80 (SD=1.26)
Anticipatory	*All*	256.41 (SD=443.31)	6.29 (SD=1.99)
	Correct	51.03 (SD=195.64)	5.40 (SD=1.72)
	Incorrect	516.04 (SD=527.35)	7.41 (SD=1.74)
	Neighboring, Correct	164.70 (SD=300.38)	5.80 (SD=1.85)
	Neighboring, Incorrect	587.97 (SD=673.68)	8.06 (SD=1.40)

We also found that the ability to correctly predict user intent was strongly associated with improvements in the two objective measures that resulted from the use of anticipatory control. Correlation analyses using Pearson's product-moment method showed that prediction accuracy was strongly correlated with response time, $r(118) = -0.52$, $p < .001$, and time to grasp, $r(118) = -0.50$, $p < .001$. This interdependence between prediction accuracy and objective measures highlight the importance of correctly predicting user intent for achieving efficient human-robot collaboration.

Subjective measures — Table II and Figure 4 summarize the results from our subjective measures, particularly the perceived *awareness* and *intentionality* of the robot. Participants rated the anticipatory system to be significantly more aware of their intended choices than the reactive system. However, no significant differences were found in how intentional participants found the two robot systems to be.

Post-experiment interview — In the post-experiment interview, we asked participants open-ended questions about their perceptions of how the two systems behaved in preparing their orders. Several participants described the proactive behavior

TABLE II
STATISTICAL TEST RESULTS FOR THE MANIPULATION CHECK, OBJECTIVE MEASURES, AND SUBJECTIVE MEASURES

Control method	**Objective Measures**		
	Response time *(ms)*	Time to grasp *(s)*	
Reactive	482.71 (SD=551.33)	8.80 (SD=1.26)	
Anticipatory	256.41 (SD=443.31)	6.29 (SD=1.99)	
	F(1,46)=12.96, p<.001 95% CI [49.77, 175.99] d=0.452	F(1,46)=147.88, p<.001 95% CI [1.05, 1.47] d=1.507	
Control method	**Subjective Measures**		
	Manipulation check	Awareness	Intentionality
Reactive	6.79 (SD=0.41)	3.91 (SD=1.56)	4.54 (SD=1.73)
Anticipatory	4.13 (SD=2.31)	5.09 (SD=1.29)	4.66 (SD=1.58)
	F(1,46)=31.01, p<.001 95% CI [-1.82, -0.85] d=1.603	F(1,46)=8.24, p=.006 95% CI [0.18, 1.01] d=0.824	F(1,46)=0.06, p=.812 95% CI [-0.43, 0.54] d=0.072

of the anticipatory robot as being efficient, which was in line with the findings from our response time and time to grasp measures, as illustrated in the excerpts below:

> P3: *"[The anticipatory robot] seemed like it's moving toward what I was going to order, so I thought it knew... I guess that would be more time efficient if it already knew."*

> P4: *"[The anticipatory robot] just moved the arm closer to the fruit before I said something and so it was faster... it was preparatory... it was being more efficient."*

> P5: *"[The anticipatory robot] was going for, I guess, what my eyes were looking towards before I even made a decision."*

Eight participants explicitly mentioned that they preferred the anticipatory system over the reactive one because of the perceived efficiency and proactivity of the robot. On the other hand, two participants preferred the reactive system, one participant describing the robot's anticipatory actions as "freaky" and reporting feeling "unnerved" and "bothered:"

> P1: *"I could tell [the anticipatory robot] was watching my gaze or aware of my gaze... It has awareness... and that almost felt kind of freaky... that it almost could guess what I wanted... I didn't like it as much."*

The other participant who preferred the reactive system over the anticipatory one cited an instance of the anticipatory robot making a wrong prediction and moving toward the opposite direction as the primary basis of this preference:

> P8: *"[The anticipatory robot] shouldn't move before I said what I wanted... so I guess that's [its] fault..."*

V. DISCUSSION

In this paper, we present a novel "anticipatory control" method that enables a robot system to monitor its user's gaze patterns to predict their task intent and perform anticipatory actions based on these predictions in human-robot collaboration scenarios. We implemented this method as a robot system that integrated an arm manipulator, eye tracker, dialogue manager, and a trained intent-prediction component. A human-robot interaction study demonstrated that our method improves the effectiveness of the robot in responding to user requests—resulting in shorter response and task times—and user perceptions of the awareness of the robot of its user. Below, we discuss the design and research implications of the findings from our study and the limitations of the presented work.

A. Anticipatory action for efficient teamwork

Our evaluation demonstrated that the anticipatory system, compared to the reactive system, provided on average a 2.5-second advantage in reaching toward the correct item and completing the task. This advantage resulted from our proposed method for intention prediction and proactive motion planning and execution. While the reactive control method enabled the robot to respond to user requests in less than 500 milliseconds, the anticipatory control method further reduced task times and improved user perceptions, as demonstrated by data from subjective measures as well as open-ended interviews. We expect these improvements to significantly benefit human-robot teams, resulting in more efficient and fluent teamwork, and have a compounding positive effect in repeated interactions, such as assembly work in manufacturing.

B. Intention prediction in practice

Several practical issues arose in realizing intention prediction in a real-time interactive robot system. First, inferring what items participants were looking toward during interactions involved inherent uncertainties. In order to alleviate some of this uncertainty and accurately link gaze fixations to items, we utilized projective transformation between the task space captured by the eye tracker and real-world task space. The findings from the evaluation showed that our implementation incorrectly inferred gaze targets 18.75% of the time, which subsequently affected intent prediction and the anticipatory execution of actions. While we expect future implementations to achieve higher levels of accuracy and better reasoning regarding uncertain observations, prior studies of human-human (e.g., [28]) and human-robot (e.g., [29], [30]) interactions have reported a constant error rate in observers' ability to accurately determine the gaze targets of humans or robots. Future work must explore how such error can be alleviated, for example, by integrating information about the sequence of gaze patterns as well as domain knowledge to help determine priors on what items are likely to be gaze targets.

Further, we modeled the collaboration as a sequence of episodic exchanges (e.g., one for each requested ingredient) and predicted user intent in each exchange independent of prior exchanges or likely future exchanges. While this assumption simplifies the modeling problem and the required solutions, it underutilizes information that could benefit predictions of user intent, as actions taken across different episodes are likely to be highly interdependent and linked to an overarching plan. Our evaluation showed that among 53 incorrect predictions, eight instances did not involve any identifiable gaze targets and another 18 instances involved participants looking toward alternative items. These observations highlight violations of our independence assumption and suggest a more complex process of choosing and communicating items that our model did not capture. To overcome this limitation, future work must build more detailed models of decision making and communication in collaborative interactions.

Although our anticipatory system imperfectly predicted user intent, many of the errors directed the robot toward items that neighbored the correct gaze target (i.e., immediately above, below, to the left, or to the right), and the robot could re-plan when the correct gaze target was determined with minimal delay. Therefore, even many of the erroneous predictions helped the robot more efficiently respond to user requests.

C. Other Limitations

In addition to the discussion provided above and the system limitations described in Section III-F, this work has a number of limitations that motivate future research. First, as in most data-driven machine learning approaches that are attuned to training data, the performance and the generalizability of our SVM-based intent-prediction component are constrained by the training data used and the specific flow and context of the interaction from which the data were collected. Further research is needed to achieve robust prediction algorithms that are generalizable to a wide range of contexts. Second, while participants perceived the anticipatory robot system as being more aware of their actions and intents, we see many possibilities for how the robot can better communicate its awareness to its user, for instance, by displaying "legible" motion [31], that we did not explore in this work. Other potential solutions include the robot changing its movement velocity based on the confidence of its predictions or, when confidence is low, moving toward a location that is more optimal for re-planning rather than moving toward an incorrect target. Finally, future work may draw on other user behaviors, such as facial expressions, gestures, and linguistic cues, to achieve more accurate and robust prediction of user intent.

VI. Conclusion

To achieve fluid, efficient collaboration, robots need to understand and anticipate their human partners' intentions and to act accordingly. In this paper, we proposed an *anticipatory control* method that allows robots to proactively prepare and execute actions toward a shared goal based on anticipation of their human partners' intentions. We developed an autonomous robot system that implemented anticipatory control to engage users in a collaborative task. The system monitored the users' gaze, predicted their task intent, and acted proactively in response to the predicted intent. We demonstrated the effectiveness of the anticipatory control method and the implemented robot system in contributing to efficient teamwork and positive user experience in human-robot collaboration. This work highlights the promise that anticipatory control holds for realizing fluent and efficient human-robot teamwork in day-to-day settings.

Acknowledgments

This work was supported by the National Science Foundation awards 1149970, 1208632, and 1426824. The authors would like to thank Christopher Bodden, Catherine Steffel, and Xiaoyu Wang for their help with this work.

References

[1] G. Pezzulo and D. Ognibene, "Proactive action preparation: Seeing action preparation as a continuous and proactive process," *Motor control*, vol. 16, no. 3, pp. 386–424, 2012.

[2] N. Sebanz, H. Bekkering, and G. Knoblich, "Joint action: bodies and minds moving together," *Trends in Cognitive Sciences*, vol. 10, no. 2, pp. 70–76, 2006.

[3] N. Sebanz and G. Knoblich, "Prediction in joint action: what, when, and where," *Topics in Cognitive Science*, vol. 1, no. 2, pp. 353–367, 2009.

[4] A. N. Meltzoff and R. Brooks, "'like me' as a building block for understanding other minds: Bodily acts, attention, and intention," in *Intentions and intentionality: Foundations of social cognition*, B. F. Malle, L. J. Moses, and D. A. Baldwin, Eds., 2001, pp. 171–191.

[5] R. S. Johansson, G. Westling, A. Bäckström, and J. R. Flanagan, "Eye–hand coordination in object manipulation," *The Journal of Neuroscience*, vol. 21, no. 17, pp. 6917–6932, 2001.

[6] M. Land, N. Mennie, J. Rusted *et al.*, "The roles of vision and eye movements in the control of activities of daily living," *Perception*, vol. 28, no. 11, pp. 1311–1328, 1999.

[7] G. Butterworth, "The ontogeny and phylogeny of joint visual attention." in *Natural theories of mind*, A. Whiten, Ed. Blackwell, 1991.

[8] M. Tomasello, *Why we cooperate*. MIT Press, 2009.

[9] H. Koppula and A. Saxena, "Anticipating human activities using object affordances for reactive robotic response," in *In RSS*, 2013.

[10] J. Mainprice and D. Berenson, "Human-robot collaborative manipulation planning using early prediction of human motion," in *Proceedings of IROS*, 2013, pp. 299–306.

[11] G. Hoffman and C. Breazeal, "Effects of anticipatory action on human-robot teamwork efficiency, fluency, and perception of team," in *Proceeding of HRI*, 2007, pp. 1–8.

[12] K. P. Hawkins, S. Bansal, N. N. Vo, and A. F. Bobick, "Anticipating human actions for collaboration in the presence of task and sensor uncertainty," in *Proceedings of ICRA*, 2014, pp. 2215–2222.

[13] H. Bekkering, E. R. De Bruijn, R. H. Cuijpers, R. Newman-Norlund, H. T. Van Schie, and R. Meulenbroek, "Joint action: Neurocognitive mechanisms supporting human interaction," *Topics in Cognitive Science*, vol. 1, no. 2, pp. 340–352, 2009.

[14] C. D. Frith and U. Frith, "How we predict what other people are going to do," *Brain research*, vol. 1079, no. 1, pp. 36–46, 2006.

[15] V. Gallese and A. Goldman, "Mirror neurons and the simulation theory of mind-reading," *Trends in cognitive sciences*, vol. 2, no. 12, pp. 493–501, 1998.

[16] M. Tomasello, "Joint attention as social cognition," in *Joint attention: Its origins and role in development*, C. Moore and P. Dunham, Eds., 1995, pp. 103–130.

[17] S. E. Brennan, X. Chen, C. A. Dickinson, M. B. Neider, and G. J. Zelinsky, "Coordinating cognition: The costs and benefits of shared gaze during collaborative search," *Cognition*, vol. 106, no. 3, pp. 1465–1477, 2008.

[18] A. Doshi and M. M. Trivedi, "On the roles of eye gaze and head dynamics in predicting driver's intent to change lanes," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 3, pp. 453–462, 2009.

[19] W. Yi and D. Ballard, "Recognizing behavior in hand-eye coordination patterns," *International Journal of Humanoid Robotics*, vol. 6, no. 03, pp. 337–359, 2009.

[20] C.-M. Huang, S. Andrist, A. Sauppé, and B. Mutlu, "Using gaze patterns to predict task intent in collaboration," *Frontiers in psychology*, vol. 6, p. 1049, 2015.

[21] J. Gray, C. Breazeal, M. Berlin, A. Brooks, and J. Lieberman, "Action parsing and goal inference using self as simulator," in *Proceedings of RO-MAN*. IEEE, 2005, pp. 202–209.

[22] C. Pérez-D'Arpino and J. Shah, "Fast target prediction of human reaching motion for cooperative human-robot manipulation tasks using time series classification," in *Proceedings of ICRA*. IEEE, 2015, pp. 6175–6182.

[23] S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, and N. Hagita, "How to approach humans?-strategies for social robots to initiate interaction," in *Proceedings of HRI*. ACM/IEEE, 2009, pp. 109–116.

[24] D. Ognibene and Y. Demiris, "Towards active event recognition," in *Proceedings of IJCAI*. AAAI, 2013, pp. 2495–2501.

[25] 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*, vol. 6, p. 3, 2012.

[26] J. J. Kuffner and S. M. LaValle, "RRT-connect: An efficient approach to single-query path planning," in *Proceedings of ICRA*. IEEE, 2000, pp. 995–1001.

[27] H. Seltman, *Experimental Design for Behavioral and Social Sciences*. Carnegie Mellon University, 2012. [Online]. Available: http://www.stat.cmu.edu/~hseltman/309/Book/

[28] A. D. P. James J. Gibson, "Perception of another person's looking behavior," *The American Journal of Psychology*, vol. 76, no. 3, pp. 386–394, 1963.

[29] M. Imai, T. Kanda, T. Ono, H. Ishiguro, and K. Mase, "Robot mediated round table: Analysis of the effect of robot's gaze," in *Proceedings of RO-MAN*, 2002, pp. 411–416.

[30] B. Mutlu, F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita, "Nonverbal leakage in robots: communication of intentions through seemingly unintentional behavior," in *Proceedings of HRI*, 2009, pp. 69–76.

[31] A. D. Dragan, K. C. Lee, and S. S. Srinivasa, "Legibility and predictability of robot motion," in *Proceedings of HRI*, 2013, pp. 301–308.