



UNIVERSITY OF LEEDS

This is a repository copy of *Performance and Usability Evaluation Scheme for Mobile Manipulator Teleoperation*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/200851/>

Version: Accepted Version

Article:

Wan, Y, Sun, J, Peers, C et al. (3 more authors) (2023) Performance and Usability Evaluation Scheme for Mobile Manipulator Teleoperation. IEEE Transactions on Human-Machine Systems. ISSN 2168-2291

<https://doi.org/10.1109/THMS.2023.3289628>

This is an author produced version of an article accepted for publication in IEEE Transactions on Human-Machine Systems, made available under the terms of the Creative Commons Attribution License (CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Performance and Usability Evaluation Scheme for Mobile Manipulator Teleoperation

Yuhui Wan¹, Jingcheng Sun¹, Christopher Peers¹, Joseph Humphreys¹, Dimitrios Kanoulas² and Chengxu Zhou^{1*}

Abstract—This paper presents a standardised Human-Robot Teleoperation Interface (HRTI) evaluation scheme for mobile manipulators. Teleoperation remains the predominant control type for mobile manipulators in open environments, particularly for quadruped manipulators. However, mobile manipulators, especially quadruped manipulators, are relatively novel systems to be implemented in the industry compared to traditional machinery. Consequently, no standardised interface evaluation method has been established for them. The proposed scheme is the first of its kind in evaluating mobile manipulator teleoperation. It comprises a set of robot motion tests, objective measures, subjective measures, and a prediction model to provide a comprehensive evaluation. The motion tests encompass locomotion, manipulation, and a combined test. The duration for each trial is collected as the response variable in the objective measure. Statistical tools, including mean value, standard deviation, and T-test, are utilised to cross-compare between different predictor variables. Based on an extended Fitts' Law, the prediction model employs the time and mission difficulty index to forecast system performance in future missions. The subjective measures utilise the NASA-task load index and the System Usability Scale to assess workload and usability. Lastly, the proposed scheme is implemented on a real-world quadruped manipulator with two widely-used HRTIs, the gamepad and the wearable motion capture system.

Index Terms—Fitts' law, Human-Robot Interaction, Man-Machine systems, Teleoperation, Mobile Manipulator

I. INTRODUCTION

AS modern control methods evolve, robotic agents (RAs) have become increasingly powerful and intelligent. With the assistance of artificial intelligence, today's robot systems are nearly fully autonomous in factories and warehouses. However, the variability and complexity of tasks in open environments remain beyond the capability of autonomous RAs, particularly in emergencies. These tasks, such as hazardous materials (HAZMAT) rescue, HAZMAT decontamination, and explosive ordnance disposal (EOD), not only depend on complex real-time operations but also require the professional knowledge and experience of human agents (HAs).

This work was supported by the Engineering and Physical Sciences Research Council [grant numbers EP/R513258/1-2441459, EP/V026801/2], the Advanced Machinery and Productivity Institute [Innovate UK project number 84646] and the UKRI Future Leaders Fellowship [grant number MR/V025333/1] (RoboHike). For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Business, Environment and Social Sciences Research Ethics Committee of the University of Leeds (Reference No. MEEC 21-009).

¹School of Mechanical Engineering, University of Leeds, UK.

²Department of Computer Science, University College London, UK.

*Corresponding author, c.x.zhou@leeds.ac.uk

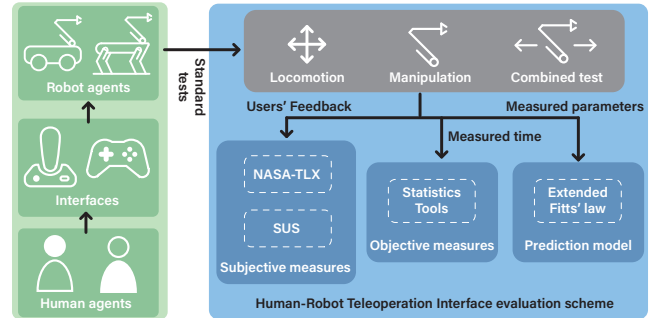


Fig. 1. Structure of the HRTI evaluation scheme for mobile manipulator applications.

Conversely, these tasks can potentially cause harm to HAs on the scene; thus, physical HA involvement must be minimised. Therefore, teleoperation at the motion level is one solution to bridge this gap, where HAs and RAs perform the mission to their advantages [1].

Nowadays, RAs serve public safety agencies in HAZMAT [2] and EOD [3] missions, primarily as mobile manipulators. In certain instances, mobile manipulators, especially quadruped manipulators, have advantages over HAs. For example, a human first-responder in a level-A HAZMAT suit with a self-contained breathing apparatus (SCBA) enclosed has their operation time limited by the size of the oxygen tank in SCBA, which is further affected by equipment weight and physical workload [4]. In contrast, a quadruped robot's operation time is mainly limited by its battery life (e.g., 2.5 - 4.5 hours for the Unitree AlienGo quadruped robot) and can be extended by an external power source. Moreover, the long-term operation cost of robots is lower than that of humans, and the size of quadruped manipulators can be smaller than human first responders. The smaller size offers increased manoeuvrability in confined spaces, which is crucial in HAZMAT missions. Thus, quadruped manipulators can be more suitable than human first responders for specific tasks. To achieve peak performance from such a robot system, operating with human intelligence, a Human-Robot Teleoperation Interface (HRTI) is key to leveraging the advantages of both robots and humans.

However, limited research explicitly focuses on teleoperation strategies for mobile manipulators and HRTIs [5]. Moreover, in recent years, many intriguing new technologies have been introduced into HRTIs, for instance, inertial measurement units (IMU) and visual recognition. With all these varied types of HRTIs, it is impossible to compare them directly side by side. Therefore, a standard HRTI evaluation scheme is critical for developing a mobile manipulator's teleoperation system.

In this study, a standard HRTI evaluation scheme for mobile

manipulators is designed. The scheme provides a comprehensive evaluation through a set of robot motion tests, both objective and subjective measures, and a quantified prediction model, as shown in Fig. 1. These measures comprise statistical side-by-side time comparisons for different types of motions, as well as first-hand user feedback. The prediction model takes both human and robot systems into account by utilising existing data to predict the performance of robot systems with HRTIs in future real-world tasks. Subsequently, an experiment on two HRTIs for a quadruped manipulator was conducted to test and refine the scheme. The detailed contributions include:

- 1) A standard HRTI evaluation scheme for mobile manipulators, which consists of three parts: (i) A set of standard motion tests, which examine locomotion and manipulation functionalities individually, and their combined performance; (ii) A separate objective measure using statistical tools to analyse the operator's motion time for performance evaluation of each motion; and (iii) A standardised model, extended from Fitts' Law, for predicting performance in future missions with existing standard test data.
- 2) Standardised subjective measures, containing NASA-task load index (NASA-TLX) and System Usability Scale (SUS), for workload and usability evaluation.
- 3) Evaluation and comparison of quadruped manipulator teleoperation performance and usability of two widely-used HRTIs, the conventional gamepad and the novel wearable motion capture system (WMCS), through the proposed scheme with experiments.

This article is organised as follows. First, the related works in robot teleoperation and HRTI evaluation are reviewed in Section II. Section III introduces the HRTI evaluation scheme. The proposed extended Fitts' Law model is detailed in Section III-C1. Then, Section IV presents the experimental hardware and design used to assess the HRTI evaluation scheme. Section V presents the actual experimental setup and user composition. Next, the results are demonstrated in Section VI, followed by their discussion in Section VII. Finally, Section VIII concludes this article.

II. RELATED WORK

Numerous scholars have conducted empirical studies on integrating human and robot intelligence in human-machine collaboration with HRTIs [1]. Over the years, various types of control interfaces have been developed and applied to robotic systems. However, determining the most effective control method remains a challenge. As a result, scholars advocate for the establishment of a standardised method to assess HRTI performance [6], motivating this research to explore a coherent approach for evaluating HRTI performance by introducing an extended Fitts' Law.

There are two primary types of interfaces between HAs and RAs. One allows HAs to utilise remote controllers, such as gamepads or keyboards, to interact with RAs [7]. Another permits HAs to use body movement captured by motion capture technology to interact with RAs [8]. In recent years, there has been a growing interest in applying motion capture technology in the robotic teleoperation context.

A. Gamepad Technologies

As one of the most widespread methods for controlling RAs, many researchers employ gamepads to control RAs in various applications, including nursing and assistive robots [9], [10]. Researchers also study the performance of gamepad teleoperation and compare it with alternative control methods, such as hand gesture control [11] and touch screen control [12]. Furthermore, most commercial quadruped systems utilise gamepads as the primary control method. However, no study provides explicit evidence on the performance of gamepads in quadruped manipulator teleoperation applications or compares gamepad control with motion capture technologies in mobile manipulator applications.

B. Motion Capture Technologies

In addition to traditional gamepad controllers, motion capture systems have emerged as a prevalent teleoperation technology among HRTIs, utilising input from cameras [13] or inertial measurement units [14]. Current motion capture technologies typically adopt a range of approaches, such as optical, inertial mechanical, magnetic, and acoustic techniques, while also employing programming by demonstration methodologies like keyframing and clustering to enhance their capabilities [15]. The majority of these studies focus on the development of motion capture technology itself but require further analysis of their performance. Moreover, very few works assess the performance of developed interfaces in robotic applications and open environment tasks. Consequently, additional evaluation is necessary to gain a better understanding of the practicality of motion capture systems as HRTIs in real-world missions.

1) *Vision-based Motion Capture Technologies:* Recent studies investigate using camera images as input for motion capture systems. The work in [16] exemplifies a camera-based motion capture method, in which a Microsoft Kinect V2 is adopted for human-body motion analysis. A baseline performance evaluation for the Kinect's depth tracking capabilities is conducted.

2) *Wearable Motion Capture Technologies:* Compared to vision-based motion capture systems, wearable systems are reported to have higher stability and better resistance to environmental disturbances, including changes in lighting and moving objects in the background. Additionally, studies achieve tracking of full-body motion through wearable motion capture suit systems and map it to an RA in real-time [17], [18]. Workspace mapping and path planning are accomplished by setting virtual obstacles to constrain the RA's motion, making them more user-friendly [19].

C. Evaluation

A task-based evaluation framework for teleoperation is presented in [20]. The evaluation framework comprises a task-based measurable parameter based on successful and unsuccessful movements, and user-opinion data is obtained through a questionnaire. More recent works offer a relatively comprehensive overview of methods that assess the performance and usability of operators in robotic scenarios [6], [21]. Although these studies propose using a standardised model to evaluate

HRTIs in robot applications, they are based on movement models in lower-dimensional, which can be oversimplified to describe real-world missions accurately. Therefore, further attention is required for evaluating robot teleoperation in real-world applications.

1) *Performance Prediction*: The performance of an HRTI has a significant impact on the system's efficiency in missions [22]. For fields other than robotics, there are numerous existing standardised measurement methods for evaluating a human-machine system's performance, including Fitts' Law [23], power model formulation [24], electroencephalograms (EEGs) [25], and electrocardiograms [26].

One of the most renowned analysis models is Fitts' Law [23], which was developed from research on the performance of HAs interacting with computing systems by P.M.Fitts. Fitts' Law is a widely used predictive model for a human-machine interface's performance. It predicts the motion time (MT) for HAs to complete a motion with a specific interface through the index of difficulty (ID), as shown below:

$$MT = a + b \cdot ID, \quad (1)$$

where a and b are constants based on the system, and b measures the rate of change of motion time with the change in motion difficulty. ID ranges from 0 to infinity, and due to the linear relationship, the motion becomes impossible at infinite ID . The original ID comprises two parts: the target distance (d) and the target width (w).

An alternative to Fitts' Law, proposed by Kvålseth, is called power model formulation [24]. It has three empirically determined constants, while Fitts' law only has two, thus can provide higher multiple correlations. However, it has not been widely adopted due to its complexity.

Throughout the years, there have been many modified versions of modelling ID in Fitts' Law. One of the most well-known models is proposed by MacKenzie [27], also known as the Shannon formulation:

$$ID = \log_2\left(\frac{d}{w_t} + 1\right). \quad (2)$$

However, Fitts' Law is a one-dimensional predictive model to measure motion. A recent line of research focuses on extending the application to two-dimensional target acquisition [28]. Motivated by the Shannon formulation (2), Stoelen and Akin combined both translation and rotational motion in the ID [29]. In their model, respective rotational distance (α) and rotational tolerance (θ) of the probe are added into consideration, as shown below:

$$ID = \log_2\left(\frac{d}{w_t} + 1\right) + \log_2\left(\frac{\alpha}{\theta} + 1\right). \quad (3)$$

However, although both translational and rotational difficulty are considered in the total ID , the translational and rotational movements were performed independently with two different cursors.

Cha and Myung's version of the ID is also based on the Shannon formulation (2), but it took into account the size of probe (f), the finger pad size of HAs in their case [30]:

$$ID = \log_2\left(\frac{2d}{w_t + f}\right). \quad (4)$$

Although these contributions to Fitts' Law make the model more flexible and adaptive, the application is still limited to direct manipulation of probes (mice and fingers) and oversimplified for real-world robot missions. It is necessary to have a new model that better represents the real-world robot mission characteristics for HRTIs.

2) *Usability and Workload Evaluation*: For subjective response measurements, there are two main topics: the system's mental [31] and physical workload on the operator, and its usability [32]. The NASA-TLX is most prominently used for measuring subjective cognitive demand [33]. NASA-TLX employs a questionnaire with asymptotic performance evaluation and the assessment of various aspects (e.g., mental and physical demand, temporal demand, and effort). Research indicates that NASA-TLX is more popular than other models in real-world engineering tests [34]. Another similar questionnaire, the NASA situation awareness rating technique (SART), is more commonly employed in evaluating teleoperation with video feedback [35]. However, the SART focuses on the HA's awareness of their surrounding environment rather than machinery operation. To enrich the subjective understanding of the HRTI, some research also benefits from usability tests. To standardise the usability test, the SUS [36] is introduced. It is shown to be easy to understand for regular users and is widely used across all industries [37], [38].

III. METHODS OF EVALUATION SCHEME

The proposed HRTI evaluation scheme has four major components, standard tests, objective measures, prediction model, and subjective measures, as shown in Fig. 1. The standard tests collect time-related data for mathematical models to measure performance and generate first-hand user experience for usability measurements.

A. Standard Tests

The standard tests provide data for quantitative analysis of the robot's performance for different motions. The "standard" indicates the tests should contain three categories for evaluation: locomotion tests, manipulation tests, and combined tests. The presented extended Fitts' Law models systems based on the variation of mission difficulty, so different missions and HRTIs for the same RA can still be cross-compared. At least one test for each category is required, and additional tests with diverse difficulty can increase the particulars and accuracy of the result. For evaluation, the study was designed with three standard tests and examined the prediction model accuracy with a real-world exercise, which is discussed in Section IV.

B. Objective Measure

The quantified model provides data on the performance of the system. First, the system measures the motion time to complete each trial as response variables. Then, the motion time is cross-compared between different predictor variables, such as HRTIs and user groups, through mean value, standard deviation, and T-test or ANOVA. From these statistical comparisons, the performance characteristics of each system

in different motions are understood. Furthermore, the number of attempts and other measurements according to mission circumstances contribute to evaluating the targeted systems.

C. Prediction Model

The prediction model is based on the Shannon formulation [27] and Fitts' Law [23]. The new model provides an overall forecast for system performance in real-world missions based on standard tests. It evaluates mission difficulty through RA and target position information. It explores the relationship between motion time and the mission's difficulty.

1) *Extended Fitts' Law*: Fitts' Law and its recent iterations still struggle to model systems with complex control in real-world robot applications. Specifically, the relationship between the facing direction of the RA and the orientation of the target cannot be represented by the ID. This work develops a new model that can better reflect motion difficulty in the ID. Similar to Stoelen and Akin's work, which treats figures as a probe [29], this work considers the RA's end-effector as a probe. Additionally, the size, location, and orientation of both the target and RA, as shown in Fig. 2, are taken into account.

In detail, the translation index of difficulty (ID_{trans}) considers the translation part of the motion, including effective distance to the target (d), the target width (w_t), the RA width (w_e), and the mission requirement:

$$ID_{trans} = \log_2\left(\frac{d}{w_t \pm w_e} + 1\right). \quad (5)$$

The relationship between RA width (w_e) and target width (w_t) is defined by each mission. For example, a locomotion mission requires any part of the RA to reach the target area, the RA width (w_e) is the full diagonal width of the robot body, and the (5) uses $w = w_t + w_e$. Conversely, in a manipulation mission, the arm needs to fit inside a target hole. Therefore, the RA width (w_e) becomes the diagonal width of the end-effector, and the (5) uses $w = w_t - w_e$.

The orientation index of difficulty (ID_{ori}) takes into account the target tolerance angle (θ) and the angle between the RA's starting location from the target and the target facing direction (α):

$$ID_{ori} = \log_2\left(\frac{\alpha}{\theta} + 1\right). \quad (6)$$

The facing direction index of difficulty (ID_{dir}) considers the angle between the RA facing direction and the target direction (β in degrees):

$$ID_{dir} = \log_2\left(\frac{\beta}{180^\circ} + 1\right). \quad (7)$$

Therefore, the complete standardised prediction model using Fitts' Law (1) can be calculated using the extended ID as

$$ID = \sum_1^n ID_{trans_i} + ID_{ori} + ID_{dir}. \quad (8)$$

In this new model, when the target is facing the RA's starting point ($\alpha = 0^\circ$), the RA is facing the target ($\beta = 0^\circ$), and the RA size is small enough to be neglected ($w_e \rightarrow 0$), the extended version of the ID in (8) becomes the same as the Shannon formulation (2).

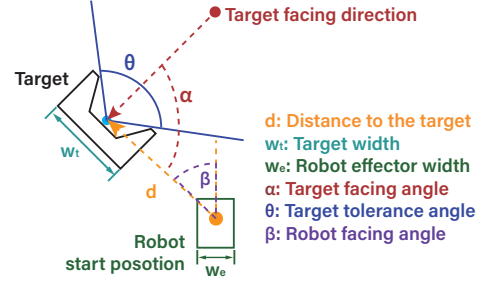


Fig. 2. Parameters used to calculate the extended ID.

Although each mission is treated as a whole, the total ID can contain more than one translation index of difficulty ($\sum_1^n ID_{trans_i}$). For example, in the combined test, the RA first locomotes into the arm's reachable distance (l) and then manipulates the arm to reach the target. Since it is not feasible to predict where each user will stop the RA, locomotion motion cannot be separated from manipulation motion. However, the RA always stops within the arm's reachable distance from the target and then performs manipulation. Therefore, the combined motions are simplified, where the locomotion part uses the total distance minus the arm length ($d - l$) as the target distance in (5), and the manipulation part uses the arm length (l) as the target distance in (5).

Furthermore, multiple steps motions only consider orientation and facing direction once in the ID. Since the RA already stops within the target tolerance angle (θ) from the locomotion step, the orientation index of difficulty (ID_{ori}) is only calculated once at the initial position. Also, due to the manipulator being more flexible than the trunk, the manipulator's starting facing direction (ID_{dir}) is overlooked. Therefore, in the combined test, IDs consider multiple translations, but only one orientation and one facing direction ID.

The new model reflects the HRTI and its relationship with the field environment in the mobile manipulator teleoperation tasks. Later on, this extended Fitts' Law is deployed to analyse a real-world system and provide evidence on the performance characteristics of different HRTIs for a quadruped manipulator.

D. Subjective Measures

The subjective measures analyse the usability and workload of the system. NASA-TLX is employed for workload measurement, including mental task load and physical task load. NASA-TLX was initially designed to evaluate comprehensive workload on pilots in aircraft. Therefore, only pertinent questions on mental and physical demand are selected based on short-term robot teleoperation applications, and questions that require longer-term experience are removed.

SUS provides details on usability with ten standard statements and 5 response options (scoring from 1 "Strongly disagree" to 5 "Strongly agree") for each statement, as shown in Table VI. Half of these statements in SUS are positive, and half are negative. This unique structure reduces acquiescent bias and extreme response bias. However, this makes the comparison of results less intuitive. To better analyse the results, the SUS user responses are converted into a converted score (the higher, the better).

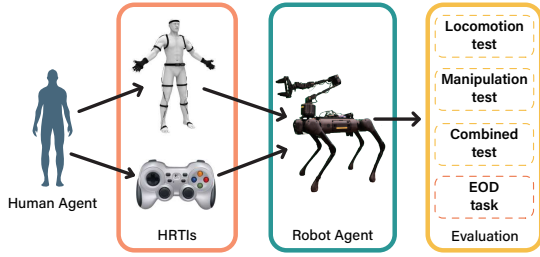


Fig. 3. Structure of the quadruped manipulator HRTI evaluation experiment.

IV. EXPERIMENT DESIGN

To conduct a pilot test of the HRTI evaluation scheme, an experiment involving a quadruped robot with two HRTIs is designed and implemented. The experiment is organised with standard tests and an additional real-world task. The results are evaluated using objective measures, the prediction model, and subjective measures. The standard tests consist of one locomotion test, one manipulation test, and one combined test. The real-world task is a simulated EOD operation. All missions are independent of one another and are reset after each trial. Moreover, the IMU sensor on the WMCS is calibrated between missions.

The hardware for this experiment is divided into two parts: the robot hardware and the teleoperation hardware. The overall experiment structure is shown in Fig. 3.

A. Robot and HRTI Hardware

This study employs the Unitree AlienGo quadruped robot with a Trossen Robotics ViperX 300 robot arm as the platform [39], featuring an arm length of 75 cm. Additionally, the ViperX 300 robot arm has been redesigned to reduce its overall weight [40]. The integrated legged manipulator is controlled using a customised whole-body controller [41]. We compare two types of HRTIs for quadruped manipulators: a traditional gamepad and a WMCS. Consequently, the Logitech F710 wireless gamepad and the Noitom Perception Neuron inertia-based motion capture suit (selected for its stability) are used as the interfaces in this study. Both HRTIs map human inputs to the teleoperation strategies detailed in Section IV-B.

B. Teleoperation Strategies

Since the HA and the RA are not kinematically similar, directly mapping the HA's body joints to the RA's joints is unfeasible. Therefore, a set of robot teleoperation strategies are designed to provide intuitive HRTI control. The HRTI control logic is divided into two groups of robot strategies: trigger and argument. The trigger strategies switch between different modes, and argument strategies provide the magnitudes of the motions, as shown in Table I. Both HRTIs share these strategies to minimise variables during the comparison.

1) Trigger Strategies:

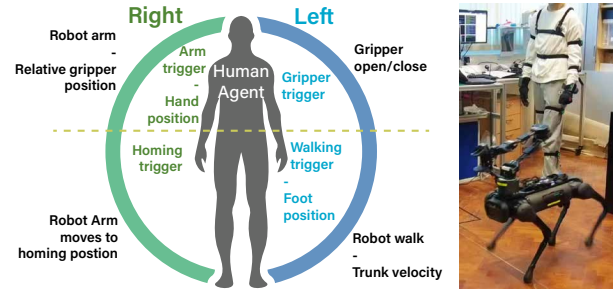
- *Walking trigger*: This trigger activates walking mode. The robot may perform locomotion motion only when the walking mode is activated.
- *Arm trigger*: This trigger activates manipulation mode. The arm can only move once this trigger is activated.

TABLE I
TELEOPERATION STRATEGIES AND HRTIS

Robot strategies	HRTIs		
	Gamepad	WMCS	
Trigger strategies	Walking trigger	LT	Right hand down
	Arm trigger	LB	Left hand up
	Gripper trigger	RT	Right hand up
	Homing trigger	RB	Left hand down
Argument strategies	Walking arguments	LS + RS	Relative foot position
	Arm arguments	LS + RS	Relative hand position



(a) Gamepad



(b) WMCS

Fig. 4. Details of mapping from interfaces to trigger and argument strategies, and experiment operation example: (a) gamepad, (b) WMCS. For the WMCS, each trigger is active by the user closing his/her hand.

- *Gripper trigger*: This trigger activates the closing motion of the gripper on the end of the robotic arm manipulator, and the gripper will remain closed until this trigger is released.
- *Homing trigger*: This trigger activates the arm homing command, which returns the arm to its home position.

2) *Argument Strategies*: While the WMCS collects 3D motion, the joysticks on the gamepad only collect 2D motion. Therefore, a pair of joysticks, left stick (LS) and right stick (RS), on a gamepad, are used, as shown in Fig. 4.

- *Walking arguments*: In walking mode, the arguments of trunk velocity are sent to the robot. The velocity has three components: forward/backward, left/right, and counter-clockwise/clockwise. The magnitude of these arguments is linear to the relative velocity components.
- *Arm arguments*: This trigger activates manipulation mode. In this mode, the arm can move its end-effector in position, including moving forward/backward, up/down, and rotating the base joint counter-clockwise/clockwise.

C. Modelling Standard Tests

Three tests are designed based on the parameters in the IDs of extended Fitts' Law (8), as shown in Table II. In detail, the locomotion test requires the RA to walk from the

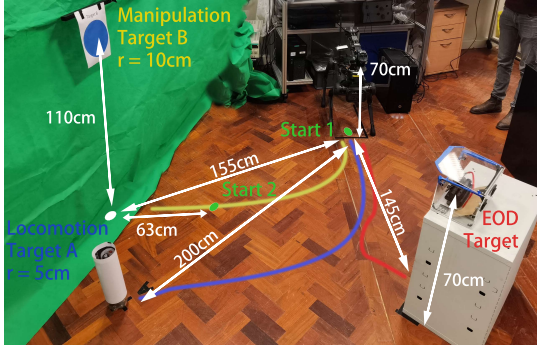


Fig. 5. Experiment setup of start point “1” and “2”, locomotion test target “A”, manipulation and combined test target “B”, and “EOD” target.

TABLE II
PARAMETERS OF THE EXPERIMENT DESIGN

Mission:	d	w_t	w_e	α	θ	β	l	ID
Locomotion	200	10	69.46	0	360	45	0	2.14
Manipulation	75	20	12.53	75	180	75	75	2.73
Combined	160	20	12.53	20	180	70	75	4.20
EOD	145	6	2	45	90	5	75	10.08

(d, w_t, w_e, l) in cm, (α, θ, β) in degree.

starting position to cylinder target “A”, as shown in Fig. 5 (path illustrated as a blue line). The manipulation test starts with the RA standing next to target “B”, and the RA moves the robotic arm to use the end-effector to touch target “B”, as shown in Fig. 5. During the combined test, the RA first walks from the starting position towards target “B”. After the RA stops at a convenient position, it uses the end-effector to touch target “B”, as shown in Fig. 5 (path illustrated as a yellow line).

The IDs of the locomotion test and the manipulation test can directly apply designed parameters into (5) to (7), using

$$ID = ID_{trans} + ID_{ori} + ID_{dir}. \quad (9)$$

The combined test has multiple translation steps and consequently requires a summation of multiple translation indices of difficulty ($\sum_1^n ID_{trans_i}$). The RA approaches the target with locomotion motion until the end-effector can reach the target:

$$ID_{trans_1} = \log_2\left(\frac{d-l}{w_t+w_e} + 1\right), \quad (10)$$

then completes the mission with manipulation motions:

$$ID_{trans_2} = \log_2\left(\frac{l}{w_t+w_e} + 1\right). \quad (11)$$

Therefore, the combined test has two steps of translation motion, as shown below:

$$ID_{combined} = ID_{trans_1} + ID_{trans_2} + ID_{ori} + ID_{dir}. \quad (12)$$

D. Modelling Real-world Exercise

An EOD task has been simulated in real-world exercises, which requires the RA to disable a “bomb”. This task has three steps. Step one: the RA walks from the starting point “1” towards the “EOD target”, as shown in Fig. 5 (path illustrated as a red line). Step two: the RA uses its arm and end-effector to open the “bomb” box. Step three: the RA unplugs a red wire from the “bomb” to disable it. The parameters in the EOD

task also follow the standard test procedures and are used to calculate the ID for each step, as shown in Table II.

Therefore, the EOD task has three steps of translation motion: locomotion to approach the target:

$$ID_{trans_1} = \log_2\left(\frac{d-l}{w_{t_3}-w_e} + 1\right), \quad (13)$$

manipulation to open the box:

$$ID_{trans_2} = \log_2\left(\frac{l}{w_{t_2}-w_e} + 1\right), \quad (14)$$

and manipulation to pull out the wire:

$$ID_{trans_3} = \log_2\left(\frac{l}{w_{t_3}-w_e} + 1\right). \quad (15)$$

In manipulation motions, the end-effector (w_e) needs to fit inside the box opening (w_{t_2}) as well as into the gap between the wire and the “bomb” body (w_{t_3}), which limits the available target width in the model. In locomotion motion (ID_{trans_1}), the RA needs to stop at a location suitable for the most difficult manipulation motion. Otherwise, they have to readjust. Consequently, the locomotion parameters should consider the most challenging motion. Thus, the EOD task’s ID becomes:

$$ID_{EOD} = ID_{trans_1} + ID_{trans_2} + ID_{trans_3} + ID_{ori} + ID_{dir}. \quad (16)$$

V. EXPERIMENT PARTICIPATION

The volunteer users first complete a pre-test questionnaire to provide a baseline of their background experiences. At the beginning of the experiment, basic training is provided. They then undertake the experiment with two cameras recording the entire process. After the experiment, users receive another questionnaire to evaluate their experience with the HRTI.

A. Basic Training

Initially, the users watch a brief demonstration video showing real-world exercises performed by an expert using the WMCS, giving them an overview of the system and operation. After the video, the users are instructed on the manoeuvres for the gamepad and the WMCS. Subsequently, they are briefed on upcoming missions. Additionally, a physical copy of the graph of instructions for both types of interfaces is made available to the users, as illustrated in Fig. 4(a) and Fig. 4(b), to help them memorise the commands during the missions.

B. Experiment Performing

After ensuring users understand teleoperation strategies and mission requirements, they proceed to perform experiments. To minimise bias between the two HRTIs due to the learning curve, five randomly selected users are required to perform the standardised tests and the real-world exercise using the gamepad first, as shown in Fig. 4. They then repeat the same process using the WMCS, as demonstrated in Fig. 4. The remaining five users perform these experiments in reverse order, with the WMCS first, followed by the gamepad.

The user stands next to the robot and can move around while performing teleoperation. However, it is ensured that they do not move into the RA’s trajectory. Moreover, there is no time limit for each trial.

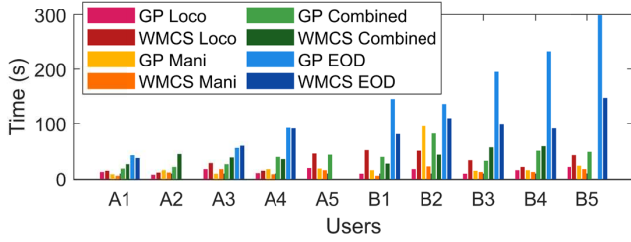


Fig. 6. Motion time of users complete missions with the gamepad (GP) and WMCS. Users A1—A5 are group A, and users B1—B5 are group B.

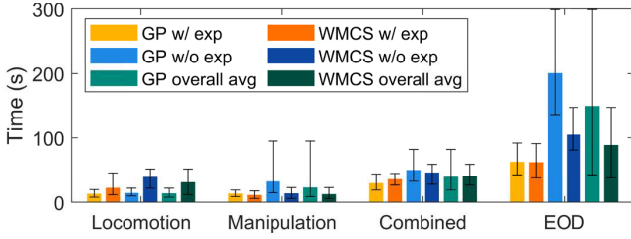


Fig. 7. The side-by-side comparison of the motion time to complete each mission between user group A with past gamepad experience, user group B without past gamepad experience, and the total average of all the users.

C. Volunteer Constitution

Ten randomly selected volunteers with various backgrounds participated in the experiment as HAs. Among them, there were six males and four females, ranging in age from 20 to 32. Additionally, half of the volunteers had experience with gamepads. The volunteers with gamepad experience were organised into group A, listed as users A1 - A5 in Fig. 6. The rest of the volunteers belonged to group B, listed as users B1 - B5. None of them had experience with WMCSs in the past. Furthermore, three of the users with gamepad experience had engineering or robotics backgrounds (users A1, A3 and A4).

VI. RESULT

The proposed HRTI evaluation scheme was implemented to evaluate the performance and usability of HRTI systems with different user groups. The results reveal a noticeable difference in some missions between the two HRTIs.

A. Objective Measure

The objective measure, through Matlab and Excel, utilises the time taken for all ten users to complete the experiment missions as response variables, as shown in Fig. 6. The time is measured by three personnel individually through video recording and then averaged. There are two predictor variables in this experiment: the HRTI used and user's prior experience with gamepads. Fig. 6 compares all users' completion times for each mission side-by-side with the two HRTIs.

Fig. 7 displays the mean and range of time results for user groups A and B to compare the performance of users with different experiences. The study uses statistical tools to compare the objective measurements, including mean value, standard deviation, and P-value from T-tests, as shown in Table III. Moreover, during the EOD task, on average, users completed the task in 3.4 attempts while using the gamepad and in 1.8 attempts while using the WMCS.

TABLE III

STATISTICAL ANALYSIS SHOWS THE MEAN, STANDARD DEVIATION (STD), AND P-VALUE COMPARING TWO HRTIS AND TWO GROUPS OF USERS. THE MOST REPRESENTATIVE RESULTS (P-VALUE <0.1) APPEAR IN BOLD, AND SIGNIFICANT RESULTS (P-VALUE <0.05) APPEAR IN ITALIC.

Users	HRTI	Result	Locomotion	Manipulation	Combined	EOD
All users	GP	Mean	14.60	23.80	40.70	149.88
		Std	4.59	24.07	17.20	82.90
	WMCS	Mean	31.90	13.40	41.88	89.50
		Std	14.73	5.22	11.14	30.41
		P-value	0.002	0.086	0.364	0.016
Group A (w/ exp)	GP	Mean	14.00	14.40	30.40	63.67
		Std	4.43	4.13	9.89	20.73
	WMCS	Mean	23.40	12.20	36.75	63.00
		Std	12.75	4.40	6.50	21.74
		P-value	0.050	0.237	0.089	0.411
Group B (w/o exp)	GP	Mean	15.20	33.20	51.00	201.60
		Std	4.66	31.07	16.79	60.04
	WMCS	Mean	40.40	14.60	47.00	105.40
		Std	11.25	5.68	12.39	22.84
		P-value	0.007	0.119	0.379	0.008
Group A VS. Group B	GP	P-value	0.360	0.141	0.057	0.005
	WMCS	P-value	0.062	0.206	0.094	0.093
Group B ALL	GP	P-value	0.054	0.102	0.016	0.012
	WMCS	P-value				

The study sets the most representative results (p-value <0.1) and significant results (p-value <0.05) based on the commonly used value for human-robot interfaces [42]. As seen in Table III, for all users, there were no statistically significant results for the two HRTIs in the combined test. However, for all other missions, the result shows the most representative. Furthermore, the significant results of two HRTIs in the locomotion test and EOD task indicate more significant performance differences between the two HRTIs. It is also interesting to observe changes in statistically significant results between two HRTIs within group A compared to within group B in the following metrics: (i) higher performance advantage with the gamepad in the locomotion test (p-value = 0.05 vs. p-value = 0.007); (ii) lower performance advantage with the WMCS in manipulation test (p-value = 0.237 vs. p-value = 0.119); (iii) reversed result in the combined test; and (iv) higher performance advantage with the WMCS in manipulation test (p-value = 0.411 vs. p-value = 0.008).

The standard deviation indicates the variance in user performance. From Table III, for all users, the gamepad has smaller standard deviations in the locomotion test (std = 4.59 vs. std = 14.73), and the WMCS has smaller standard deviations in the manipulation test (std = 5.22 vs. std = 24.07) and in the EOD task (std = 30.41 vs. std = 82.90). Additionally, user group A has smaller standard deviations in most missions, except for the locomotion test with the WMCS.

The difference between the performance of the two user groups using the same interface is also intriguing. The study performs a T-test on the results of the two user groups in each standard test and the EOD task, as shown in the last three columns of Table III. There is a more considerable performance difference between the two user groups with the WMCS in the locomotion test than with the gamepad (p-value = 0.062 vs. p-value = 0.36). Intuitively, in a more complex EOD task, the performance advantage of gamepad users is more significant while operating with the gamepad (p-value = 0.005 vs. p-value = 0.093). Also, for all users, significant results appear in the combined test and the EOD tasks, showing greater performance differences between the users in these missions.

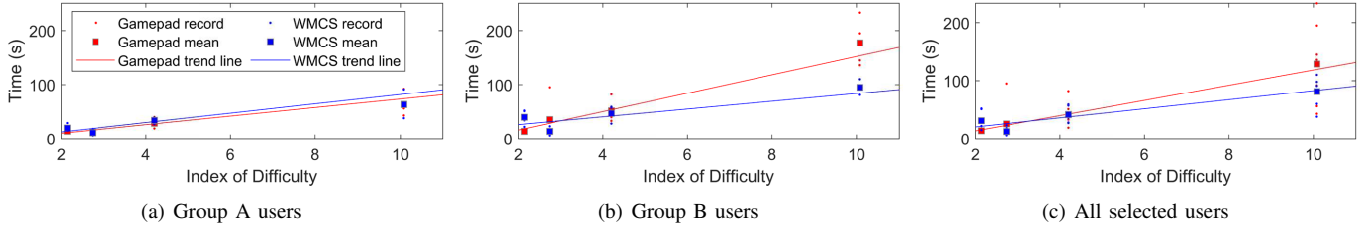


Fig. 8. The motion time of users took to complete each mission, which represents by different IDs, and the fitted linear polynomial line for average motion time: (a) user group A, (b) user group B, and (c) all 7 selected user. (The lower the motion time, the better performance)

B. Prediction Model using the Extended Fitts' Law

This research explores an extended Fitts' Law with a more detailed ID for real-world robot teleoperation applications. Fitts' Law suggests that the time required to complete a motion is positively correlated with the ID. A motion with a higher ID will take longer to complete.

Due to technical issues, three users did not complete specific missions during the experiment. The performance of the remaining seven users who completed all four missions with both HRTIs was selected to examine the proposed prediction model. Specifically, they are users A1, A3, A4, and B1 - B4, from Fig. 6. This research utilises Matlab to build the model and predict future mission performance. Firstly, the previously calculated extended IDs (9) (12) (16) for each mission in the experiment, based on their environmental and targeting characteristics, were revisited, as shown in Table II.

Then, the extended ID values were plotted alongside the user group's average motion time for the standard tests, as shown in Fig. 8. In the graphs, four missions from left to right are the locomotion test, manipulation test, combined test, and EOD task. A linear polynomial line was fit to the average time from three standard tests, with constants a and b from extended Fitts' Law shown in Table IV, and the motion time for the EOD task was predicted using this line. Firstly, the extended Fitts' Law was used to model the three users from group A. The linear polynomial curves fitted to the data by Matlab have root-mean-square deviation (RMSE) values of 4.65 and 10.12 for the gamepad and WMCS. The model also predicts the performance of the gamepad will be better than WMCS with user group A, as shown in Fig. 8(a). This prediction reflects real-world experience, since users from group A were more familiar with the gamepad than the WMCS.

Next, the four users from group B were modelled. The linear polynomial curves fitted to the data by Matlab have RMSE values of 8.81 and 22.26 for the gamepad and WMCS. In Fig. 8(b), the lines for HRTIs cross each other around 3.1 ID. This means the gamepad was better in missions with lower difficulty, and the WMCS performed better in more complex missions for user group B. Additionally, it is observed that user group A has a smaller RMSE than user group B (4.65 vs. 8.81 and 10.12 vs. 22.26), which indicates more accurate modelling. This accurate modelling leads to a more precise prediction in WMCS, with less difference in measured MT (-30.79% vs. 11.42%) from Table IV.

Although groups A and B users have different gamepad experiences, they still share many other similar characteristics.

TABLE IV
CONSTANT a AND b IN FITTS' LAW AND THE DIFFERENCE BETWEEN PREDICTED MOTION TIME AND AVERAGE MEASURED TIME.

Users:	HRTI	Measured MT	a	b	Predicted MT	Difference
Group A	Gamepad	63.67s	-5.504	7.884	73.98s	-16.18%
	WMCS	63s	-4.513	8.622	82.40s	-30.79%
Group B	Gamepad	177.25s	-17.89	17.00	153.45s	13.43%
	WMCS	95s	11.80	7.178	84.16s	11.42%
All users	Gamepad	128.57s	-12.58	13.09	119.39s	7.14%
	WMCS	81.29s	4.807	7.797	83.40s	-2.6%

TABLE V
MEAN SCORES FOR NASA TASK LOAD INDEX, ON A SCALE OF 0 TO 100. (THE LOWER SCORE, THE LOWER WORKLOAD, MARKED IN BOLD.)

NASA-TLX		Gamepad	WMCS
Mentally demanding:	Locomotion	52	51
	Manipulation	67	35
	Combined and EOD task	69	49
	Average workload	63	45
Physically demanding:	Locomotion	27	55
	Manipulation	39	32
	Combined and EOD task	43	45
	Average workload	36	44

Therefore, finally, all seven selected users were treated as a whole group, and the proposed prediction model was applied, as shown in Fig. 8(c). The linear polynomial curves fitted to the data by Matlab have RMSE values of 3.04 and 17.06 for the gamepad and WMCS. The extended Fitts' Law has two linear lines for two HRTIs intersecting at around 3.5 ID. This indicates the gamepad was better in easier missions, and the WMCS performed better in more difficult missions. As the number of users increases, it is evident that the difference between predicted and measured motion time is reduced, as shown in Table IV. The predictions for both HRTIs have less than a 10% difference. Hence, a larger sample size leads to a more accurate prediction model.

The proposed extended Fitts' Law successfully predicted the performance of different user groups in the experiment. It is understood that the accuracy is related to the sample size of targeted user groups and the experience of the users. Furthermore, from the power-analysis result (power of 0.8 with a 0.1 type I error rate), this group of seven users is sufficient to distinguish the performance of the two HRTIs in quadruped manipulator teleoperation.

C. Subjective Measure

Subjective measures were collected through questionnaires from users after the experiment. Standardised forms employed in the measures include NASA-TLX and SUS. In general, 58% of users prefer to use the gamepad over WMCS in

TABLE VI
USERS' AVERAGE SCORES FOR SYSTEM USABILITY SCALE (ON A SCALE OF 1 (STRONGLY DISAGREE) TO 5 (STRONGLY AGREE)).

No.	System Usability Scale statements	Response type	User response		Converted score	
			Gamepad	WMCS	Gamepad	WMCS
1	I think that I would like to use this system frequently.	Positive	3.2	3.4	2.2	2.4
2	I found the system unnecessarily complex.	Negative	2.5	2.6	2.5	2.4
3	I thought the system was easy to use.	Positive	4.3	2.7	3.3	1.7
4	I think that I would need the support of a technical person to be able to use this system.	Negative	2.1	4	2.9	1
5	I found the various functions in this system were well integrated.	Positive	3	3.4	2	2.4
6	I thought there was too much inconsistency in this system.	Negative	2.2	3.3	2.8	1.7
7	I would imagine that most people would learn to use this system very quickly.	Positive	3.3	3.1	2.3	2.1
8	I found the system very cumbersome to use.	Negative	2.7	2.9	2.3	2.1
9	I felt very confident using the system.	Positive	3.3	3.6	2.3	2.6
10	I needed to learn a lot of things before I could get going with this system.	Negative	1.7	2.5	3.3	2.5
(For odd-numbered questions, subtract 1 from the score. For even-numbered questions, subtract the score from 5.)			Total score:		64.75	52.25

the locomotion test, 44% prefer to use the WMCS for the manipulation test, and 28% prefer to use the WMCS for the EOD task.

In detail, two questions on mental and physical demand were selected from the NASA-TLX to assess users' workload in each mission. Table V shows the result of this index. It indicates that the WMCS had a lower mental workload and a noticeable average advantage in all the operations.

From the SUS scoring in Table VI, most users thought the WMCS was more complex than the gamepad. However, they exhibited more confidence in using the wearable system.

VII. DISCUSSION

From Table III, the most representative and significant results appear in most missions. This indicates greater differences in the performance of the two HRTIs in locomotion and manipulation motions. In the real world, most crisis management missions involve locomotion and manipulation motion in a single task [2]. Therefore, having a comprehensive evaluation system is essential for selecting suitable HRTIs for such missions. Moreover, an HRTI that benefits from the intuitiveness of WMCSs, and the accessibility of gamepads can potentially have an advantage during the operation.

The gamepad with joysticks provides only linear commands in 2D, while the WMCS offers position input in 3D. Consequently, it was easier for the joystick on the gamepad in the locomotion motion. Conversely, it was more natural for HAs to map their arm motion directly to the manipulator in 3D space. In practice, it was observed that users made more mistakes when controlling the manipulator with the joystick.

The extended Fitts' Law demonstrates that motion time increases as task difficulty increases, which aligns with the original Fitts' Law [23] and its modifications [27], [29], [30]. However, both HRTIs performed better than the predicted result, as shown in Table IV. One explanation is that users with gamepad experience also had gaming experience, and they organised their motions more efficiently than predicted. For example, the fastest user saved time by pushing out the wire connector in the EOD task instead of pulling the wire as recommended. There is another explanation for the linear line's slope differences in Fig. 8(c). The users experimented with both interfaces in the same order, from the lowest to the greatest difficulty. Although no user had experience with the WMCS, they still gained experience as they practised during the experiments. More mistakes were present in their WMCS

operation in earlier than later missions. This suggests that the WMCS is harder to operate at first contact, even for low-difficulty missions, but users can gain proficiency more quickly with practice. In interviews, users also indicated difficulty coordinating trigger strategies with argument strategies when using the WMCS at first contact, supporting this theory.

The usability results indicate that the gamepad has a higher mental workload but a lower physical one. In practice, users occasionally struggled to remember the function of each button and joystick on the gamepad, increasing their mental demand. However, operating the WMCS was more straightforward but required full-body motion rather than just finger movement. The SUS suggests that the gamepad had higher usability due to its simplicity in terms of total scoring. One reason was that some users were already familiar with gamepads. Additionally, the WMCS requires battery charging, system setup, and calibration before usage, making it less simple and more complicated to maintain.

Direct message feedback was also received from users. One user stated that "the gamepad was more sensitive and user-friendly." Another thought: "The motion-capture suit had more straightforward controls." One user pointed out the low accuracy of the WMCS, which was resolved with recalibration of the IMU sensors and did not affect the experience result. Furthermore, in more in-depth interviews, users suggested that the most optimal setup would combine both systems: "In an ideal world, I'd have a hybrid system with a joystick for locomotion and hand controls for the arm." A new design with a gamepad for locomotion control and a WMCS for manipulation control could optimise both systems.

The comparison of the two interfaces reveals a discrepancy between performance and usability. This phenomenon suggests a possible separation of performance from usability. Therefore, evaluating both aspects is essential for a comprehensive understanding of the system.

VIII. CONCLUSION

Mobile manipulators are valuable due to the combination of locomotion and manipulation functions. The development of quadrupedal manipulators enables various applications in different fields. Teleoperation will remain the dominant approach for missions in open environments in the near future. Although various teleoperation methods have been developed, a standard evaluation method needs to be included to compare their performance and usability. This work is the first

systematic attempt to fill this gap with a standardised HRTI evaluation scheme for mobile manipulation. This evaluation scheme comprises a set of standard motion tests, standardised objective measures, and subjective measures. This work also extends the ID in Fitts' Law in objective measures to make it more suitable for real-world applications with complex control methods. The scheme was practised and analysed through an experiment on a quadrupedal manipulator with two different HRTIs, revealing the differences between the two interfaces.

Although this work extends Fitts' Law by considering the position and orientation of both RA and target, the model still has limitations in representing difficulty in 3D space, particularly in orientation difficulty. In future work, this model will be expanded into real 3D space. Additionally, future work will consider different terrains of the operation field and the 3D position and tolerance angle of the manipulating target.

The experiment provided an example of applying the presented HRTI evaluation scheme to real RA and HRTIs. From the results of the experiment with the gamepad and WMCS interfaces, the proposed model can predict system performance in future missions. However, the HAs in the experiment had limited experience in robot teleoperation compared to professional operators in actual missions. Therefore, the results from the experiment only represent the user group with limited robot teleoperation experience, and a professional user group may produce different results. Also, the system uses linear polynomial lines in the prediction model, and future research can explore non-linear approaches to model the relation.

REFERENCES

- [1] C. Yang *et al.*, "A review of human-machine cooperation in the robotics domain," *IEEE Trans Hum Mach Syst*, vol. 52, no. 1, pp. 12–25, 2022.
- [2] J. C. Jurmain *et al.*, "HazBot: Development of a telemanipulator robot with haptics for emergency response," *American Journal of Disaster Medicine*, vol. 3, no. 2, pp. 87–97, 2008.
- [3] J. Carpenter, *Culture and human-robot interaction in militarized spaces: A war story*. Routledge, 2016.
- [4] S. L. Murray *et al.*, "The effects of chemical protective suits on human performance," *Journal of Loss Prevention in the Process Industries*, vol. 24, no. 6, pp. 774–779, 2011.
- [5] R. Chipalkatty *et al.*, "Human-in-the-loop: MPC for shared control of a quadruped rescue robot," in *IEEE/RSJ IROS*, 2011, pp. 4556–4561.
- [6] E. Triantafyllidis *et al.*, "Considerations and challenges of measuring operator performance in telepresence and teleoperation entailing mixed reality technologies," in *ACM CHI*, 2021, pp. 1–5.
- [7] V. R. Garate *et al.*, "A scalable framework for multi-robot tele-impedance control," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 2052–2066, 2021.
- [8] C. Stanton *et al.*, "Teleoperation of a humanoid robot using full-body motion capture, example movements, and machine learning," in *Australasian Conf. on Robotics and Automation*, vol. 8, 2012, p. 51.
- [9] T. L. Chen *et al.*, "Lead me by the hand: Evaluation of a direct physical interface for nursing assistant robots," in *ACM/IEEE HRI*, 2010, pp. 367–374.
- [10] T. I. Chowdhury *et al.*, "VR disability simulation reduces implicit bias towards persons with disabilities," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 6, pp. 3079–3090, 2019.
- [11] J. Zhao *et al.*, "Comparing hand gestures and a gamepad interface for locomotion in virtual environments," *International Journal of Human-Computer Studies*, vol. 166, p. 102868, 2022.
- [12] M. Oshita *et al.*, "Gamepad vs. touchscreen: a comparison of action selection interfaces in computer games," in *Special Interest Group on Computer Graphics and Interactive Techniques Asia*, 2012, pp. 27–31.
- [13] S. Gaglio *et al.*, "Human activity recognition process using 3-d posture data," *IEEE Trans Hum Mach Syst*, vol. 45, no. 5, pp. 586–597, 2015.
- [14] W. Takano, "Annotation generation from IMU-based human whole-body motions in daily life behavior," *IEEE Trans Hum Mach Syst*, vol. 50, no. 1, pp. 13–21, 2020.
- [15] M. Field *et al.*, "Human motion capture sensors and analysis in robotics," *Industrial Robot*, vol. 38, no. 2, pp. 163–171, 2011.
- [16] F. Schlagenhaut *et al.*, "Comparison of single-kinect and dual-kinect motion capture of upper-body joint tracking," in *Asian Control Conference*, 2017, pp. 256–261.
- [17] H. Zhou *et al.*, "IoT-enabled dual-arm motion capture and mapping for telerobotics in home care," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 6, pp. 1541–1549, 2019.
- [18] J. Humphreys *et al.*, "High utility teleoperation framework for legged manipulators through leveraging whole-body control," *Journal of Intelligent & Robotic Systems*, 2023.
- [19] L. Penco *et al.*, "Robust real-time whole-body motion retargeting from human to humanoid," in *IEEE-RAS Humanoids*, 2018, pp. 425–432.
- [20] N. Mavridis *et al.*, "A novel evaluation framework for teleoperation and a case study on natural human-arm-imitation through motion capture," *International Journal of Social Robotics*, vol. 4, no. 1, pp. 5–18, 2012.
- [21] E. Rosen *et al.*, "Communicating and controlling robot arm motion intent through mixed-reality head-mounted displays," *The International Journal of Robotics Research*, vol. 38, no. 12–13, pp. 1513–1526, 2019.
- [22] J. A. Marvel *et al.*, "Test methods and metrics for effective HRI in collaborative human-robot teams," in *ACM/IEEE HRI*, 2019, pp. 696–697.
- [23] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," *Journal of experimental psychology*, vol. 47, no. 6, p. 381, 1954.
- [24] T. O. Kvävseth, "An alternative to Fitts' law," *Bulletin of the psychonomic Society*, vol. 16, no. 5, pp. 371–373, 1980.
- [25] S. Sanei *et al.*, *EEG signal processing*. John Wiley & Sons, 2013.
- [26] S. A. Israel *et al.*, "ECG to identify individuals," *Pattern recognition*, vol. 38, no. 1, pp. 133–142, 2005.
- [27] I. S. MacKenzie, "Fitts' law as a research and design tool in human-computer interaction," *Human-Computer Interaction*, vol. 7, pp. 91–139, 1992.
- [28] E. Triantafyllidis *et al.*, "The challenges in modeling human performance in 3D space with Fitts' law," in *ACM CHI*, 2021, pp. 1–9.
- [29] M. F. Stoelen *et al.*, "Assessment of Fitts' law for quantifying combined rotational and translational movements," *Human Factors*, vol. 52, no. 1, pp. 63–77, 2010.
- [30] Y. Cha *et al.*, "Extended Fitts' law for 3D pointing tasks using 3D target arrangements," *Int. J. Ind. Ergon.*, vol. 43, no. 4, pp. 350–355, 2013.
- [31] J. Zhang *et al.*, "Recognition of mental workload levels under complex human-machine collaboration by using physiological features and adaptive support vector machines," *IEEE Trans Hum Mach Syst*, vol. 45, no. 2, pp. 200–214, 2014.
- [32] K. Stowers *et al.*, "A framework to guide the assessment of human-machine systems," *Human factors*, vol. 59, no. 2, pp. 172–188, 2017.
- [33] S. Hart *et al.*, "Development of NASA-TLX (task load index)," in *Advances in psychology*, 1988, vol. 52, pp. 139–183.
- [34] R. A. Grier, "How high is high? A meta-analysis of NASA-TLX global workload scores," in *The Human Factors and Ergonomics Society Annual Meeting*, vol. 59, no. 1, 2015, pp. 1727–1731.
- [35] M. R. Endsley *et al.*, "A comparative analysis of SAGAT and SART for evaluations of situation awareness," in *Proc Hum Factors Ergon Soc Annu Meet*, vol. 42, no. 1, 1998, pp. 82–86.
- [36] J. Brooke, *SUS: A 'quick and dirty' usability scale*. CRC Press, 1996, ch. 21.
- [37] A. Huber *et al.*, "Developing human-robot interaction for an industry 4.0 robot: How industry workers helped to improve remote-HRI to physical-HRI," in *ACM/IEEE HRI*, 2017, pp. 137–138.
- [38] R. Bevilacqua *et al.*, "Robot-Era project: Preliminary results on the system usability," in *International conference of design, user experience, and usability*, 2015, pp. 553–561.
- [39] C. Zhou *et al.*, "TeLeMan: Teleoperation for legged robot loco-manipulation using wearable IMU-based motion capture," 2022, arXiv:2209.10314.
- [40] C. Peers *et al.*, "Development of a teleoperative quadrupedal manipulator," in *The 4th UK-RAS Conference*, 2021, pp. 17–18.
- [41] J. Li *et al.*, "Whole-body control for a torque-controlled legged mobile manipulator," *Actuators*, vol. 11, no. 11, 2022.
- [42] C. Pinheiro *et al.*, "Wearable biofeedback improves human-robot compliance during ankle-foot exoskeleton-assisted gait training: a pre-post controlled study in healthy participants," *Sensors*, vol. 20, no. 20, p. 5876, 2020.