Assistive Teleoperation of Robot Arms via Automatic Time-Optimal Mode Switching

Laura V. Herlant, Rachel M. Holladay, and Siddhartha S. Srinivasa
The Robotics Institute, Carnegie Mellon University
herlant@cmu.edu, rmh@andrew.cmu.edu, siddh@cs.cmu.edu

Abstract—Assistive robotic arms are increasingly enabling users with upper extremity disabilities to perform activities of daily living on their own. However, the increased capability and dexterity of the arms also makes them harder to control with simple, low-dimensional interfaces like joysticks and sip-and-puff interfaces. A common technique to control a high-dimensional system like an arm with a low-dimensional input like a joystick is through switching between multiple control modes. However, our interviews with daily users of the Kinova JACO arm identified mode switching as a key problem, both in terms of time and cognitive load. We further confirmed objectively that mode switching consumes about 17.4% of execution time even for able-bodied users controlling the JACO. Our key insight is that using even a simple model of mode switching, like time optimality, and a simple intervention, like automatically switching modes, significantly improves user satisfaction.

I. INTRODUCTION

People with upper extremity disabilities are gaining increased independence through the use of assisted devices such as robotic arms [1]–[4]. However, the increased capability and dexterity of these arms also makes them harder to control. For example, the Kinova assistive arm in Fig. 1 has 6 independently controllable joints and 1 controllable gripper. Even under Cartesian control of the end-effector, assistive arms have 3 position and 3 orientation freedoms to independently control.

Paradigms for controlling such high dimensional systems with much lower dimensional interfaces like a joystick [5], sip-and-puff [6], or a brain-computer interface [7] fall into two categories.

The first treats the user as an indicator of goals by reducing the user’s input to one of a finite set of predefined configurations or goals [8]. While these systems are easy to operate, removing autonomy from the user often decreases their satisfaction with the system [9], [10].

The second treats the user as a provider of motion by mapping the lower dimensional input to some subset of the arm’s degrees of freedom, called a mode. Users then switch modes to control a different subset [11]–[17], a method known as modal control. Thus the user is able to move the arm everywhere, but not using all of the degrees of freedoms at the same time. Studies with people with disabilities revealed that the numerous and frequent mode switches required for performing everyday tasks made such systems difficult to operate [18], [19].

Our own interviews with current users of the Kinova arm pinpointed that the struggles with modal control relate back to the need to constantly change modes. Our users found switching between the various control modes, seen in Fig. 1, to be slow and burdensome, noting that there were “a lot of modes, actions, combination of buttons”. Each of these mode changes requires the user to divert their attention away from accomplishing their task to consider the necessary mode change [19], [20]. The cognitive action of shifting attention from one task to another is referred to as task switching. Task switching slows down users and can lead to increased errors regardless of the interface they are using to make this switch [21]–[25]. Simply the need to change modes is a harmful distraction that impedes efficient control.

In this paper, we alleviate the problem of mode switching by automatically switching modes. We achieve this by building a model of the the user’s desired robot motion and mathematically computing the optimal switching points according to that model. Doing so creates a level of shared control that allows the user to
freely express their intent in the full configuration space of the robot while simultaneously receiving assistance from the arm [26]–[28]. Our key insight is that even a simple model, like time-optimality, combined with a simple intervention, like automatically switching modes, significantly improves user satisfaction while maintaining performance.

We therefore make the following contributions:

1. **Identifying Mode Change Disruption.** Based on interviews with current JACO users and Kinova employees we identify that mode switching is a common problem in assistive arms. A study with able-bodied users performing a variety of instrumental activities of daily living [29] confirmed that a significant portion of the execution time was spent changing modes.

2. **Time-Optimal Model.** A study using a 2D simulated mobile robot allowed us to determine that we can model the mode changing behavior of the users as time-optimal. Using time as the cost to optimize, we can use Dijkstra’s algorithm to predict when the robot should automatically change modes for the user.

3. **User Performance.** Having developed a policy for when to change modes, we tested how able-bodied users performed with the robot changed modes automatically. Users reported feeling comfortable with the assistance, and that the modes were changed at the same times and places that they would have wanted to change modes themselves.

We discuss several limitations of our work in Sec. V. Key among them is how to generalize the time-optimal model to higher dimensional spaces, like the 6-DOF JACO arms. Algorithms like Dijkstra’s, that compute the optimal cost-to-go, incur combinatorial computational cost with respect to the size of the search space. Our preliminary results on approximate algorithms like Weighted A* and LPA* [30] which trade-off optimality for speed, have been encouraging.

II. **CHALLENGES OF MODAL CONTROL**

To understand the challenges of modal control, we interviewed the developers and current users of the Kinova assistive arm, and conducted a pilot study with able-bodied users operating Kinova’s MICO arm to perform 3 basic tasks (Fig. 2). This section details our findings.

A. **The Kinova MICO and JACO arms**

The JACO is a 6-DOF wheelchair-mounted or workstation-mounted arm with an actuated gripper which presently has over 150 users. The arm can be teleoperated using the same interface that controls the powered wheelchair, typically a joystick. The joystick provided with the arm has 3 DOFs: tilting the stick forward or backward, tilting the stick left or right, and twisting the stick clockwise or counterclockwise. To control the robot arm’s hand location and orientation, 6 DOFs are needed, and a 7th is required to open and close the gripper. Controlling 7 DOFs with a 3-DOF joystick requires at least 3 control modes to divide the robot DOFs into groups of 3, such as: translation mode, wrist mode, and finger mode (Fig. 1). The control mode is switched via buttons on the joystick, and a 4 LED light pattern indicates the current mode.

B. **User Interviews**

We interviewed three current users of the JACO assistive arm who operate the device on a daily basis within their homes. All users expressed difficulty in changing modes. Specifically they noted two key problems. First, users found that “too many mobilizations and too many movements” are needed to perform elementary tasks such as eating. As a result, they often abandoned these tasks completely and relied on their caregivers. Second, users said that remembering which mode the arm was in was difficult and required a lot of concentration. In our interviews, Kinova employees acknowledged the second problem, stating that not remembering the mode can be “a real nightmare” for the user. Kinova is currently working on an LCD screen interface to display the modes more easily. Although this alleviates the second problem, it does not address the first.

C. **Study 1: CAHAI Tasks with the MICO Robot**

To objectively measure the impact of mode switching, we ran a study with able-bodied users performing household tasks with the Kinova MICO arm using a joystick interface.

Experimental Setup Users sat behind a table on which the MICO arm was rigidly mounted (Fig. 1). They used the standard Kinova joystick to control the arm.

Tasks The tasks we chose are modified from the Chedoke Arm and Hand Activity Inventory (CAHAI), a validated, upper-limb measure to assess functional recovery of the arm and hand after a stroke [31]. The CAHAI has the advantage of drawing tasks from instrumental activities of daily living, which are representative of our desired use case, rather than occupational therapy assessment tools such as the 9-Hole Peg Test [32] or the Minnesota Manipulation Test [33] that also evaluate upper extremity function, but do not place the results into context.

Manipulated Factors We manipulated which task the user performed. The three tasks we used were: calling 911, pouring a glass of water from a pitcher, and

<table>
<thead>
<tr>
<th></th>
<th>Translation</th>
<th>Wrist</th>
<th>Finger</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translation</td>
<td>-</td>
<td>1.98 ± 0.15s</td>
<td>1.94 ± 0.16s</td>
</tr>
<tr>
<td>Wrist</td>
<td>2.04 ± 0.51s</td>
<td>-</td>
<td>3.20 ± 1.85s</td>
</tr>
<tr>
<td>Finger</td>
<td>1.30 ± 0.13s</td>
<td>0.98 ± 0.24s</td>
<td>-</td>
</tr>
</tbody>
</table>

**TABLE I: Mean mode switching times**
unscrewing the lid from a jar of coffee. These three tasks were chosen from the CAHAI test because they could easily be modified from a bimanual task to being performed with one arm. The three tasks are shown in Fig. 2.

Procedure After a five minute training period, each user was given a maximum of ten minutes per task. The order of tasks was counterbalanced across the users. The joystick inputs and the robot’s position and orientation were recorded throughout all trials. After all the tasks were attempted, we asked the users to rate the difficulty of each task on a 7-point Likert scale and what aspects make performing tasks difficult with this robot.

Participants and Allocation We recruited 6 able-bodied participants from the local community (4 male, 2 female, aged 21-34). This was a within subject design, and each participant performed all three tasks with a counterbalanced ordering.

Analysis On average, 17.4 ± 0.8% of task execution time is spent changing control modes and not actually moving the robot. The mode changing times were calculated as the total time the user did not move the joystick before and after changing control mode. The fraction of total execution time that was spent changing modes was fairly consistent both across users and tasks as seen in Fig. 3. If time spent changing mode could be removed, users would gain over a sixth of the total operating time.

The tasks the users performed were not of equal difficulty. Users responded that the pitcher pouring was the most difficult task (M=5.5, SD=0.7), followed by unscrewing the jar (M=5.2, SD=0.7), and the easiest task was dialing 911 (M=4, SD=0.6). The total execution time shown in Fig. 3a mirrors the difficulty ratings, with harder tasks taking longer to complete. Difficulty could also be linked to the number of mode switches, mode switching time, or ratio of time spent mode switching, as shown in Fig. 5. The hardest and easiest tasks are most easily identified when using switching time as a discriminating factor. The pitcher and jar tasks both rated as significantly more difficult than the telephone task, which may be due to the large number of mode changes and small adjustments needed to move the robot's hand along an arc — as one user pointed out: “Circular motion is hard.”

One might argue that we are basing our findings on novice users, and their discomfort and hesitation switching modes will diminish over time. However, over the course of half an hour using the arm, and an average
of more than 100 mode switches, users did not show any significant decrease in the time it takes to change mode (Fig. 4). The continued cost of mode switching is further supported by our interviews, in which a person using the JACO for more than three years stated “it’s really hard with the JACO because there are too many mobilizations and too many movements required.”

The users had three possible modes and used two buttons on the top of the joystick to change between them. The left button started translation mode, the right button started orientation mode, and pressing both buttons simultaneously started gripper mode. Changing into gripper mode was particularly since the timing between the two buttons had to be very precise lest the user accidentally press the left or right button when releasing and switch to translation or wrist mode. The cost to change from one mode to another was not constant across the modes; Table I shows the average time it took to change from the mode in the row to the mode in the column. While in this case the difference can be explained by the chosen interface, it could be important to consider if switching from one particular control mode to another causes a larger mental shift in context. Such differences would require the cost of mode switches to be directional, which we leave for future work.

III. Mode Switching Observation

The users of the JACO arm identified that frequently changing modes was difficult. We objectively confirmed the difficulty of mode changing by having able-bodied users perform everyday tasks with the MICO arm. Having identified mode switching as a problem in this complex scenario, we tried to model the problem in a much simpler scenario and provide the foundations for scaling the solution back up to the full space of the MICO arm.

A. Study 2: 2D Mode Switching Task

Study 1 demonstrated that people using modal control spend a significant amount of their time changing modes and not moving the robot. The next step is to model when people change modes so that the robot can provide assistance at the right time. We identified certain behaviors from Study 1 that could confound our ability to fit an accurate model. We observed that different people used very different strategies for each of the tasks, which we postulated is because they were performing multi-step tasks that required several intermediate placements of the robot’s gripper. In some trials, users changed their mind about where they wanted to grab an item in the middle of a motion, which we could detect by the verbal comments they made. To gather a more controlled set of trajectories under modal control, we ran a second study in which we more rigidly defined the goal and used only two modes. To fully constrain the goal, we used a simulated robot navigating in two dimensions and a point goal location. We kept all the aspects of modal control as similar to that of the robot arm as possible. Using a 2D simulated robot made it simpler to train novice users and removed confounds, allowing us to more clearly see the impacts of our manipulated
factors as described below.

**Experimental Setup** In this study, the users were given the task of navigating to a goal location in a planar world with polygonal obstacles. We had each user teleoperate a simulated point robot in a 2D world. There were two control modes: one to move the robot vertically, and one to move it horizontally. In each mode, the users pressed the up and down arrow keys on the computer keyboard to move the robot along the axis being controlled. By using the same input keys in both modes, the user is forced to re-map the key’s functionality by pressing the spacebar. This is a more realistic analogy to the robot arm scenario, where the same joystick is being used in all of the different control modes to control different things.

**Manipulated Factors** We manipulated two factors: the delay when changing modes (with 3 levels) and the obstacles in the robot’s world (with 3 levels). To simulate the cost of changing modes, we introduced either no delay, a one second delay, or a two second delay whenever the user changed modes. Different time delays are analogous to taking more or less time to change mode due to the interface, the cognitive effort necessary, or the physical effort. We also varied the world the robot had to navigate in order to gather a variety of different examples. The three tasks are as follows: (1) an empty world, (2) a world with concave and convex polygons obstacles, and (3) a world with a diagonal tunnel through an obstacle, and are shown in the top row of Fig. 6.

**Procedure** This was a within users study design. Each user saw only one task, but they saw all three delay conditions. Each user had a two trial training period with no delay to learn the keypad controls, and then performed each of the three delay conditions twice. Five users performed each task. The goal remained constant across all the conditions, but the starting position was randomly chosen within the bottom left quadrant of the world. We collected the timing of each key press, the robot’s trajectory, and the control mode throughout each of the trials.

**Measures** To measure task efficiency, we used three metrics: the total execution time, the number of mode switches, and the total amount of time switching modes. We also recorded the path the user moved the robot and which control mode the robot was in at each time step.

**Participants** We recruited 15 participants through Amazon Mechanical Turk (aged 18-60).

**Analysis** When the cost of changing modes increases, people choose different strategies in particular situations. This is best seen in Task 2, where there were two different routes to the goal, whereas in Task 1 and Task 3 the map is symmetrical. When there was no mode delay, nearly all users in Task 2 navigated through the tunnel to get to the goal, Fig. 7. When the delay was one second, some users began to navigate around the obstacles completely, and not through the tunnel. While navigating the tunnel was a shorter Euclidean distance, it required more mode changes than navigating around the tunnel entirely. Therefore we saw that when the cost of mode changes increased, people optimized for fewer mode changes.

We noticed that the user trajectories could be very well modeled by making the assumption that the next action they took was the one that would take them to the goal in the least amount of time. Since switching modes is one of the possible actions, it becomes possible to use this simple model to predict mode switches. The next section discusses the time-optimality model in more detail.

IV. **Time-Optimal Mode Switching**

The time-optimal policy was found by assigning a cost to changing mode and a cost to pressing a key. These costs were found by empirically averaging across the time it took the users from Study 2 to perform these actions. Using a graph search algorithm, in our case Dijkstra’s algorithm [34], we can then determine how much time the optimal path would take. By looking at each (x,y) location, we can see if the optimal path is faster if the robot is in x-mode or y-mode. The time-optimal mode for each particular (x,y) location is the mode which has a faster optimal path to the goal. A visualization of the optimal mode can be seen in Fig. 6 for each of the tasks. Time-optimal paths change into the optimal mode as soon as the robot enters one of the x-regions or one of the y-regions. By plotting the user
trajectories over a map of the regions, we can see where users were suboptimal. If they were moving vertically in the x-region or horizontally in the y-region, they were performing sub-optimally with respect to time.

In Task 1, users were in the time-optimal mode 93.11% of the time. In Task 2, users were in the time-optimal mode 73.47% of the time. In Task 3, users were in the time-optimal mode 90.52% of the time. Task 2 and Task 3 require more frequent mode switching due to the presence of obstacles.

A. Study 3: 2D Automatic Mode Switching

Once we determined that people often switch to be time-optimal, we tested how people would react if the robot autonomously switched modes for them. Using the same tasks from Study 2, we used the time-optimal region maps (Fig. 6), to govern the robot’s behavior.

Manipulated Factors We manipulated two factors: the strategy of the robot’s mode switching (with 3 levels) and the delay from the mode switch (with 2 levels). The mode switching strategy was either manual, automatic or forced. In the manual case, changing the robot’s mode was only controlled by the user. In the automatic case, when the robot entered a new region based on our optimality map, the robot would automatically switch into the time-optimal mode. This change would occur only when the robot first entered the zone, but then the user was free to change the mode back at any time. Within each region in the forced case, by contrast, after every action the user took, the robot would switch into the time-optimal mode. This meant that if the user wanted to change to a suboptimal mode, they could only move the robot one step before the mode was automatically changed into the optimal mode. Hence the robot effectively forces the user to be in the time-optimal mode.

Similar to Study 2, we had a delay condition, however we considered the following three cases: (1) no delay across all assistance types, (2) a two second delay across all assistance types, and (3) a two second delay for manual switching but no delay for auto and forced switching. The purpose of varying the delay was to see if the users’ preference was impacted equally by removing the imposed cost of changing mode (delay type 3), and by only removing the burden on the user to decide about changing mode (delay type 1 and 2).

Hypotheses

H1: People will prefer when the robot provides assistance.
H2: Forced assistance will frustrate users because they will not be able to change the mode for more than a single move if they do not accept the robot’s mode switch.
H3: People will perform the task faster when the robot provides assistance.

Procedure After giving each user two practice trials, we conducted pairs of trials in which the user completed the task with the manual mode and either the forced or automatic mode. Testing the automatic assistance and forced assistance across the three delay conditions led to six pairs. For each pair, users were asked to compare the two trials on a forced choice 7-point Likert scale with respect to the user’s preference, perceived task difficulty, perceived speed, and comfort level. At the conclusion of the study, users answered how the felt about the robot’s mode switching behavior overall and to rate on a 7-point Likert scale if the robot changed modes at the correct times and locations.

Participants We recruited 13 able-bodied participants from the local community (7 male, 6 female, aged 21-58).

Analysis People responded that they preferred using the two types of assistance significantly more than the manual control, \( t(154) = 2.96, p = .004 \), supporting H1. The users’ preference correlated strongly with which control type they perceived to be faster and easier (\( R = 0.89 \) and \( R = 0.81 \) respectively).

At the conclusion of the study, users responded that they felt comfortable with the robot’s mode switching (\( M=5.9, SD=1.0 \)) and thought it did so at the correct time and location (\( M=5.7, SD=1.8 \)). Both responses were above the neutral response of 4, with \( t(24)=4.72, p < .001 \) and \( t(24)=2.34, p = .028 \) respectively. This supports our finding that mode switching can be predicted by following a strategy that always places the robot in the time-optimal mode.

Since this was an experiment, we did not tell participants which trials the robot would autonomously changing modes in. As a result, the first time the robot switched modes automatically, many users were noticeably taken aback. Some users immediately adjusted, with one saying “even though it caught me off guard that the mode automatically switched, it switched at the exact time I would have switched it myself, which was helpful”. While others were initially hesitant, all but two of the participants quickly began to strongly prefer when the robot autonomously changed for them, remarking that it saved time and key presses. They appreciated that the robot made the task easier and even that “the program recognized my intention”.

Over time people learned where and when the robot would help them and seemed to adjust their path to maximize robot assistance. People rarely, if ever, fought against the mode change that the robot performed. They trusted the robot’s behavior enough to take the robots suggestions [35]–[37]. We found no significant difference between the forced and automatic mode switching in terms of user preference \( t(76) = 0.37, p = 0.71 \). Some users even stated that there was no difference at all between the two trials. Therefore we did not find evidence to support H2.

Task efficiency, measured by total execution time and
the resulting trajectories do not differ greatly in terms of model that humans already closely follow. It follows that techniques are choosing when to switch modes based on a however, this is not surprising as the assistance techniques are choosing when to switch modes based on a. The optimal mode regions will become optimal conditions. Therefore we were not able to support manual control, auto switching, and forced switching the robot), was not significantly different between the total time spent changing modes (as opposed to moving the robot), was not significantly different between the manual control, auto switching, and forced switching. Therefore we were not able to support h3. However, this is not surprising as the assistance techniques are choosing when to switch modes based on a model that humans already closely follow. It follows that the resulting trajectories do not differ greatly in terms of path length nor execution time.

V. Discussion

While this work proposes a useful intervention for controlling assistive robotic arms, there remain many avenues to explore.

Generalization While our studies involved exclusively able-bodied subjects, we want to see how these results, particularly those relating to acceptance, generalize to people with disabilities. Study 3, as described in sec. IV-Å, was restricted to a 2D point robot and we are looking to reconduct this experiment on the MICo robot arm. The optimal mode regions will become optimal mode volumes, and the assisted mode switching will occur when the robot enters a new volume.

Priority List of Modes How best to improve modal control is an open question. Here we presented one technique: having the robot perform mode switches automatically. An alternative form of assistance would be to have a priority ordered list of modes. With a single button press the user could transition to the most likely next mode, as estimated by the robot. In the case of an incorrect ordering the user would cycle through the list. This kind of assistance would complement Kinova’s LCD screen as described in sec. II-B.

Extramodal Control We have suggested a method for more easily switching modes. An alternative approach is to remove this problem entirely by having the user only control one mode. The remaining modes would be controlled by the robot, in a type of assistance known as extramodal assistance. While this eliminates the burden on the user in general, mistakes made by the robot become much more costly. In the case of a robot mistake, the user would have to change modes manually, correct the mistake and then revert back to their original task.

Goal uncertainty Often the exact goal the user wants to achieve is not known in advance. In these situations, the system must simultaneously predict the user’s goals and assist them to complete the task [26], [27]. Robot policies for each goal are key to most formalisms that address goal uncertainty. Our framework provides that policy and can be naturally plugged into a policy blending formalism for goal uncertainty.

By investigating modal control and helpful interventions we strive to close the gap between what users want to do with their assistive arms and what they can achieve with ease.

References


