2023 IEEE International Conference on Robotics and Automation (ICRA 2023)
May 29 - June 2, 2023. London, UK

Design and Evaluation of an Augmented Reality Head-Mounted Display User Interface for Controlling Legged Manipulators

Rodrigo Chacón Quesada and Yiannis Demiris[1]

Abstract— **Designing an intuitive User Interface (UI) for controlling assistive robots remains challenging. Most existing UIs leverage traditional control interfaces such as joysticks, hand-held controllers, and 2D UIs. Thus, users have limited availability to use their hands for other tasks. Furthermore, although there is extensive research regarding legged manipulators, comparatively little is on their UIs. Towards extending the state-of-art in this domain, we provide a user study comparing an Augmented Reality (AR) Head-Mounted Display (HMD) UI we developed for controlling a legged manipulator against off-the-shelf control methods for such robots. We made this comparison baseline across multiple factors relevant to a successful interaction. The results from our user study ($N = 17$) show that although the AR UI increases immersion, off-the-shelf control methods outperformed the AR UI in terms of time performance and cognitive workload. Nonetheless, a follow-up pilot study incorporating the lessons learned shows that AR UIs can outpace hand-held-based control methods and reduce the cognitive requirements when designers include hands-free interactions and cognitive offloading principles into the UI.**

I. INTRODUCTION

Recent advances in the design of AR HMD UI offer exciting possibilities for collaboration between people and assistive robots. This article builds on our previous work to explore to what extent, if any, AR HMD UIs improve Human-Robot Interaction (HRI) factors that are relevant for a successful collaboration between legged manipulators and their operators. Although inspections tasks, typical for legged manipulators, are repetitive and expected to be executed autonomously by the robot, a human operator always has to define the robot's movements and its location-based inspection actions (visual inspection, thermal sensing, laser scanning, gauge reading, leak detection, etc.) in the environment first. However, such definition is currently done using traditional UIs, such as joysticks and 2D screen-based UIs [1], [2].

Such interfaces not only limit the operators' availability to use their hands for other tasks; they also take the operator's attention away from the task by diverting attention to the UI, bringing safety concerns to an application motivated by the need to reduce the risk associated with dangerous jobs such as working on offshore platforms and nuclear sites. Furthermore, traditional interfaces could prove unintuitive, discouraging the interaction with the robot partner. By exploiting the users' existing abilities, AR HMD UIs offer the possibility of hands-free interactions through their multi-modal input capabilities, such as eye-gaze and voice commands. Thus, potentially making these interactions more intuitive and easier to perform [3], [4], [5], and increasing the probability of acceptance by non-expert users.

Fig. 1: This paper discusses a multi-dimensional evaluation of our proposed AR HMD UI using validated metrics relevant to a successful human-robot interaction. We compared our AR HMD UI against off-the-shelf hand-held UIs provided by the manufacturers. Moreover, our user study involved navigation and manipulations tasks with varying difficulty levels using a Boston Dynamics' Spot®, a 7 DoF Kinova® robot arm, and a Robotiq® 2F-85 gripper that we integrated into a legged manipulator. See our complementary video.

Furthermore, by leveraging recent advances in vision-based colocalisation capabilities [6], [7], these UIs also remove the need to instrument the environment with fiducial markers. Such markers are commonly used to align the map built by the robot during the definition of the inspection task with the world around it [8]. Colocalisation is the process that allows localising the robots and the HMD under a global coordinate system. We discussed our vision-based colocalisation approach in subsection III-B.

Our contributions are derived from evaluating our proposed AR HMD UI for controlling legged manipulators. We present a user study with 17 participants analysing the effects of our approach on validated metrics for a successful interaction. The participants executed navigation and manipulations tasks with varying difficulty levels using a Boston Dynamics (BD) Spot®, a 7 DoF Kinova® robot arm, and a Robotiq® 2F-85 gripper that we integrated into a legged manipulator (see Figure 1). Our results show that although the AR UI increases immersion, off-the-shelf control methods outperformed the AR UI in time performance and cognitive workload. Nonetheless, a follow-up pilot study incorporating the lessons learned shows that AR UIs can outpace hand-held-based control methods and reduce cognitive requirements when designers include hands-free interactions and cognitive offloading principles into the UI.

---

[1] The authors are with the Personal Robotics Laboratory, Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, United Kingdom {`r.chacon-quesada17`, `y.demiris`}`@imperial.ac.uk`. Rodrigo Chacón is supported by the University of Costa Rica. This research was supported in part by UKRI grant EP/V026682/1, and by a Royal Academy of Engineering Chair in Emerging Technologies to Yiannis Demiris.

979-8-3503-2365-8/23/$31.00 ©2023 IEEE

II. RELATED WORK

A. *User studies with AR UIs for robot control*

Although not in the field of legged manipulators, several researchers have previously considered the problems involved in AR-based robot control [9], [10]. Applications in this domain can be classified into two categories: remote interactions (teleoperation) [11], where the user sends control commands from outside the robot's location, and proximal interactions, where the user and the robot share the same location [12], [13], [14]. However, most previous user studies in this domain are limited to teleoperation and programming applications [15], [16], [17], [18], [13]. Examples of user studies outside these applications demonstrating the benefits in the interaction derived from using AR UIs to control single and multirobot platforms are presented in [3], [4], [19], [20].

In [3], the authors evaluated an AR HMD UI for controlling a smart wheelchair. Their results show that the task load reported by participants is lower when controlling the smart wheelchair with their UI compared to using the joystick. In [4], the authors presented a user study evaluating an AR HMD UI for controlling a multi-robot platform. Their results showed a significant improvement in fluency perception derived from their UI. In [19], the authors evaluated an AR application that disambiguates item references. They compared against physical actions for reducing robot uncertainty and found their approach to be more accurate, faster, and improve usability, trust, and workload. Finally, [20] developed a user study simulating a collaborative manufacturing task using a robotic arm. Their results show that their UI feels more novel to users and reduces physical demand and task completion time while increasing robot utilisation.

B. *Human-Robot Interaction Metrics*

Following standard practice in HRI and interaction design, we evaluated our platform using established metrics:

1) Cognitive Workload: The cognitive workload is the level of measurable mental effort put forth by an individual in response to one or more cognitive tasks [21]. It can be evaluated using the NASA Task Load Index (TLX) to assess the relative importance of six factors that determine how much workload user experiences while performing a task. These six factors are: 1) Mental Demand (MD), 2) Physical Demand (PD), 3) Temporal Demand (TD), 4) Performance (P), 5) Effort (E), and 6) Frustration level (F). These are rated using six scales. For each of these scales, the user has to evaluate the task performed by marking the scale's location that matches their experience.

2) Technology Acceptance: The Technology Acceptance Model (TAM) is used to assess the acceptance of people towards technology products [22], [23], [24]. It includes a total of 10 items with a 5-point Likert scale from which we were interested in the following: a) Perceived usefulness, which is the people's tendency to use or not an application to the extent they believe it will help them perform their tasks better. b) Perceived ease of use; even if potential users found an application useful, they may also believe it is hard to use. Thus, the effort required outweighs the performance benefits of usage. This effort refers to perceived ease of use. c) Intention of use, which refers to the likelihood of using new technology in the future.

Fig. 2: We used the off-the-shelf UIs developed by BD (top) and Kinova (bottom) to control Spot and the arm from the hand-held controller provided by BD with Spot. See Spot controller configurations and Kinova Kortex Web App for further details.

3) Subjective Fluency Metrics for Human-Robot Interaction: When humans collaborate on a shared activity, they can reach a high level of coordination, resulting in a well-synchronised meshing of their actions. Their timing is precise and efficient. They alter their plans and actions appropriately. This quality of interaction is denoted as the fluency of the shared activity [25], [26], [27].

The subjective assessment of fluency includes a total of 30 items with a five-point Likert scale based on the following areas: a) Human-robot fluency, b) Robot relative contribution, c) Trust in a robot, d) Positive teammate traits, e) Improvement, f) Working alliance for Human-Robot teams, and g) Individual measures. These areas include a three-item scale evaluating fluency directly (a) and six possible downstream outcomes of collaborative fluency (b–g). Scale f) is an adaptation of an existing instrument, the "Working Alliance Inventory", adapted to human-robot teamwork.

4) System Usability: Usability focuses on how well users can learn and use a device to achieve their goals. It also refers to how satisfied users are during that process [28]. Usability is evaluated using the Hybrid System Usability Scale (H-SUS) [29], a ten-item scale that combines pictorial and verbal information on the same scale. The pictorial information consists of two visual representations depicting the extreme points of a bipolar scale. An avatar is presented, interacting with a mobile device in a specific usage situation (negative vs positive experience). A five-point Likert scale is provided to give the ratings. The verbal content is placed above the pictorial scale, containing a statement for the subjective assessments of different usability items.

5) Immersion: Immersion is a form of cognitive and emotional absorption that promotes enjoyment and engagement in a task or while learning [30]. For Location-aware AR applications, immersion is measured using ARI, a 21-item seven-point Likert-type instrument based on a multi-levelled immersion model with multidimensionality at each level [30]. These levels are: a) Engagement; this first level is divided into the constructs of *interest*, which measures the user's interest in the activity, and *usability*, which measures the user's perception of the usability of the application. b) Engrossment; this second level is divided into the constructs of *Emotional Attachment*, which measures the emotional attachment to the activity, and *Focus of Attention*, which measures the user's focus during the activity. c) Total Immersion; this third level is divided into the constructs of *Presence*, which measures the user's sense of feeling surrounded by a blended, yet realistic physical/virtual environment, and *Flow*, which measures the user's absorption in the activity.

III. SYSTEM DESCRIPTION

A. *Robot platform*

We used SpoK as the robot platform for our user study [7]. SpoK is shown in Figure 1. SpoK integrates BD's Spot, a 7 DoF Kinova robot arm, and a Robotiq 2F-85 gripper into a legged manipulator (see [31] for a similar Spot-Kinova setup). We provided the on-board computing processing of the robot platform using Spot Core, a computer equipped with an Intel® i5 CPU, 16 GB of RAM, 512 GB SSD, Ubuntu Desktop 18.04. We developed our code using ROS Noetic and Docker. For the UIs, we used the Microsoft HoloLens 2 as HMD and Unity 2019.4.10f1, Mixed Reality Toolkit v2.5.3 and ROS Sharp to build the AR HMD UI. In addition, we used the off-the-shelf UIs developed by the manufacturers to control Spot and the Kinova arm separately from the hand-held controller provided by BD. We show these UIs in Figure 2.

Spot's UI allows the user to control the robot using the right and left joysticks for position and orientation. In addition, the UI shows views from the robot's cameras, which the user can change using a D-pad. Using the action menu or mode buttons, the user can access different gait modes available, e.g., stairs mode. The arm's UI provides multiple virtual joysticks and control modes for the arm and the gripper. Furthermore, the UI allows the user to select between predefined poses or sequences of poses.

B. *Colocalisation*

Colocalisation is the process that allows localising the robots and the HMD under a global coordinate system. The 3D poses included in the UI are relative to the HMD's frame of reference and can not be directly used by the robots. To find the transformation between their frames, we use a vision-based colocalisation method that extracts sparse visual features from images and 3D poses provided by the SLAM system of the HMD and the robots. These features are then used to localise the HMD and the robots against a spatially and temporally stable 6 DOF pose relative to these features. This pose is known as an anchor; several cloud-based services exist that can create these anchors for AR applications. For this work, we use Azure Spatial Anchors (ASA) [32], but others are available [33], [34].

Nonetheless, vision-based colocalisation approaches commonly generate an offset between the robot's and the HMD's estimate of the robot's pose. This offset increases as the robot moves further away from the anchor. As we will describe in the following sections, our experiment required navigating along a corridor and between floors. Therefore, we extended the colocalisation capabilities of the UI presented in [7] to support defining multiple anchors. Thus, reducing the effect of the colocalisation offset.

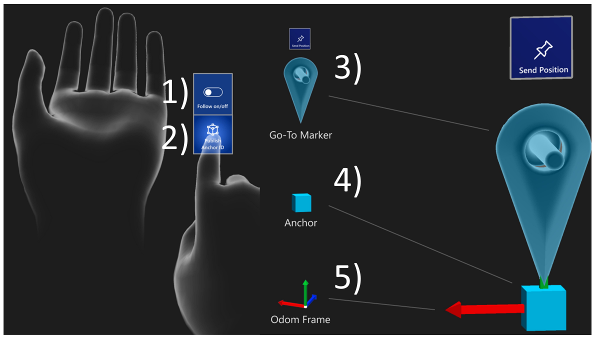

Fig. 3: AR HMD UI elements Left: raising the left palm displays two options for interacting with SpoK: 1) Follow and 2) Grasp. Right: UI components 3) GoTo marker, 4) Anchor, and 5) Spok's odometry frame icons.

The general workflow begins with the HMD defining the reference coordinate system by anchoring a reference frame to the environment, which allows it to persist in that location. Then, the user moves to a second location while the HMD tracks the relative movement. When the user creates a new anchor at this position, the HMD creates a connection by maintaining a relationship using this relative movement. Following the same procedure, the user creates as many anchors as needed for a particular application.

The HMD can localise the first anchor again in the future using image features and poses from its head-tracking system. Then, the HMD can query for nearby anchors using the connections created in the previous step. SpoK can localise against the same anchors by using grayscale images and camera intrinsics from one of its front-facing cameras and its respective poses relative to SpoK's vision odometry frame. Once the HMD and SpoK find the same anchors, we can find the transformation between their respective frames. The AR application uses the anchor closest to the user as a reference frame to minimise the colocalisation offset.

C. *AR HMD UI components*

The UI components provide several visual cues, control commands, and configuration options for the user (Figure 3). First, raising the left palm displays two options for interacting with SpoK: Follow and Grasp (see 1 and 2 in Figure 3). We use a coordinate system symbol to represent the origin of SpoK's vision odometry frame, i.e., the position where SpoK was when turned on (see 5 in Figure 3). We also included a GoTo marker to represent target poses for SpoK. The user can move and rotate this icon using hand gestures to place it at the desired location with the desired orientation (see 3 in Figure 3 and Figure 4a). Then, the user can press the button on top of the icon to send the desired pose to SpoK. Finally, we use a small blue cube to represent an anchor (see 4 in

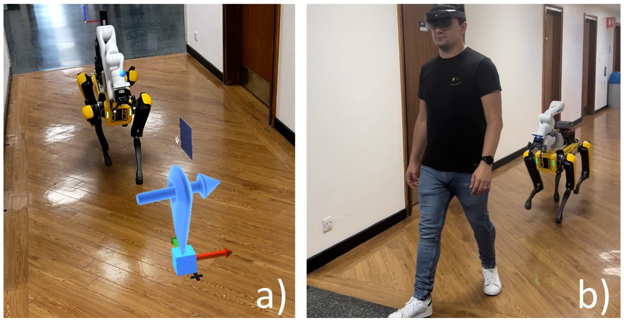

Fig. 4: AR HMD UI interaction methods. a) Operators can use the GoTo marker to define target positions and orientations for SpoK. b) The UI also incorporates a ”follow” behaviour that allows the user to naturally navigate its environment while SpoK maintains the distance between them within a fixed range.

Figure 3). This cube allows the user to visually verify that the HMD found and placed the anchor correctly. The UI also offers a ”follow” behaviour (Figure 4b). When using this behaviour, the user navigates its environment while SpoK maintains the distance between them within a fixed range. We provide further details about this UI in [7].

IV. USER STUDY

In this section, we describe our user study, which was designed to evaluate to what extent the proposed AR HMD UI affects factors relevant to a successful HRI. We hypothesised that, when compared to off-the-shelf control methods, our AR UI will improve HRI metrics on technology acceptance, cognitive workload, fluency, usability, and immersion.

A. Experimental design

We explored one main factor as an independent variable: UI (hand-held controller and AR HMD UI). Nevertheless, the user trials also involved various tasks categories: 1) navigating with the robot across an obstacle-free corridor, 2) navigating with the robot through wide and narrow door openings as well as going up a flight of stairs, and 3) picking an object from the floor while the robot remains still.

The primary dependent variables were the participants' level of acceptance of technology products, their perceived cognitive workload and fluency while performing the tasks, and their assessment regarding system usability and immersion. We collected these using the validated metrics introduced in subsection II-B. Furthermore, we used the time to complete the tasks to measure their performance. We also asked the participants to report their previous experiences with virtual reality (VR), AR, computer games, and robots.

B. Study Procedure

This study was approved by the Research Governance and Integrity Team from Imperial College London. We recruited 17 participants through email advertisements on our university campus. We did not offer financial incentives for participating in the study. We conducted the experiments in our laboratory. After signing a consent form, completing a safety screening questionnaire, and providing standard demographic data, we introduced the participants to the robot's parts and capabilities and explained the different tasks involved in the study. Next, we continued with a training session with both UIs. Then, participants continued with the experiments in five stages:

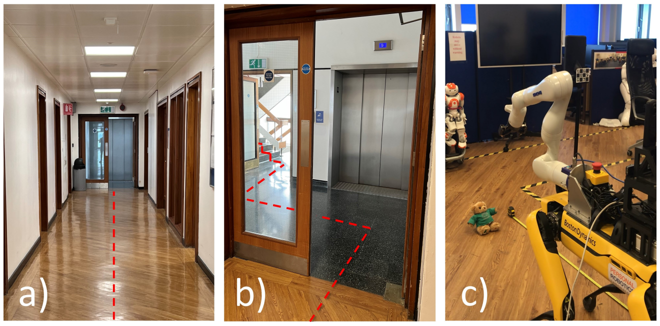

Fig. 5: User experiment tasks. a) Obstacle-free navigation involved navigating with the robot across an obstacle-free corridor. b) Stairs navigation involved navigating through doors and taking the robot up a flight of stairs. c) Object grasping involved picking an object from the floor while the robot remained still.

a) Obstacle-free navigation: this task involved navigating with the robot across an obstacle-free corridor. The start position was always the same and involved three waypoints, placed 5 m apart, along a straight line (see Figure 5a).

b) Stairs navigation: this second task involved navigating through a 90 cm wide door opening, a 180 cm door opening, and taking the robot up a flight of stairs along three waypoints. When taking the robot upstairs, we told participants not to stand above or below the robot. We illustrate the course for this task in Figure 5b.

c) HRI Questionnaires: after completing the first two tasks, we asked participants to complete a set of five questionnaires (see subsection II-B).

d) Object grasping: this last task involved picking an object from the floor while the robot remained still. The participants had a maximum of five minutes to successfully grasp an object placed on the floor 130 cm from the robot's base link (see Figure 5c).

e) Post-experiment questionnaire: finally, the participants completed a free text questionnaire elucidating their experience with the study.

The participants repeated steps a, b and c twice, once per UI. We counterbalanced for the UI the participants started with. For step d, participants chose their preferred UI to complete the task. Nonetheless, they had to remove whatever UI they had last before completing the questionnaires in stage c. This was to avoid participants basing their decisions on wanting to finish the experiment as fast as possible. We asked participants to justify their selection during step e. In addition, for a and b, we gave the participants a secondary task where they had to take pictures using a 360 camera while the robot was rotated perpendicular to the corridor at each waypoint. The 360 camera was onboard the robot, but the participants had to use an app running on a separate device to take the picture. We gave participants this secondary task to simulate an inspection task where they needed to use their hands to take the pictures in addition to controlling the robot.

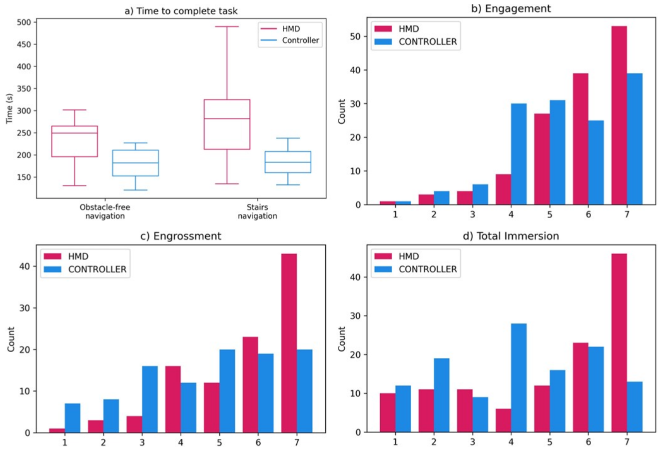

Fig. 6: a) Time to complete tasks for each interface. b) ARI engagement histogram. c) ARI engrossment histogram. d) ARI total immersion histogram. The "x" axis for b)-d) are response values in the Likert scale (1 Very Strongly Disagree, 7 Very Strongly Agree).

TABLE I: Demographics of the participants and their reported experience with other technologies ($N = 17$).

Factor	n	%
Gender		
Female	7	41
Male	10	59
Age (years)		
18–24	5	29.4
25–34	10	58.8
35–44	2	11.8

Baseline	n	%
VR	13	76.5
AR	13	76.5
Computer games	17	100
Robots	16	94.1

We performed the statistical analysis using two group comparison tests, i.e., T-Test and Mann-Whitney U-Test. We represented normally distributed values with means and standard deviations (SD), and non-normal distributed values as medians, modes, and inter-quartile ranges (IQR).

V. RESULTS AND DISCUSSION

We show the demographics of the participants and a summary of their reported experience with VR, AR, and computer games in Table I. In Figure 6a we summarise the time-related results for both groups. Regarding the task load, in Figure 7 and Table II, we show a summary of the results. In Table III, we present a summary of the results for all the remaining metrics. Finally, in Figure 6b-d, we provide a more in-depth analysis of the results of the ARI questionnaire.

From a performance perspective, the participants were significantly faster (p-value < 0.05) completing the tasks when using the controller than they were when using the AR UI: 178.27 s (SD =30.95) and 233.26 s (SD = 49.14) for obstacle-free navigation, and 179.84 s (SD = 46.93) and 269.31 s (SD = 83.55) for stairs navigation, respectively. We believe that the main reasons for this were: 1) the level of familiarity all participants had with game controllers, higher than their experience with AR and VR (see Table I), and 2) the difficulty some participants experienced positioning and rotating the GoTo marker, as well as while interacting with some UI buttons, which we noticed during the experiments and while analysing the video recordings.

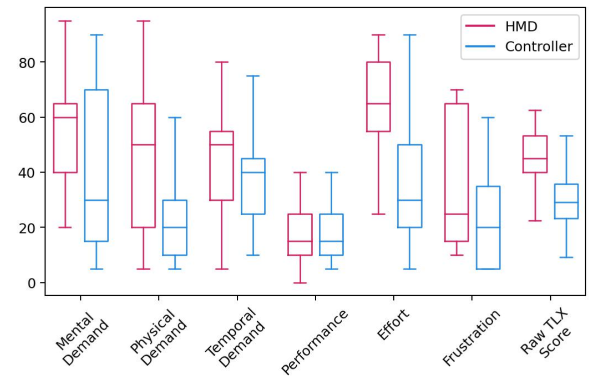

Fig. 7: Experiment results. NASA-TLX score box-plots. The six dimensions of the questionnaire are compared separate for both UIs. In addition, the raw NASA-TLX score is shown as well.

TABLE II: TLX score reported by participants (ns = no significant).

	MD	**PD**	**TD**	**P**	**E**	**F**	**Raw**
HMD							
Mean	55.3	43.8	44.1	23.8	62.1	35.6	44.1
SD	21.1	26.8	19.2	21.1	20.2	22.3	13.9
Controller							
Mean	40.9	24.4	38.8	23.5	36.2	24.1	31.3
SD	26.6	17.3	20.4	23.2	22.0	17.5	14.4
p value	ns	0.021	ns	ns	0.0015	ns	0.015

Our intuition is that these reasons might also have affected cognitive workload. The raw TLX score reported by the participants when using the controller was lower than when using the AR UI (see Table II). The statistical analysis of the raw TLX score revealed that the difference in task load reported for each control method is significant at $p < 0.05$. Nonetheless, when analysing the results for task load separately for each dimension in the questionnaire, we found that the main contributors to this difference are physical demand (PD) and effort (E).

We believe this increased PD derives from the inherently different approaches for interacting with the UIs. For the controller, the participants simply need to move a joystick in the direction they want the robot to move. In contrast, the AR UI requires engaging more body parts to use the "follow" behaviour or the GoTo marker and define the desired position and orientation for the robot. Moreover, the users had to press AR buttons to trigger the robot behaviours. In addition, the fact that some participants found manipulating certain UI elements challenging could explain the increased E reported.

Regarding the remaining NASA-TLX dimensions, we did not find a significant difference between control interfaces. Crucially, participants found the AR UI as mentally and temporally demanding (MD, TD) as the controller. They felt they performed (P) just as well with both UIs, and reported similar levels of frustration (F) between the two. We obtained a similar result for technology acceptance (TAM), fluency, and system usability (H-SUS). The statistical analysis revealed that the differences in these metrics reported by the participants are not significant. Thus, both UIs performed just as well in these dimensions.

Fig. 8: Cognitive offloading. a,b) When using the controller, some users rotated their bodies to align themselves with the robot and facilitate its positioning. c) We added new functionality to the AR HMD UI that allows the users to position themselves as required. d) After a voice command, the robot will position itself to match the user's pose. Watch our complementary video for more details.

Regarding immersion, the responses reported by the participants (ARI) were higher when using the AR UI than when using the controller. The analysis of the ARI results revealed that the difference in immersion reported for each control method is significant at $p < .05$. We found this to be consistent with the participants' qualitative answers. Furthermore, when analysing the results for immersion separately for each of the ARI questionnaire's dimensions (engagement, engrossment, and total immersion), we found that the difference remained significant for all of them. Therefore, resulting in higher interest, usability, emotional attachment, focus of attention, presence, and flow than when the participants used the controller (Figure 6b-d).

Regarding the object grasping task, eight of the participants (47%) chose to complete this task using the AR UI. The reasons given for this were obtained from the post-experiment questionnaire and included: enjoyment, ease of use, engagement, immersion, feeling focused, involved, and interested in the activity, finding it more intuitive, precise, novel, accurate, and less mentally demanding. Some participants explicitly mentioned finding their hands free for the secondary task convenient. The reasons for choosing the controller included performance, ease of use, reliability, sense of control, a more gentle learning curve, and their level of familiarity with this kind of UI. From this last group of participants, five expressed regretting their decision due to the complexity associated with controlling a robot arm using a virtual joystick. This only happened once with the AR UI.

Improving the UI and further user evaluation: We acknowledge that performance and cognitive workload are crucial metrics for the successful deployment of AR HMD UIs in the type of applications typical for legged manipulators, such as industrial applications. Therefore, aiming at improving these metrics and derived from our experience with the user study, we improved the functionality of the AR UI by exploiting the concept of cognitive offloading. Cognitive offloading refers to using physical activity to alter the information processing requirements of a task to reduce cognitive demand [35].

When using the controller, we noticed that some users rotated their bodies to align themselves with the robot and facilitate its positioning (see Figure 8a-b). Similarly, we added new functionality to the AR UI that allows the users to position themselves where they want the robot to go and orient themselves in the direction they want the robot to face. Then, after receiving a voice command, the robot positions itself, matching the required pose (see Figure 8c-d).

TABLE III: Results for technology acceptance (TAM), fluency, usability (H-SUS), and immersion (ARI) (ns = no significant).

	TAM	**Fluency**	**H-SUS**	**ARI**
HMD				
Median	4	4	4	6
Mode (count)	5 (65)	5 (82)	5 (71)	7 (142)
IQR	1.0	2.0	1.0	2.0
Controller				
Mean	4	4	5	5
SD	5 (75)	4 (85)	5 (88)	7 (72)
IQR	1.0	2.0	1.0	2.0
p value	ns	ns	ns	$2.56 \cdot 10^{-5}$

TABLE IV: Time to complete (s) and reported raw TLX score (lu) when using the added cognitive offloading feature. We report the values as (obstacle-free navigation, stairs navigation), respectively.

ID	Controller	AR UI	Improv. (%)	TLX
B1	(192.93, 289.85)	(162.94, 121.65)	(16, 58)	12.5
B6	(210.43, 183.05)	(89.68, 136.13)	(57, 26)	13.33
B8	(147.81, 211.39)	(96.37, 111.95)	(35, 47)	35.83
P5	(181.99, 179.25)	(106.65, 85.50)	(41, 52)	19.17
P6	(213.06, 159.56)	(103.25, 87.15)	(52, 35)	5.83

To demonstrate that this functionality positively impacts time performance and cognitive workload metrics, we performed further evaluation by asking five participants to repeat the navigation tasks once more using the added functionality. We recruited these participants at random. We summarised the results in Table IV. The added functionality drastically reduced the time to complete the task by an average of 40% (SD = 16%) and 44% (SD = 13%) for the obstacle-free and stairs navigation tasks, respectively.

Furthermore, the participants reported a lower raw TLX score, averaging 17.33 (SD = 11.37), 45% lower than with the controller (see Table II). Nevertheless, the fact that participants completed the trials before is an important confound and future work will focus on corroborating these results.

VI. CONCLUSIONS

This article provides a user study comparing an AR HMD UI we developed for controlling a legged manipulator against off-the-shelf control methods for such robots. We made this comparison baseline across multiple factors relevant to a successful interaction. Our results showed that although the AR UI increases immersion, off-the-shelf control methods outperformed the AR UI in time performance and cognitive workload. Nonetheless, a follow-up pilot study incorporating the lessons learned shows that AR UIs can outpace hand-held-based control methods and reduce the cognitive requirements when hands-free interactions and cognitive offloading principles are incorporated into the UI. Future work will investigate how the level of trust assigned by the operator to the robot transfers across tasks and environmental contexts and how AR HMD UIs affect such transfer.

REFERENCES

[1] C. D. Bellicoso, K. Krämer, M. Stäuble, D. Sako, F. Jenelten, M. Bjelonic, and M. Hutter, "Alma-articulated locomotion and manipulation for a torque-controllable robot," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 8477–8483.

[2] T. Klamt, D. Rodriguez, M. Schwarz, C. Lenz, D. Pavlichenko, D. Droeschel, and S. Behnke, "Supervised autonomous locomotion and manipulation for disaster response with a centaur-like robot," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 1–8.

[3] R. Chacon Quesada and Y. Demiris, "Augmented reality controlled smart wheelchair using dynamic signifiers for affordance representation," *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2019)*, pp. 4812–4818, 2019.

[4] ——, "Augmented reality user interfaces for heterogeneous multirobot control," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 11 439–11 444.

[5] ——, "Proactive robot assistance: Affordance-aware augmented reality user interfaces," *IEEE Robotics & Automation Magazine*, vol. 29, no. 1, pp. 22–34, 2022.

[6] J. Delmerico, R. Poranne, F. Bogo, H. Oleynikova, E. Vollenweider, S. Coros, J. Nieto, and M. Pollefeys, "Spatial computing and intuitive interaction: Bringing mixed reality and robotics together," *IEEE Robotics Automation Magazine*, pp. 2–14, 2022.

[7] R. Chacon Quesada and Y. Demiris, "Holo-SpoK: Affordance-aware augmented reality control of legged manipulators," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 856–862.

[8] F. Muhammad, A. Hassan, A. Cleaver, and J. Sinapov, "Creating a shared reality with robots," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2019, pp. 614–615.

[9] Z. Makhataeva and H. A. Varol, "Augmented reality for robotics: A review," *Robotics*, vol. 9, no. 2, p. 21, 2020.

[10] R. Suzuki, A. Karim, T. Xia, H. Hedayati, and N. Marquardt, "Augmented reality and robotics: A survey and taxonomy for ar-enhanced human-robot interaction and robotic interfaces," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–33.

[11] K. Ruffo and P. Milgram, "Effect of stereographic + stereovideo 'tether' enhancement for a peg-in-hole task," in *[Proceedings] 1992 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 1992, pp. 1425–1430.

[12] C. Hutton, N. Sohre, B. Davis, S. Guy, and E. S. Rosenberg, "An augmented reality motion planning interface for robotics," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, pp. 1313–1314.

[13] D. S. D. Caetano, C. Valentini, F. Mattioli, P. Camargos, T. Sá, A. Cardoso, E. Lamounier, and E. Naves, "The augmented reality telerehabilitation system for powered wheelchair user's training," *Journal of Communication and Information Systems*, vol. 35, no. 1, pp. 51–60, 2020.

[14] C. Reardon, K. Lee, J. G. Rogers, and J. Fink, "Augmented reality for human-robot teaming in field environments," in *Virtual, Augmented and Mixed Reality. Applications and Case Studies: 11th International Conference, VAMR 2019, Orlando, FL, USA, July 26–31, 2019*. Springer, 2019, pp. 79–92.

[15] S. Hashimoto, A. Ishida, M. Inami, and T. Igarashi, "Touchme: An augmented reality based remote robot manipulation," in *The 21st International Conference on Artificial Reality and Telexistence, Proceedings of ICAT2011*, vol. 2, 2011.

[16] S. Lee, N. P. Lucas, R. D. Ellis, and A. Pandya, "Development and human factors analysis of an augmented reality interface for multi-robot tele-operation and control," in *Unmanned Systems Technology XIV*, vol. 8387. SPIE, 2012, pp. 254–261.

[17] H. Fang, S. Ong, and A. Nee, "Interactive robot trajectory planning and simulation using augmented reality," *Robotics and Computer-Integrated Manufacturing*, vol. 28, no. 2, pp. 227–237, 2012.

[18] V. Alvarez-Santos, R. Iglesias, X. M. Pardo, C. V. Regueiro, and A. Canedo-Rodriguez, "Gesture-based interaction with voice feedback for a tour-guide robot," *Journal of Visual Communication and Image Representation*, vol. 25, no. 2, pp. 499–509, 2014.

[19] E. Rosen, D. Whitney, M. Fishman, D. Ullman, and S. Tellex, "Mixed reality as a bidirectional communication interface for human-robot interaction," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 11 431–11 438.

[20] W. P. Chan, G. Hanks, M. Sakr, H. Zhang, T. Zuo, H. M. Van der Loos, and E. Croft, "Design and evaluation of an augmented reality head-mounted display interface for human robot teams collaborating in physically shared manufacturing tasks," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 11, no. 3, pp. 1–19, 2022.

[21] J. A. Gomer and C. C. Pagano, "Nasa task load index for human-robot interaction workload measurement," *International Test and Evaluation Association Journal*, vol. 32, pp. 210–214, 2011.

[22] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS quarterly*, pp. 319–340, 1989.

[23] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, "User acceptance of computer technology: A comparison of two theoretical models," *Management science*, vol. 35, no. 8, pp. 982–1003, 1989.

[24] V. Venkatesh and F. D. Davis, "A theoretical extension of the technology acceptance model: Four longitudinal field studies," *Management science*, vol. 46, no. 2, pp. 186–204, 2000.

[25] G. Hoffman and C. Breazeal, "Cost-based anticipatory action selection for human–robot fluency," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 952–961, Oct 2007.

[26] G. Hoffman, "Evaluating fluency in human–robot collaboration," *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 3, pp. 209–218, June 2019.

[27] M. Paliga and A. Pollak, "Development and validation of the fluency in human-robot interaction scale. a two-wave study on three perspectives of fluency," *International Journal of Human-Computer Studies*, vol. 155, p. 102698, 2021.

[28] J. Brooke *et al.*, "Sus-a quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

[29] J. Baumgartner, N. Ruettgers, A. Hasler, A. Sonderegger, and J. Sauer, "Questionnaire experience and the hybrid system usability scale: Using a novel concept to evaluate a new instrument," *International Journal of Human-Computer Studies*, vol. 147, p. 102575, 2021.

[30] Y. Georgiou and E. A. Kyza, "The development and validation of the ARI questionnaire: An instrument for measuring immersion in location-based augmented reality settings," *International Journal of Human-Computer Studies*, vol. 98, pp. 24–37, 2017.

[31] S. Zimmermann, R. Poranne, and S. Coros, "Go Fetch!-Dynamic grasps using boston dynamics spot with external robotic arm," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 4488–4494.

[32] Microsoft, "Spatial anchors," https://azure.microsoft.com/en-us/services/spatial-anchors/#overview, 2021, accessed: 2021-09.

[33] Apple, "More to explore with arkit 5," https://developer.apple.com/augmented-reality/arkit/, 2021, accessed: 2021-09.

[34] Google, "Cloud anchors overview for android," https://developers.google.com/ar/develop/java/cloud-anchors/overview-android, 2021, accessed: 2021-09.

[35] E. F. Risko and S. J. Gilbert, "Cognitive offloading," *Trends in cognitive sciences*, vol. 20, no. 9, pp. 676–688, 2016.