Know Thy User: Designing Human-Robot Interaction Paradigms for Multi-robot Manipulation*

Bennie Lewis and Gita Sukthankar

Abstract—This paper tackles the problem of designing an effective user interface for a multi-robot delivery system, composed of robots with wheeled bases and two 3 DOF arms. There are several proven paradigms for increasing the efficacy of human-robot interaction: 1) multimodal interfaces in which the user controls the robots using voice and gesture; 2) configurable interfaces which allow the user to create new commands by demonstrating them; 3) adaptive interfaces which reduce the operator’s workload as necessary through increasing robot autonomy. Here we study the relative benefits of configurable vs. adaptive interfaces for multi-robot manipulation. User expertise was measured along three axes (navigation, manipulation, and coordination), and users who performed above threshold on two out of three dimensions on a calibration task were rated as expert. Our experiments reveal that the relative expertise of the user was the key determinant of the best performing interface paradigm for that user, indicating that good user modeling is essential for designing a human-robot interaction system meant to be used for an extended period of time.

I. INTRODUCTION

Designing an effective human-robot interaction paradigm is particularly important for complex tasks, such as multi-robot manipulation, which require the human and robot to work in tightly coupled fashion [12]. In our delivery task, the human driver commands a pair of robots to move stacks of objects from starting piles to the goal location in a lightly cluttered environment [15]–[17]. Each robot is composed of a non-holonomic wheeled base and two 3-DOF arms for lifting objects. Some of the objects are large enough to require two robots to lift, necessitating tight coordination between the robots (see Figure 1). The human operator’s role is to align the robots, grasp the objects, and oversee the delivery process. The task was created to be simple to learn, but challenging to do within the allotted time. It models many types of real-world robotic usage scenarios, including delivering supplies to urban search and rescue victims, explosive ordinance disposal, robotic butler chores, and multi-robot office delivery systems.

There is a rich body of pre-existing work on reducing operator workload in human-robot interaction [10], [21]. Multimodal interfaces [22] that replace finicky keyboard/mouse commands with speech and gesture can improve the operator experience and reduce the time required to perform individual actions. Adaptive interfaces that intelligently increase the robots’ autonomy can be highly effective on multi-robot tasks [26], particularly when they are coupled with good user modeling mechanisms [17]. Learning from demonstration (LfD) has been successfully used in many robotics applications to create policies based on a set of demonstrations provided by the user [1]. Embedding LfD within a configurable user interface empowers the user to create customized primitives that the system can execute autonomously, enabling the automation of frequently repeated sections of the task. However, it may be difficult for inexperienced users to demonstrate complicated sections of the tasks, where help is desirable, effectively enough to be automated using learning from demonstration.

This paper evaluates the existence of interaction effects between the expertise of the user and their performance with a specific user interface (adaptive vs. configurable) on a multi-robot manipulation task. Users controlled the movement of the robots with a combination of Nintendo Wii commands and arm gestures. Both styles of user interfaces were equally effective at reducing dropped objects and task completion time, but our experiments revealed a strong interaction between the expertise of the user, as measured on a set of calibration tasks, and their performance using a particular class of user interface. Only experts were able to effectively leverage the configurable interface, while novice users performed better with the adaptive interface. This suggests that using a real-time user modeling system to activate or deactivate user interface functions could be an effective way to support long-term users whose expertise is constantly improving. For instance, Francois et al. demonstrated a system that classified changing human-robot tactile interaction styles in real-time [8].

*This research was supported in part by NSF award IIS-0845159.
1B. Lewis and G. Sukthankar are with the Department of EECS, University of Central Florida, 4000 Central Florida Blvd, Orlando, FL 32816-2362 {blewis,gitars}@eecs.ucf.edu

Fig. 1. The multi-robot delivery task requires the operator to control the pair of robots to move objects to a goal location. Objects too large to be lifted by a single robot (such as the box shown in the photo) must be delivered by the team of robots driving in tandem under the direction of the human operator.
II. RELATED WORK

In cases where the user is likely to anthropomorphize the robot, imbuing the robot with social skills such as turn-taking, intention recognition, and emotion models, can yield higher dividends in improved human-robot interaction and increased trust in automation [2]. This can be particularly important in manipulation tasks that involve the human sharing the robot workspace and handing objects directly to the robot (e.g., [7], [19]). However, the multi-robot delivery scenario described in this paper does not involve a shared workspace. Our task is extremely labor intensive for the human operator; in the extreme case, the operator can opt to teleoperate the robots directly through all phases of the task, rather than relying on the robot’s autonomous behaviors. Therefore our system focuses on reducing operator workload.

Improved visualization can improve the operator’s situational awareness and reduce the amount of time required to make decisions [24]. The effort required by the user to communicate with the robot can be minimized through the use of voice, gesture, or EEG controls [11]. Gesture-based interfaces, in particular, may be more intuitive for manipulation and remotely controlling robot arms [4]. Pieska et al. developed a gesture-based system for assisting inexperienced operators at industrial robot programming tasks; they note that their interface was useful for users with varying levels of robot programming expertise [23]. For this paper we developed a bimanual gesture-based interface for modifying the arm position of the robots.

Another option is to shift workload from the operator to the robots by increasing the robots’ autonomy. The robot's autonomy can be adjusted intelligently by encoding the user’s interruption preferences as rewards, allowing the robot to make optimal use of the human’s time [25]. Here, we evaluate an adaptive user interface that modifies the robot’s autonomy based on a user model learned during the calibration period [17]. Muszynski and Behnke [20] presented an interface for personal service robots that allowed the users to select the appropriate level of autonomy based on their perceived workload.

For multi-robot tasks, reasoning about teamwork decisions, such as role allocation and communication timing, can be as important for the success of the task as taskwork decisions. The Multiagent Adjustable Autonomy Framework (MAAF) was designed specifically for facilitating teamwork in large mixed human-robot teams [9]. Coactive design [12] focuses on understanding and managing interdependencies between team members in RAP (Robots, Agents, and People) systems. Unlike RAP systems in these large-scale military scenarios, our multi-robot delivery scenario is very sensitive to small execution differences; even good role allocations often fail due to minor timing failures that lead to dropped objects.

Intrinsically these paradigms require the designer to make decisions on behalf of the users, based on the experiences of a small test group. An alternative approach is to create a configurable interface that allows each user to customize the behavior of the system [6], [18]. One method to achieve this personalization is to allow the user to program the system directly, using learning from demonstration. Learning from demonstration has been used successfully to learn manipulation primitives [5], [14], and even watching human action strategies for dual-arm manipulation tasks has proven to be useful in robot controller design [13]. Alternatively, increased personalization can be achieved by allowing the user to modify the retargetting function used to map the user teleoperation sequences onto the robot [6].

In this paper, we hypothesize that expert-novice differences can affect the efficacy of configurable interfaces. Chen and Barnes’s study on multi-robot supervisory control indicates that individual differences in spatial ability, attentional control and video gaming can affect the quality of human-robot interaction [3]. In our previous work [17], we created a user modeling system specifically for this scenario, in which short teleoperation sequences were used to classify user expertise levels on three aspects of the task: navigation, manipulation, and coordinating the robots. Here, we use the same modeling paradigm to predict which users are likely to prefer an adaptive version of the interface that modifies robot autonomy vs. a configurable version in which the user creates autonomous behavior macros that can be activated on command.

III. ROBOT PLATFORM

For our multi-robot delivery task, we designed a robot, the Home and Urban Intelligent Explorer (HU-IE), for lifting and transporting lightweight objects in an indoor environment. Our robot was constructed using the following commercially available components: an iRobot Create, Acer Aspire One netbook, two NXT 2.0 Robotics kits, a Logitech Communicate STX webcam, Turtlebot shelves, and Tetrax Robotics parts. For this paper, we upgraded our system to include two arms on each robot, in contrast to our previous system which only had one arm per robot.

Each of the two arms on the HU-IE robot was created using the LEGO NXT Robotic Kit; arms are 1.2 feet long and extend 8 inches in front of the robot. They are actuated using three motors and have an operating range of $-45^\circ$ to $90^\circ$ in elevation. At the end of each arm is a four tong claw with rubber grips capable of grasping objects sized for a human hand. Having two arms gives the users more choices on how to parallelize the task, but makes the teleoperation problem slightly more difficult. Even with two arms, some objects are still too large to be carried by a single robot. We added an ultrasonic sensor to the claw of the robot to determine the distance between the claw and the pickup object, along with an accelerometer to measure the arm angle. A small webcam mounted on one of the robot arms presents a first-person perspective to the user during teleoperation. The Acer netbook relays sensor information from the Create sensors and webcam to the user interface.

1A complete design for our system is available at: http://ial.eecs.ucf.edu/HUIE.php.
IV. USER INTERFACE

In this paper, we evaluate an adaptive version of the interface that adjusts the robot’s autonomy based on the user’s expertise [17] vs. a configurable version that allows the user to create macros for autonomous behaviors using a simple learning from demonstration paradigm [16]. The operator interacts with the robots through a combination of arm gestures, tracked using a Microsoft Kinect sensor, plus gamepad controls executed with a Nintendo Wii remote. To facilitate tandem operation, the system supports special mirror modes in which the teleoperation commands are copied by the unmanaged robot [15]. Single mirror mode allows the user to simultaneously control both arms on a single HU-IE robot. Dual mirror mode is used for simultaneously controlling both the robots including the arms; it is particularly useful for instructing the robots to cooperatively lift and deliver objects.

The Wii remote controller is configured as follows. Buttons One and Two are used to toggle between the two robots and to activate the mirror mode in the unmanaged robot. The + and - buttons are used to switch between controlling a single robot arm and activating the mirror mode for the unmanaged arm on the robot currently being controlled by the user. The B button (while pressed) enables arm motion tracking, allowing the user to control the HU-IE robotic arms, and the claw can be toggled open and close using the A button. The mobile base is driven with the D Pad. In the configurable version of the interface, the Home button activates the macro mode, and the D Pad is then used to select the macro to execute.

The user interface runs on a separate desktop computer, connected to two Microsoft Kinect sensors. The first Kinect sensor is oriented toward the user and tracks the user’s arm positions in order to control the elevation of the robotic arms. The Kinect SDK provides access to the user’s skeletal tracking information for twenty joint track states every frame. We then extrapolate the data and map it to the active arm on the HU-IE robot. Since the human’s arm has more degrees of freedom and a greater operating range, we only take the Y axis position data into consideration and ignore anything outside the operating range of the robotic arm. This allows the user to comfortably move their arm up and down without strain while controlling the corresponding arm on the active HU-IE robot. The system monitors the location of the HU-IE robots and delivery objects using a second Kinect sensor, oriented toward the workspace. The user interface utilizes a grid-based localization technique to track the robot positions and an A* path planner to select paths for the robots under autonomous operation. Figure 3 shows the user’s view of the interface.

Fig. 2. A panel on the interface displays the camera’s view of the user. The user gestures up and down to modify the elevation of the robotic arms. Joint tracking information is obtained from a Kinect sensor; only the joints marked in green are used to control the arms.

A. Adaptive

The adaptive interface (described in more detail in [17]) adjusts the robots’ autonomy based on the user’s teleoperation skill level, providing less autonomy to expert users and more autonomy to novices. User expertise is assessed using a set of calibration tasks, designed to measure the user’s ability to drive the robots, manipulate objects, and control both robots simultaneously; semi-supervised k-means is then used to group the users as being expert or novice on the three separate dimensions of teleoperation proficiency. Based on this initial expertise assessment, the adaptive interface selectively adjusts the autonomy of the robots. Users who were rated as novices on the navigation dimension receive more help during sections of the task that involve driving the robots. We implemented two autonomous navigation behaviors which the adaptive interface invokes when the user starts driving: 1) auto goal return and 2) nearest object seeking. Any time that a HU-IE robot is under autonomous operation, the user can retake control of the HU-IE robot by canceling the drive command. For novice users, the system will reactivate the drive command during robot idle times. If the user is classified as being expert at navigation, the system does not reactivate the drive command.

For users who are classified as novices at manipulation, the adaptive interface autonomously adjusts the arm and the claw to help the user when the robots are near objects. Based on arm accelerometer sensor data and Kinect object detection, the adaptive interface attempts to calculate the angle required for a successful pickup and adjusts the arm accordingly when an object is within a certain radius of the robot. The adaptive interface observes the incoming commands, adds the required adjustments to the end of the command string, and displays it to the user before sending it to the robot. If the ultrasonic sensor indicates that the grasp will not be successful, the adaptive interface autonomously adjusts the claw and mobile base of the HU-IE robot to improve the grasp. Note that even though it is possible to autonomously calculate reasonable base, arm, and claw positions for grasping objects, an expert user can still outperform fully autonomous operation. Users

2We have made the user interface software freely available at: http://ial.eecs.ucf.edu/IAI_Package.zip.
who perform poorly on the coordination calibration task receive assistance when the robots are within close proximity of the same pickup object; the adaptive interface attempts to adjust the arm, claw, and base of both robots to improve the alignment. This behavior is also triggered if the arms of both robots are not positioned evenly.

B. Configurable

The configurable interface (described in more detail in [16]) enables the user to elect sections of the task for the robots to execute autonomously. The user might specify sections for multiple reasons: 1) they occur repeatedly 2) are tiresome to complete 3) need to transpire while the user is occupied with other tasks, such as teleoperating the second robot. To make the process of macro acquisition simpler, the user initially performs a demonstration by teleoperating a single robot. However, during the task, the macro is automatically generalized to account for the execution state of the second robot. The macro can be propagated across both robots by invoking the mirror mode, without additional examples (Figure 4).

During the configurable interface phase, the robot’s state space trajectory is recorded, paying distinct attention to the initial and final states of the trajectory. The state includes the extracted features in absolute coordinates: drive start/end position, arm start/end, claw open/closed. Additionally, the status of all of the key on-board sensor systems (cliff, wall, and bumper sensors) are recorded. The configurable interface also tracks the present location of known movable objects in the environment and whether the user is teleoperating the second robot. The state space trajectory is then used to generate an abstract workflow of the task to generalize to new situations. To build the workflow, the state space trajectory is divided into drive, arm, and claw segments. Adjacent drive and arm segments are merged to form one long segment. The position of the robot is retained in both absolute coordinates, along with its relative position to the nearest object or robot. After the user records a macro, the macro is evaluated in multiple locations, with the HU-IE robot arms at altered angles. This user is then given a chance to verify the performance of a recorded macro. If the user is dissatisfied, the system presents a new variation of the macro according to a predefined list of macro generalization rules. If an object is dropped during the acceptance stage, it is assumed that the drop is the principal reason for the macro non-acceptance and the macro is repeated using the same abstraction but with minor modifications to its positioning relative to the object using the ultrasonic sensor. When the user is satisfied, he/she accepts the macro and maps it to one of the Wii remote D Pad buttons.

V. EXPERIMENTAL METHODOLOGY

To evaluate our research hypothesis that the users’ expertise level on the calibration tasks is a good predictor of their performance with the different versions of the interface (adaptive vs. configurable), we recruited 40 users (18 male, 22 female) between the ages of 23 and 36 to use both interfaces; a video of the subjects performing the task is included in the proceedings. Experiments were conducted using the following procedure:

1) Each participant was trained in basic usage of both systems for 10 minutes. The systems were introduced to the user in randomized order.

2) Users then performed the three 10 minute calibration tasks used to assess their expertise on our dimensions.
of teleoperation proficiency (navigation, manipulation, and coordination). The tasks included 1) navigating one robot through a course 2) lifting different types of small objects 3) lifting large objects that require both robots to move. The machine learning classifier was used to categorize the user’s performance on each of the three dimensions; the output of the classifier was then added to the configuration file of the adaptive interface.

3) Participants were then allowed to create macros using the configurable version of the interface. Each participant was allotted 30 minutes to create four macros and map them to the controller. During the macro acquisition phase, the subjects principally interacted with a single robot. After creating each macro, they described the macro on a worksheet.

4) Users were asked to complete Scenario 1 twice, using the two different interfaces. The conditions were randomly ordered, and subjects were allotted 15 minutes per condition In this scenario, the users had to use the two HU-IE robots to search an indoor area and transport small objects (movable by a single robot) to the goal. Objects were arranged in three piles (shown in the left and center panels of Figure 5). Two of the piles contained a mixture of irregularly shaped small objects, and the third pile was composed of boxes that required the robots to use both arms.

5) Users were asked to complete Scenario 2 twice, using the two different interfaces. The conditions were randomly ordered, and subjects were allotted 15 minutes per condition. This scenario stressed the user’s ability to perform bimanual manipulation. The piles contained a combination of large objects, some that could be lifted with two arms on the same robot and some that required two robots to transport to the goal location.

VI. Results

The execution traces from the calibration tasks performed at the beginning of the experiment were used to classify user expertise on three teleoperation dimensions (navigation, manipulation, cooperation) using a semi-supervised k-means classifier (k=2, novice/expert) seeded with data from a pilot study. For the purposes of this experiment, participants who were rated by the classifier as being expert in two dimensions were grouped into the expert category. Users were also asked to assess their own performance after the initial training period. There was a high agreement between the subject’s self-report of their skill on the various tasks and the classifier’s assessment. Note that there is a strong similarity between our user interface controls for driving the robots and common gaming interfaces, and the number of users expert at navigation was higher than for the other two dimensions (manipulation and cooperation).

To evaluate user performance, we examined the task completion time and the number of object drops on the two different scenarios. For task completion time on Scenario 1, a two-way analysis of variance yielded a main effect for the user’s expertise, F(1,64) = 37.2, p < .01, such that experts completed the scenario significantly faster than novices. The main effect of user interface type was non-significant, F(1,64) = 0.54. However, the interaction effect was significant, F(1,64) = 9.73, p < .01. Experts had a reduced task completion time using the configurable interface, whereas novices performed better with the adaptive one. A two-way analysis of variance yielded similar results on Scenario 2: a main effect for the user’s expertise, F(1,64) = 23.9, p < .01, such that experts completed the scenario significantly faster than novices. The main effect of user interface type was non-significant, and the interaction effect was significant, F(1,64) = 12.1, p < .01. Experts had a reduced task completion time using the configurable interface, whereas novices performed better with the adaptive one.

We also manually counted the number of times the robots dropped objects in Scenario 1. In a two-way analysis of variance, neither main effect (expertise or interface type) resulted in significant differences, but the interaction effect was significant F(1,64)=13.76, p < .01. Novices dropped fewer objects with the adaptive version of the interface, and experts dropped fewer objects with the configurable version. Scenario 2 followed the same pattern, with neither main effect being significant but with a significant interaction effect F(1,64)=20.3, p < .01.

To better understand the users’ experiences with the configurable interface, we measured the percentage of time that the robots operated autonomously (i.e. executed pre-programmed macro behaviors) out of the total scenario execution time. Figure 6 shows these results for both scenarios. The percentage of time that the robots were performing macro behaviors was comparable for novices and experts (no significant differences in the post-hoc comparison). In previous work [16], we had noted that differences may exist in the number of macros deployed and the length of macros created for more experienced vs. less experienced users, but apparently the amount of autonomy is comparable.
Fig. 5. Users operated the two robots within a 6.3’ × 6.4’ household area observed by one Kinect sensor and an overhead camera. The experimental scenarios required the subjects to use the robots to transport objects from the three piles, to the goal basket in the right corner.

VII. CONCLUSION AND FUTURE WORK

This paper presents a user modeling technique for reliably predicting which group of users will perform well with a specific interface, adaptive vs. configurable. From a developer point of view it is essential to reduce “feature creep”—the addition of new unneeded functions that add complexity to the development process, without directly benefiting the user. Therefore, our aim is to create the simplest system that can still do a good job serving the user’s needs for the specific task, while still accommodating the needs of a diverse user population. Not only can software and hardware failures sabotage performance but mismatches between the interaction paradigm and the task or between the users and the system are equally problematic. In a real-world usage scenario, the user’s expertise, and hence their interaction style, is likely to change over time. We believe that activating or deactivating features in the interface is one way to accommodate the user’s changing expertise.

REFERENCES