

Human Modeling and Passivity Analysis for Semi-Autonomous Multi-Robot Navigation in Three Dimensions

TAKESHI HATANAKA ¹ (Senior Member, IEEE), TAKAHIRO MOCHIZUKI ¹, TAKUMI SUMINO ¹,
JOSÉ M. MAESTRE ² (Senior Member, IEEE), AND NIKHIL CHOPRA ³ (Member, IEEE)

(Modeling, Control, and Learning Approaches for Human-Robot Interaction Systems)

¹School of Engineering, Tokyo Institute of Technology, Tokyo 152-8550, Japan

²Systems and Automation Engineering Department, University of Seville, 41092 Sevilla, Spain

³Department of Mechanical Engineering, University of Maryland, College Park, MD 20742 USA

CORRESPONDING AUTHOR: TAKESHI HATANAKA (e-mail: hatanaka@sc.e.titech.ac.jp).

This work was supported in part by JSPS KAKENHI under Grant 21K04104, and in part by Spanish MCIN/AEI/10.13039/501100011033 C3PO-R2D2 Project under Grant PID2020-119476RB-I00.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Human Subjects Research Ethics Review Committee in Tokyo Institute of Technology under Application No. 2023109, and performed in line with the Helsinki on Ethical Principles for Medical Research and Ethical Guidelines for Medical and Biological Research Involving Human Subjects.

This article has supplementary downloadable material available at <https://doi.org/10.1109/OJCSYS.2023.3343598>, provided by the authors.

ABSTRACT In this article, we study a one-human-multiple-robot interaction for human-enabled multi-robot navigation in three dimensions. We employ two fully distributed control architectures designed based on human passivity and human passivity shortage. The first half of this article focuses on human modeling and analysis for the passivity-based control architecture through human operation data on a 3-D human-in-the-loop simulator. Specifically, we compare virtual reality (VR) interfaces with a traditional interface, and examine the impacts that VR technology has on human properties in terms of model accuracy, performance, passivity and workload, demonstrating that VR interfaces have a positive effect on all aspects. In contrast to 1-D operation, we confirm that operators hardly attain passivity regardless of the network structure, even with the VR interfaces. We thus take the passivity-shortage-based control architecture and analyze the degree of passivity shortage. We then observe through user studies that operators tend to meet the degree of shortage needed to prove closed-loop stability.

INDEX TERMS Distributed control, human-robot collaborations, passivity, semi-autonomous multi-robot navigation, system identification, virtual reality.

I. INTRODUCTION

Most real-world robotic tasks, more or less, need human intervention due to the high capability that humans possess in reasoning and decision-making especially in unstructured/uncertain environments. A robotic system in which humans are involved in part of the overall process is called a semi-autonomous system. Needless to say, human modeling is indispensable when systematically designing a semi-autonomous robotic system. The most appropriate human model strongly depends on the role that the human plays in the system. Musić and Hirche [1] classified human roles

in human-robot interactions as either active or supervisory, depending on the required level of autonomy [2]. In the active role, human intervention may involve motion control of the robot, and accordingly, the human is involved in the control loop. Meanwhile, in the supervisory role, the human focuses only on high-level decisions based on abstracted task information, which is sometimes described as “human-on-the-loop” [3]. In this article, we focus on modeling of a human with an active role in the human-in-the-loop architecture.

A promising approach to human-in-the-loop robot control with the human having an active role has been studied

in the paradigm of bilateral teleoperation [4], [5], [6]. In this paradigm, the operator is modeled as a passive system. Beyond traditional one-human-one-robot teleoperation, teaming between a human and multiple robots has been investigated in the literature, including the scenarios of cooperative payload manipulation [7], [8], [9], [10], multi-robot navigation [11], [12], [13], and an exploration task [14], [15]. Meanwhile, relatively few papers have questioned the validity of the passivity assumption imposed on the human operator. Dyck et al. [16] examined the passivity of the human arm and revealed that it had a task-dependency. Bilateral teleoperators under milder assumptions were presented in [17], [18] to address the possibility of non-passive human operators. Non-passive components in bilateral teleoperators have also been studied based on integral quadratic constraint [19], [20], [21] and the input-to-output stability small gain theorem [22].

Previously, we addressed multi-robot navigation based on the concept of passivity, and presented a fully distributed control architecture that ensures motion synchronization under the assumption of human passivity [23]. We also examined human passivity through system identification techniques using human operation data collected on a 1-D human-in-the-loop simulator. We observed that an appropriately defined notion of human passivity strongly depends on the proficiency level of the system operation and the network structure. To address the possibility of the passivity shortage, the authors of [24], [25] also presented another distributed control architecture based on the architecture for output synchronization of passivity-short systems [26] and bilateral teleoperation [4], [5], [6]. In the present paper, we study an extension of [23], [24], [25] to the three dimensional case. It is easy to confirm that the stability analysis in [23], [24], [25] can be directly applied to the 3-D case. However, it remains unclear whether the analytical results for the 1-D experiments are applicable to the 3-D case. The selection of the interface in particular is more critical than in the 1-/2-D operations due to the limited human capability for 3-D recognition and real-time manipulability.

Virtual Reality (VR) technology is widely believed to enhance human 3-D recognition and manipulability. Indeed, there have been many publications devoted to various human-robot interactions, including operator support [27], [28], [29], simulation [30], [31], instruction [32], [33], [34], manipulation [35], [36], and teleoperation [37], [38] with VR and Augmented Reality (AR) devices as summarized in [39]. Some of these studies quantitatively revealed the benefit of VR and AR technology in terms of safety [27], [30], task completion speed [33], [34], [35], [37] and accuracy [35], [37], and human usability, workload, and experience [36], [37], [38]. The authors of [34] also conducted comprehensive subjective evaluations through user studies. Missing from these works is an analysis from a control-theoretic perspective. For example, how VR interfaces affect the human dynamic behavior remains an open question. Moreover, to the best of our knowledge, the impact of VR interfaces on human passivity has not yet been reported in the literature.

In this article, we study the scenario involving 3-D multi-robot navigation shown in Fig. 1 based on the paradigm

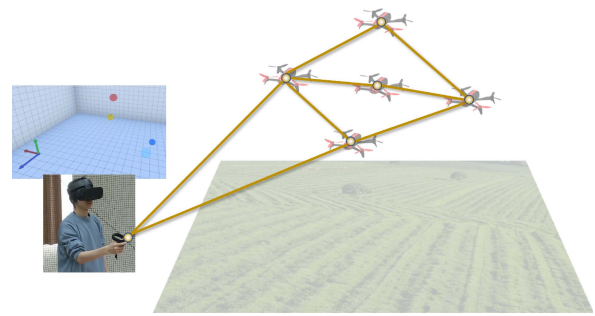


FIGURE 1. Scenario involving 3-D one-human-multiple-robot interactions.

of [23], [24], [25], and address human modeling and human passivity analysis. To this end, we begin by conducting user studies for the passivity-based architecture in [23] on a 3-D human-in-the-loop simulator with three different networks, taking three pairs of command/feedback interfaces: a traditional joystick controller and a 2-D display, a VR controller and a 2-D display, and a VR controller and a head mounted display (HMD). We then build dynamical models of a human operator and analyze human passivity, showing that VR interfaces are advantageous both in human modeling and human passivity as well as tracking performance. We also conduct the same experiments and modeling for other nine trial subjects and examine if the results of the investigation can be generalized. We also conduct the NASA TLX questionnaire in order to quantitatively evaluate the human workload for the three interface pairs. In the latter part of this article, we focus on the problem of human deviation from passivity even with VR interfaces. We conduct user studies for the passivity-shortage-based architecture in [24], [25] and reveal that all operators meet the degree of passivity shortage assumed in [24], [25].

The contributions of this article are summarized as follows.

- i) We assess the impact of VR interfaces on dynamic human properties, including dynamical model accuracy, tracking performance, and passivity.
- ii) Multiple user studies are conducted to understand the generalization capability of the aforementioned analysis, and to analyze the human workload for each interface.
- iii) The passivity-short assumption in [24], [25] is shown to be satisfied even in the 3-D operation.

Note that preliminary results only for the passivity-based architecture [23] were in part presented in a conference version in [40], however (ii) and (iii) are novel contributions added in this article. Also, the human models in [40] had undesirable resonance peaks due to overfitting, and the model developed in this article does not have such peaks.

II. PRELIMINARY: PASSIVITY

We consider a dynamical system with the same input-output dimension given as

$$\dot{x} = f(x) + g(x)u, \quad x(0) = x_0, \quad (1a)$$

$$y = h(x), \quad (1b)$$

where $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}^m$ is the control input, and $y(t) \in \mathbb{R}^m$ is the control output. The system is said to be passive if there exists a positive semi-definite function $S: \mathbb{R}^n \rightarrow \mathbb{R}$ such that

$$S(x(\tau)) - S(x_0) \leq \int_0^\tau y^T(t)u(t)dt$$

holds for all input signals u , all initial states x_0 , and all time $\tau \geq 0$. Suppose now that there exists $v \in \mathbb{R}$ such that

$$S(x(\tau)) - S(x_0) \leq \int_0^\tau y^T(t)u(t)dt + v \int_0^\tau \|u(t)\|^2 dt$$

holds for all u , all x_0 , and all $\tau \geq 0$. The minimum of v is called the input passivity index and is denoted by \bar{v} . The system is then said to be input feedforward passive if $\bar{v} > 0$ and input feedforward passivity-short if $\bar{v} < 0$. When $\bar{v} < 0$, the parameter \bar{v} is called the impact coefficient [26].

Suppose now that (1) is linear time-invariant. We then define

$$v(\omega) = \lambda_{\min}(G(j\omega) + G^H(j\omega)),$$

where $G(s)$ is the transfer function matrix from u to y and $\lambda_{\min}(A)$ is the minimal eigenvalue of a matrix A . It is then well known that $\bar{v} = \min_\omega v(\omega)$ holds [26]. The function $v(\omega)$ is thus regarded as a passivity metric corresponding to the angular frequency ω .

III. CONTROL ARCHITECTURES FOR SEMI-AUTONOMOUS MULTI-ROBOT NAVIGATION

The goal of the 3-D one-human–multiple-robot interaction in this article is for the operator to stably navigate the multiple robots to a desirable position or velocity. In this scenario, direct manual control of the robots in real time is demanding for the human operator, and may even be impossible depending on the number of robots. A promising approach to address the issue is to utilize the architecture for the so-called complementary interactions [1], where motion synchronization is left to the multi-robot system and the role of controlling the robotic group is assigned to the operator. To this end, we presented two fully distributed control architectures based on passivity [23] and passivity-shortage [25], respectively. We will now briefly review these architectures.

A. PASSIVITY-BASED CONTROL ARCHITECTURE

Let us consider a group of n robots in 3-D Euclidean space as shown in Fig. 2. The set of their IDs is denoted by $\mathcal{V} = \{1, 2, \dots, n\}$. Each robot $i \in \mathcal{V}$ is assumed to obey the kinematic model

$$\dot{p}_i = u_i, \quad (2)$$

where $p_i \in \mathbb{R}^3$ is the position of robot $i \in \mathcal{V}$ relative to the world frame and $u_i \in \mathbb{R}^3$ is the velocity input to be designed. The robots are assumed to be interconnected by a network that is modelled by a fixed and connected undirected graph $G = (\mathcal{V}, \mathcal{E})$, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. The neighbor set \mathcal{N}_i is then defined as $\mathcal{N}_i = \{j \in \mathcal{V} \mid (i, j) \in \mathcal{E}\}$.

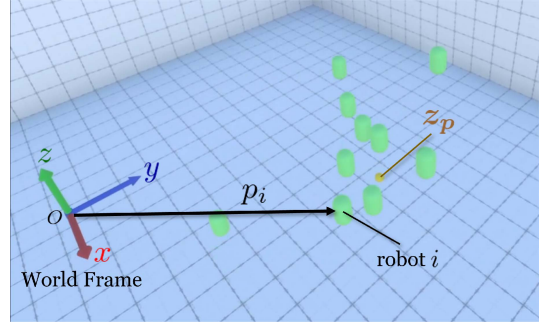


FIGURE 2. Configuration of multiple robots.

A human operator is interfaced with the robots to receive feedback and command in his or her interaction with the robots. The feedback interface translates the output of the robotic group into visual feedback for the operator, and the command interface translates the human action into a control command for the robotic group. Both interfaces are assumed to have access to a subset of robots $\mathcal{V}_h \subseteq \mathcal{V}$ through wireless communication. The information displayed on the feedback interface is denoted by y_h , and the command determined by the operator is denoted by u_h , where the vector u_h is defined in the world frame. Throughout this article, we assume that $u_h \in \mathbb{R}^3$ is a velocity command to the robots $i \in \mathcal{V}_h$. The operator drives all robots to a reference position r_p or reference velocity r_v in the world frame by manipulating the command signal.

The operator chooses whether to control the positions or velocities of the robots. For the position navigation, the control goal is formulated as

$$\lim_{t \rightarrow \infty} \|p_i - r_p\| = 0 \quad \forall i \in \mathcal{V}, \quad (3)$$

while for the velocity navigation it is formulated as

$$\lim_{t \rightarrow \infty} \|\dot{p}_i - r_v\| = 0 \quad \forall i \in \mathcal{V}, \quad (4a)$$

$$\lim_{t \rightarrow \infty} \|p_i - p_j\| = 0 \quad \forall i, j \in \mathcal{V}. \quad (4b)$$

To achieve both position and velocity navigation, the authors of [23], [25] proposed the following distributed controller based on the PI consensus algorithm.

$$u_i = \sum_{j \in \mathcal{N}_i} a_{ij}(p_j - p_i) + \sum_{j \in \mathcal{N}_i} b_{ij}(\xi_i - \xi_j) + \delta_i v_h \quad (5a)$$

$$\dot{\xi}_i = \sum_{j \in \mathcal{N}_i} b_{ij}(p_j - p_i), \quad (5b)$$

where a_{ij} and b_{ij} are positive gains, and $\delta_i = 1$ if $i \in \mathcal{V}_h$ and $\delta_i = 0$ otherwise. The symbol v_h is a signal that will be designed to reflect the human navigation objective. The signal differs between the passivity-based and the passivity-shortage-based control architecture, which will be presented later.

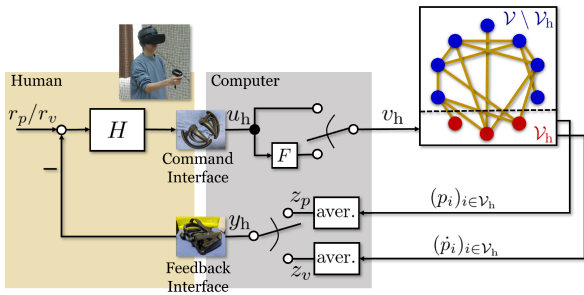


FIGURE 3. Passivity-based control architecture, where the switch is activated according to which control goal is selected by the operator, (3) or (4). The block H is a map from the tracking error e to the human velocity command u_h . The block F is a filter that ensures differentiability of the signal v_h , which is needed only for the velocity navigation. The “aver.” blocks output the average of the elements of the input into them.

Let us now define the average position and velocity of the accessible robots \mathcal{V}_h as

$$z_p = \frac{1}{|\mathcal{V}_h|} \sum_{i \in \mathcal{V}_h} p_i, \quad z_v = \frac{1}{|\mathcal{V}_h|} \sum_{i \in \mathcal{V}_h} \dot{p}_i, \quad (6)$$

respectively. The following lemma can then be proved to hold in [23].

Lemma 1: [23] Consider a group of robots with the dynamics (2) and the distributed controller (5), where the undirected graph G is assumed to be connected and $|\mathcal{V}_h| \geq 1$. Then, the collective dynamics for all robots is passive from v_h to z_p . Also, if the signal v_h is differentiable in time, then the dynamics is passive from \dot{v}_h to z_v .

The authors of [23] designed a control architecture assuming that the human behaves as a passive system. Given this assumption, Lemma 1, and the fact that feedback interconnection of two passive systems ensures closed-loop stability [6], they interconnected the human and robots with $v_h = u_h$ and had the average position $y_h = z_p$ or velocity $y_h = z_v$ fed back to the operator, where these signals are switched at the feedback interface depending on the selected control goal, (3) or (4).¹ The overall system is illustrated in Fig. 3 for the situation in which the operator determines the velocity command based on the error $e = r_p - z_p$ or $e = r_v - z_v$.

They showed that both the goals (3) or (4) with constant references r_p and r_v is achieved under human passivity together with additional assumptions, even without sharing the selected control goal among all the robots. Despite the difference in the configuration space of the considered system, the same proof can be trivially extended to the three dimensional case considered in this article and therefore is omitted. We instead focus on human behavioral analysis, including human passivity. The authors of [23] addressed this issue by studying human modeling and human passivity analysis based on operation data acquired on a 1-D human-in-the-loop simulator. However, it is unclear whether a human undertaking the 3-D robot operation would behave in the same way as in the 1-D

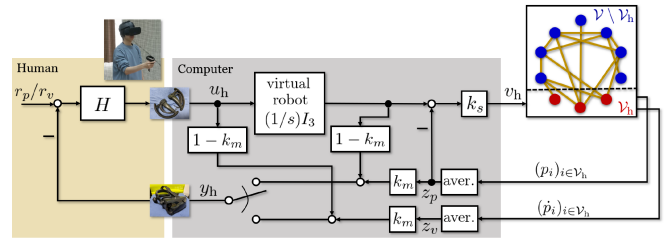


FIGURE 4. Passivity-shortage-based control architecture.

or the 2-D case. Since the tablet interface considered in [23] is unsuitable as a command interface for 3D operations due to its limited dimensionality, we must carefully select an alternative interface. Furthermore, the challenge of conveying the robot’s 3-D information to humans is an issue that demands thorough consideration.

B. PASSIVITY-SHORTAGE-BASED CONTROL ARCHITECTURE

The human passivity analysis in [23] revealed that human passivity depends on the proficiency level of the system operation and the network structure. It was also shown that the operator may fail to attain passivity for a sparse network even after training. Motivated by this work, the authors of [24], [25] presented another control architecture that accepts the human passivity shortage based on the architecture for output synchronization of passivity-short systems [26] and the bilateral teleoperation [4], [5], [6].

In our architecture, we employ a master robot to interact with the other robots, similarly to bilateral teleoperation [4], [5], [6]. Note, however, that our architecture differs from teleoperation in that the motion of the master is virtually simulated in the interface instead of there being a physical interaction between the master and the operator. Accordingly, the operator interacts with the robots by assessing visual feedback and determining a velocity command through an interface.

Denoting the position of the master by $q_m \in \mathbb{R}^3$ in the world frame, we let q_m obey the dynamics

$$\dot{q}_m = u_h. \quad (7)$$

We then interconnect the human operator with the master robot (7) and the robotic group using the architecture in [26]. Specifically, the signal y_h is designed as

$$y_h = k_m z_p + (1 - k_m) q_m, \quad (8)$$

or

$$y_h = k_m z_v + (1 - k_m) \dot{q}_m, \quad (9)$$

with $k_m \in (0, 1)$ by blending the master position and the average position. We also design v_h as

$$v_h = k_s (q_m - z_p), \quad (10)$$

where k_s is a positive gain. The overall control architecture is illustrated in Fig. 4.

¹In the velocity navigation only, v_h is set to a signal after filtering u_h to ensure the differentiability of u_h assumed in Lemma 1.

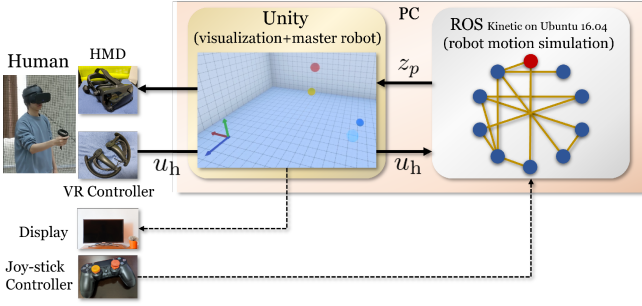


FIGURE 5. Schematic of the 3-D human-in-the-loop simulator.

Suppose now that the cascade system of the master robot and human is passivity short. Precisely speaking, we assume that the user-defined reference r_p is constant and that the cascade system is passivity short from $r_p - y_h(t)$ to $q_m(t) - r_p$ with an impact coefficient \bar{v} greater than -1 . The authors of [24] proved that (3) is achieved under $k_m \in (0, 1)$ and additional assumptions on the boundedness of the signals for position navigation. Meanwhile, in the case of velocity navigation, (4) was shown to be achieved if the cascade system is passivity short from $r_v - y_h(t)$ to $\dot{q}_m(t) - r_v$ with $\bar{v} > -1$, $k_m \in (0, 1)$, and with additional assumptions on the boundedness of the signals. They also showed through user studies that operators met the assumption of $\bar{v} > -1$. Just as in Section III-A, we forgo repeating the same proof in this article and instead focus on whether $\bar{v} > -1$ is satisfied for the 3-D operation.

IV. EXPERIMENT DESIGN

In the subsequent sections, we will address human modeling and analyze human passivity based on the models for the control architectures in Section III. In this section, we present a human-in-the-loop simulator and identification experiments conducted to collect the operation data for the human modeling. Note that we focus on position navigation for the remainder of this article, and will address velocity navigation in our future.

A. HUMAN-IN-THE-LOOP SIMULATOR

In this subsection, we present a 3-D human-in-the-loop simulator whose overview is shown in Fig. 5.

We simulate the motion of 10 robots with the dynamics (2) and the distributed controller (5). The virtual master with (7) is only used in the case of the passivity-shortage-based architecture. In view of the real implementation, the robot dynamics (2) and (5) are simulated on the robot operating system (ROS), while the master dynamics and (10) are implemented in Unity. We also take three different types of networks, shown in Fig. 6, for which we set $a_{ij} = b_{ij} = 1$ for all $(i, j) \in \mathcal{E}$. In Type 1 (left), the inter-robot network is sparse, and all robots are connected to the interface. In Type 2 (middle), the robots are interconnected by a dense network, while only robot 1 is connected to the interface. In Type 3 (right), the inter-robot

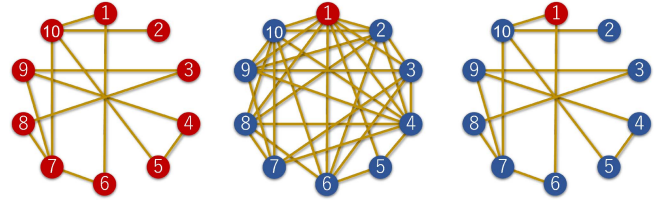


FIGURE 6. Communication network Type 1 (left), Type 2 (middle), and Type 3 (right), where the red nodes belong to \mathcal{V}_h and the blue do not.

network is the same as in Type 1, but only robot 1 is connected to the interface. The signal v_h in (5) is set to $v_h = u_h$ in the passivity-based architecture, while v_h is given by (10) with $k_s = 3$ in the passivity-shortage-based architecture. The average position z_p is sent to Unity, which then generates 3-D graphics to enable smoother interactions between the human and the robots.

We prepared two command interfaces. The first uses a DualShock 4 controller (Sony Corp.), standard joystick-based controller common for gaming that uses a pair of joysticks for input. The left stick specifies the x - and z -coordinates of the velocity command u_h , while the longitudinal operation of the right stick corresponds to the y -coordinate. The gain from the joystick angle to u_h was tuned so that the maximal angle corresponds to ± 0.15 m/s for Type 1 and ± 1.5 m/s for Types 2 and 3, which was done to enable better human operability, and due to the fact that the stationary gain from u_h to z_p for Type 1 is 10 times as large as that for the other two networks [23]. This reasonable change in the stationary gain at the interface depending on the network structure is simply determined by $|\mathcal{V}_h|/n$. The second interface used Valve Index (Valve Corp.) or Meta Quest 2 (Meta Platforms Inc.) VR controller. The operator pushes a button on the controller in the beginning of the experiments, and the controller's position at this time is set to the origin. The vector from the origin to the real-time position of the VR controller is converted to γu_h . In view of the fact that the hand motion for each coordinate is limited to around ± 30 cm, the parameter γ was set to 2 for Type 1 and 0.2 for the other networks, and consequently, $\|u_h\|_\infty$ was approximately restricted to 0.15 m/s for Type 1 and 1.5 m/s for Types 2 and 3 just as they were for the joystick controller. The command signal u_h was then directly sent to a topic through Unity, at which point ROS subscribed and substituted the signal into u_h in (5) or (10).

We used two feedback interfaces, a standard 2-D 27-inch display monitor and the Valve Index (Valve Corp.) or Meta Quest 2 (Meta Platforms Inc.) VR HMD. The HMD is connected to Unity, receiving and displaying the 3-D graphics generated by the program. The viewing angle varies depending on the behavior of the person on whom the display is mounted. The feedback information, $y_h = z_p$ and (8), is switched depending on the selected control architecture. The point y_h represented in the 3-D graphics as a yellow ball, as shown in Fig. 7. When the operator is using the VR controller, we also display the origin of the controller coordinate frame

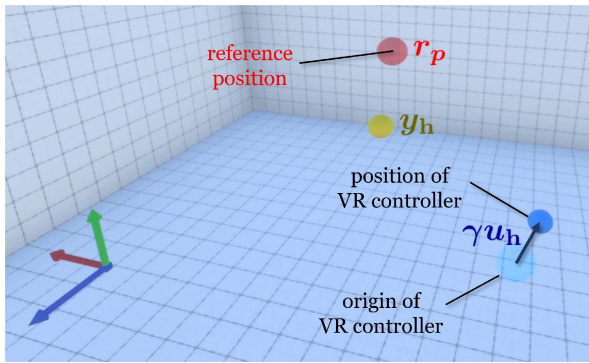


FIGURE 7. Scene viewed during system operation.

TABLE 1. Interface selections in the identification experiments.

	Feedback Interface	Command Interface
Interface #1	2-D display	Joystick
Interface #2	2-D display	VR controller
Interface #3	HMD	VR controller

and the current position of the controller using cyan and blue balls, respectively. For the 2-D monitor, the 3D graphics are seen from a fixed viewpoint, where the optical axis is in parallel to z -axis of the world frame and the viewing angle is set to 60 degrees throughout the experiments. In this setting, the robots may leave the field of view, but we arranged the experiments so that this would not happen since addressing this issue is beyond the scope of this work.

B. IDENTIFICATION EXPERIMENT

We will now detail the identification experiments we designed on the previously described simulator. The trial subject was told to use either the joysticks or the VR controller to drive z_p (yellow ball) to a reference r_p (red ball). They were told to stop when the yellow ball with a diameter of 5 cm lay inside of the red ball with a diameter of 5.5 cm. The reference jumped randomly at every 15 s to a point within a 2 m cube, including the operator whose center was located 1 m from the floor. One trial consisted of 10 jumps of the reference, taking 150 s in total. A video of the trials is found in the Multimedia Materials for this article.

The subject conducted trials for all the networks in Fig. 6 under the three different interface settings summarized in Table 1. The sampling period for the data was 0.0083 s on average. It should also be noted that we included an additional process to the acquired data in the same manner as in [23]. Specifically, the subject was told to press a button to indicate that he or she recognizes the new references and was ready to start the operation. The data were then shifted so that the initial time of the operation synchronized with the time that the button was pushed, thus excluding the delays associated with recognizing the new references. The recognition delay is a phenomenon unique to this experiment since the reference is determined by the human in practice.

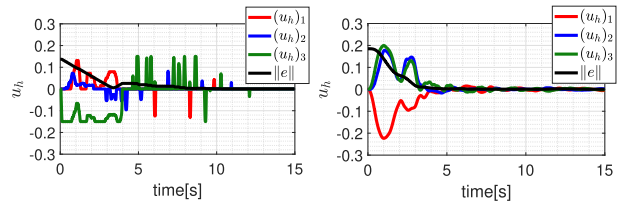


FIGURE 8. Time responses of u_h for the joystick (left) and VR controller (right), where the black curve shows $\|e\|$ with the error $e = r_p - z_p$.

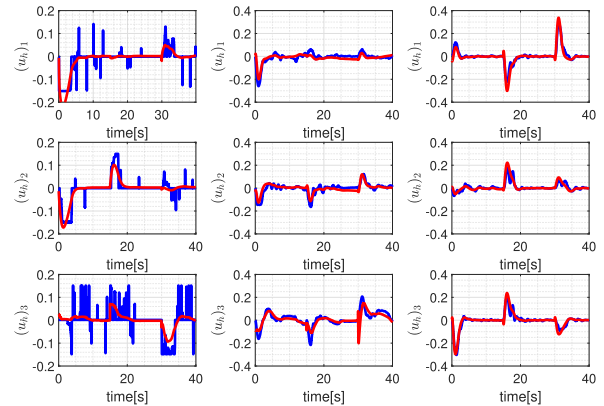


FIGURE 9. Time series data of the model outputs (red) and the identification data (blue) on u_h for the Type 1 network (left: interface #1, middle: interface #2, right: interface #3).

The time responses of u_h over 15 s for the joystick and VR controllers are illustrated in the left and right figures, respectively, of Fig. 8. The operator tended to take extreme actions for the joystick controller even in the settling phase, with a small error, because the built-in spring makes fine manipulation difficult. Meanwhile, we can see that the commands given with the VR controller are much smoother.

V. IMPACT OF VR TECHNOLOGY ON HUMAN PROPERTIES

In this section, we focus on the passivity-based architecture in Section III-A. Our goal here is to examine how the interfaces in Table 1 affect human properties including model accuracy, tracking performance, and passivity.

A. HUMAN MODELING

Let us first assess how each interface affects human modeling. To this end, we develop a human operator model using MATLAB System Identification Toolbox (Mathworks Inc.) and the so-called direct approach to closed-loop system identification [41], using the data from one trial. As a result, we obtain a model $u_h(s) = H(s)e(s)$ with a 3-by-3 transfer function matrix $H(s)$. In the system identification, we used a continuous-time model with 2 poles and 1 zero for the diagonal elements and constants for the non-diagonal elements. A discussion of the selection of the model order can be found in the Appendix. In the sequel, we use $H(s)$ identified from various operation data to analyze human properties.

The time responses of the model outputs and the identification data for the Type 1 network are shown in Fig. 9. We can

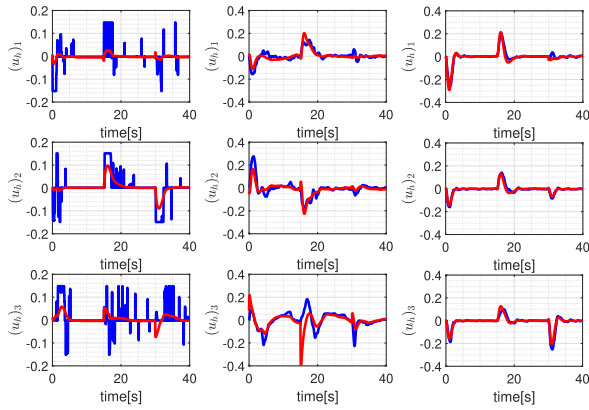


FIGURE 10. Time series data of the model outputs (red) and the verification data (blue) on u_h of the same subject as Fig. 9 for the Type 1 network (left: interface #1, middle: interface #2, right: interface #3).

TABLE 2. Average model fit ratios among three elements of u_h for the verification data (passivity-based architecture).

	Type 1	Type 2	Type 3
Interface #1	31.83%	25.73%	28.06%
Interface #2	39.37%	28.55%	35.93%
Interface #3	71.03%	54.06%	67.27%

see that the fitting performance for the traditional interface (interface #1 in Table 1) is poor, primarily because of the extreme actions pointed out in Fig. 8. Meanwhile, interfaces #2 and #3 with the VR controller achieve a better fitting performance than interface #1. The accuracy for interface #3 looks slightly better than that for interface #2, but it is difficult to discuss the superiority between these two only from this data.

Next, we conducted the cross validation by using the data from another trial as verification data. Fig. 10 shows the time responses of the model outputs and the verification data for the Type 1 network. The fitting performance for interface #1 is again worse than the other two. It is interesting to note that the fitting performance for the 3rd element of u_h is worse than the other two elements in the case of interface #2, while the model successfully fit the data for all elements in the case of interface #3. In view of the fact that the coordinate of the 3rd element corresponds to the depth from the viewpoint in the 3-D graphics, the bottom-middle figure indicates that the operator may fail to take consistent behavior against the depth errors. It is conceivable from the bottom-middle and bottom-right figures that the dominant reason why the model accuracy for this coordinate is worse than the other two coordinates is due to the limited human capacity for depth recognition in 2-D images. The fit ratios for all three networks and all interface selections are given in Table 2. We can see from them that the VR interfaces improved the model accuracy with 25~30% compared with the other interfaces, regardless of the network types. We see from this table that not only does the smoother operation of the VR controller contribute to enhancing the model accuracy but also does the enhanced 3-D recognition by the HMD.

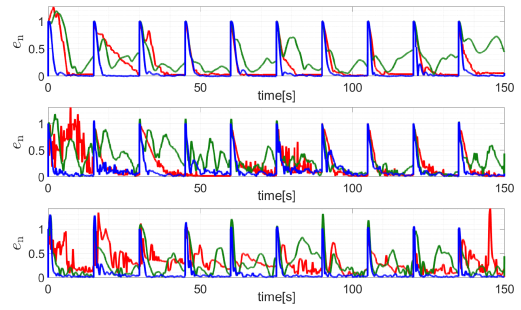


FIGURE 11. Time series data of the normalized error e_n over a trial for interface #1 (red), #2 (green), and #3 (blue) and three networks, where the top, middle and bottom figures correspond to Type 1, 2, and 3, respectively.

In summary, we can conclude that the VR interface simplifies human behavior in these tasks to the extent that it can be represented by a linear time-invariant system. This reduces the uncertainty stemming from the human factor in the loop and simplifies designing human-robot interaction systems. In the present paper, we focus only on the specific control architectures and robot dynamics. However, the above analysis of the human does not rely on the special structure of the present robot system. It is thus expected that these investigations are also valid for other semi-autonomous robot navigation systems as long as the operator gives the velocity commands and the robot dynamics are well compensated so that they follow the commands and are approximated to be linear time-invariant. We would like to leave more investigations on the issue to future work.

B. TRACKING PERFORMANCE

Let us next demonstrate the tracking performances for the above three interfaces.

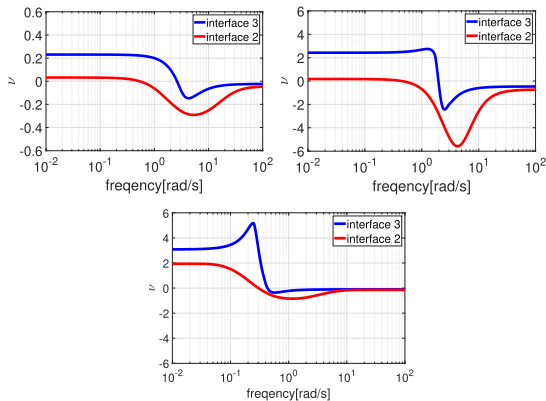
We take the tracking error e as a metric of the performance, but the absolute value of the error depends on the size of the jumps in the reference r_p . In the above trial, the reference switches at every 15 s and we denote each switching time by t_k , $k = 0, 1, 2, \dots, 10$ with $t_0 = 0$ s. Accordingly, we define the normalized error e_n by

$$e_n(t) = \frac{e(t)}{\|e(t_k)\|} \text{ if } t \in [t_k, t_{k+1})[s], \quad k = 0, 1, 2, \dots, 10.$$

The time series data of the normalized errors e_n for the three interfaces and three networks during a trial are shown in Fig. 11. Surprisingly, it is first observed that interface #2 performs worse than the other two, which indicates that the benefits of the VR interface are maximised not by the VR controller alone, but by the combination of the controller and HMD. We see from this figure that interface #3 achieves the smallest steady-state error among the three for all network types. These results indicate that the depth information provided by HMD is the key to improving the tracking performance. The accumulated values of $\|e(t)\|$ over $t \in [t_k - 5, t_k]$ ($k = 1, 2, \dots, 10$) for interface #1 and interface #3 over the two trials are summarized in Table 3. It is quantitatively

TABLE 3. Steady-state errors for interface #1 and interface #3.

1st trial			
	Type 1	Type 2	Type 3
interface #1	2.504	5.179	11.69
interface #3	0.599	2.589	1.880
2nd trial			
	Type 1	Type 2	Type 3
interface #1	3.350	5.448	16.05
interface #3	0.674	2.374	1.673


FIGURE 12. Passivity index ν for the passivity-based control architecture (top-left: Type 1, top-right: Type 2, bottom: Type 3).

confirmed that the steady-state error is reduced by the combination of the VR controller and HMD for all networks. We also see from Fig. 11 that the interface #3 achieves faster responses than #1.

In summary, we conclude that the VR interface enhances the tracking performance in terms of both steady-state and fast-response properties by providing humans with richer depth information.

C. HUMAN PASSIVITY

Let us next investigate the impact of the VR interface on human passivity. We take the model for interface #2 as a baseline for the comparison with that for interface #3, since the model for interface #1 is not always accurate enough to discuss passivity. This allows us to analyze how the higher 3-D recognition ability enabled by the HMD improves human passivity.

Fig. 12 shows the passivity index $\nu(\omega)$ for the three networks, where the blue and red curves represent $\nu(\omega)$ for the models in interface #3 and #2, respectively. We can immediately see that the index for the model with the HMD is larger than that for the model with the 2-D display over all frequencies for all networks. In view of the property of the passivity index ν stated at the end of Section II, we conclude that the HMD improves human passivity, which also means that the VR interface enhances closed-loop stability of this human-in-the-loop system.

Meanwhile, Fig. 12 indicates that the human operator fails to attain passivity for all networks even if he/she uses the VR interface. We remark that human passivity shortage does

not immediately generate closed-loop instability for a specific network. Human passivity is a condition that ensures closed-loop stability for any network. In practice, the operator was able to achieve stable operations throughout all trials. Nonetheless, it would be more reliable to design a system architecture in which the operator is able to ensure the stability for any network. This motivates us to use the passivity-shortage-based control architecture in Section III-B, which we will study in the next section.

D. VALIDATION FOR MULTIPLE SUBJECTS

In this subsection, we examine if the investigations in the above subsections are universally applied to other operators. To this end, we conduct user studies for the other nine trial subjects.

We start with remarking that the following discussions assume that subjects are pre-trained. Actually, without any training, they could not even complete the navigation task itself. In such a case, no meaningful discussion on model accuracy or other human properties can be developed. In the experiment, we thus had the subjects freely repeat the above 150 s trials with the network Type 2 three times for each interface. We next showed the operation of a well-trained operator as a reference, and then asked them to do one more trial. They finally conduct two trials for all of the three interfaces, where the data for the first trial and the second trial are used as identification data and verification data, respectively.

Let us first examine our hypothesis on human modeling in Section V-A, where it was exemplified that the VR interfaces enhance the human model accuracy. We take the same modeling method and model parameters as Section V-A. The fit ratios for the interface #1(red), #2(green), and #3(blue) are shown in Fig. 13. The fit ratios to the identification data in the left figure are higher for interface #3, #2 and #1 in this order for all participants, which validates the hypothesis formed in Section V-A. The ratios to the verification data in the right figure tend to be lower than those in the left, and, for participant 2, 7, and 8, the order for interface #1 and #2 is reversed. However, it is at least confirmed that the pair of the VR controller and HMD achieves the best fit ratio for all participants, and these results reinforce the hypothesis that VR interfaces improve the model accuracy.

We next examine our hypothesis on human passivity in Section V-C. The input passivity indices $\bar{\nu}$ for interface #1(red), #2(green), and #3(blue) are shown in Fig. 14. Although some models have questionable reliability due to their low accuracy in Fig. 13, we at least observe the tendency that interface #3 enhances the human passivity, which almost reinforces the conclusion in the previous subsection. On the other hand, all of the participants do not perfectly attain passivity even with interface #3 in the same way as Section V-C, whereas an operator attained passivity depending on the network structure in the 1-D case [23]. These results also emphasize the needs for addressing human passivity shortage.

We finally show the human workload perceived by the nine trial subjects for the three interfaces through NASA TLX

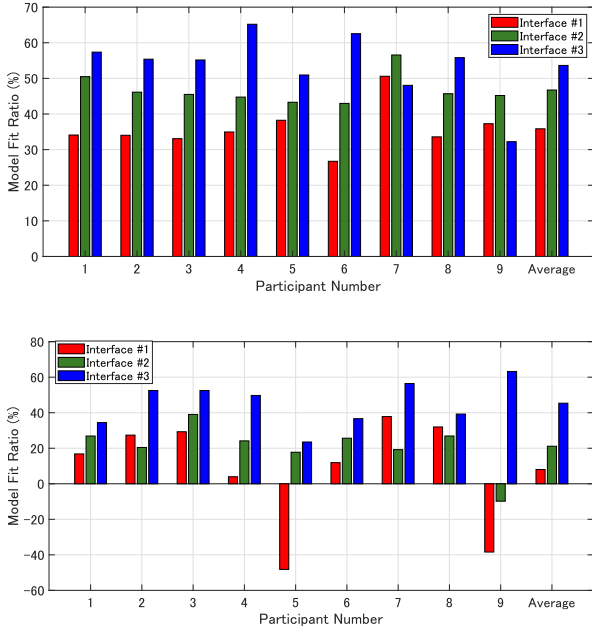


FIGURE 13. Model fit ratios for the nine trial subjects with the identification data (top) and verification data (bottom).

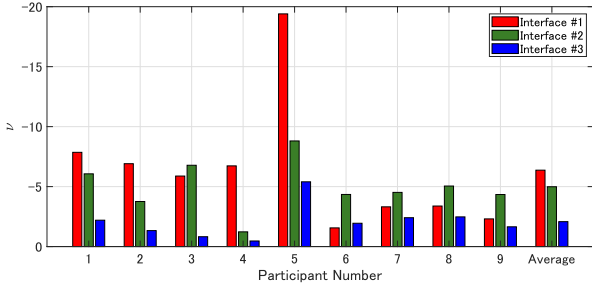


FIGURE 14. Input passivity index $\bar{\nu}$ for the nine trial subjects.

questionnaire [42] in Fig. 15. We see from this figure that the VR interfaces reduce the workload on average. In particular, the use of the VR controller requires less workload than the joystick controller for all participants. In contrast, for subjects 2, 4, and 6, the use of the HMD imposes an additional burden. The reasons for this could be the discomfort that VR images cause or the weight of the device itself. In any case, it cannot be concluded at least from these results that the HMD reduces the burden on the person.

VI. HUMAN MODELING AND PASSIVITY ANALYSIS FOR PASSIVITY-SHORTAGE-BASED CONTROL ARCHITECTURE

The experiments covered in this section are identical to those in Section IV-B except the control architecture was changed to that in Section III-B, where we take only interface #3. To analyze the impact of the parameter k_m on the human characteristics, we focus on two settings $k_m = 0.2$ and $k_m = 0.8$. Note that the human operating target is closer to the master robot for $k_m = 0.2$, while it is closer to the actual robotic group for $k_m = 0.8$. The bode diagrams of the operating targets, namely the systems from u_h to y_h , with the three

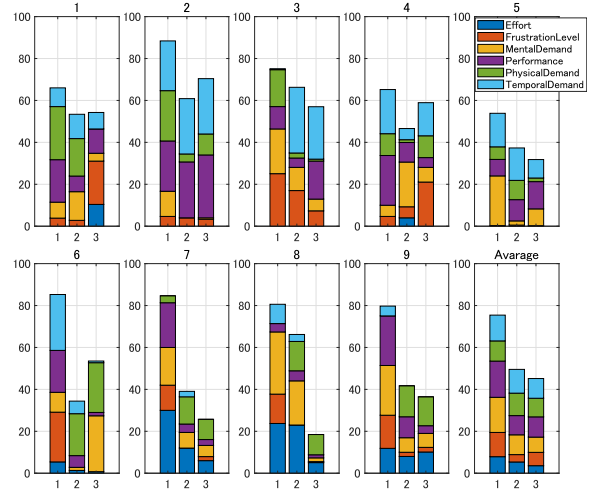


FIGURE 15. Human workload quantified by NASA TLX [42], where the number in the title of each graph is the participant number. The bars in each graph correspond to interfaces #1, #2 and #3 from left to right.

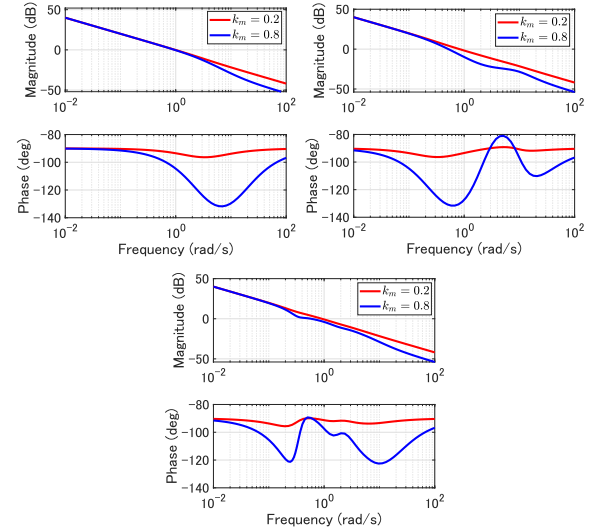


FIGURE 16. Bode diagrams of the systems from $(u_h)_i$ to $(y_h)_i$ ($i = 1, 2, 3$) for all three networks, where $k_m = 0.2$ (red) and $k_m = 0.8$ (blue).

networks for both settings of k_m are illustrated in Fig. 16. We see that $k_m = 0.8$ has more complicated dynamic properties than $k_m = 0.2$ that is close to the master robot with the single integrator dynamics.

A. HUMAN MODELING

We constructed the human models based on the operation data from the passivity-shortage-based control architecture, taken in the same way as in the previous section. We also used the same number of poles and zeros: the diagonal elements have 2 poles and 1 zero, and non-diagonal elements are constant.

The time series data of the model outputs and the verification data for $k_m = 0.2$ and $k_m = 0.8$ are shown in Figs. 17 and 18, respectively. We see from these figures that the models almost fit the verification data for both values of k_m and for all

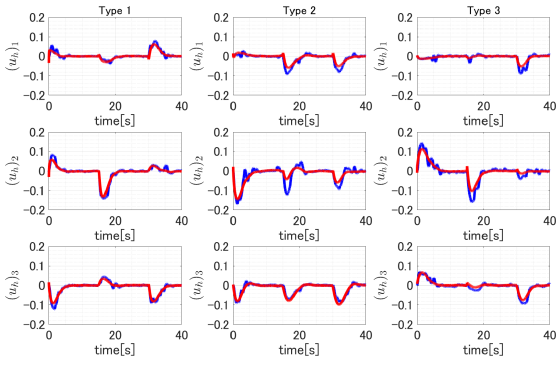


FIGURE 17. Time series data of the model outputs (red) and the verification data (blue) on u_h for $k_m = 0.2$ (left: Type 1, middle: Type 2, right: Type 3).

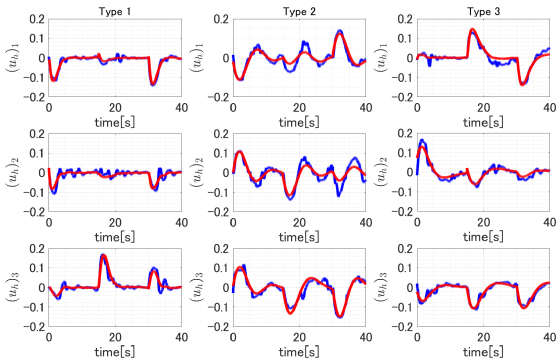


FIGURE 18. Time series data of the model outputs (red) and the verification data (blue) on u_h for $k_m = 0.8$ (left: Type 1, middle: Type 2, right: Type 3).

TABLE 4. Average model fit ratios among three elements of u_h for the passivity-shortage-based architecture.

Identification Data			
Network	Type 1	Type 2	Type 3
$k_m = 0.2$	73.07%	70.65%	65.91%
$k_m = 0.8$	71.81%	69.81%	65.74%
Verification Data			
Network	Type 1	Type 2	Type 3
$k_m = 0.2$	64.85%	60.91%	52.42%
$k_m = 0.8$	41.40%	52.32%	59.88%

network types. The average fit ratios for both the identification and verification data are summarized in Table 4. In all cases, the identification data were successfully fit by the model with the specified orders. The ratios degrade for the verification data, but they all have high values except for network Type 1 and $k_m = 0.8$. This corresponds to the left column in Fig. 18. The oscillatory behavior in the second element of u_h is not correctly fit, but the model totally fit the verification data. In view of the responses in the figures, we conclude that the model is able to correctly identify the human characteristics. On the other hand, we performed our modeling based on the hypothesis that model accuracy would deteriorate as the robot dynamics became more complex, but such a trend is not apparent in Table 4.

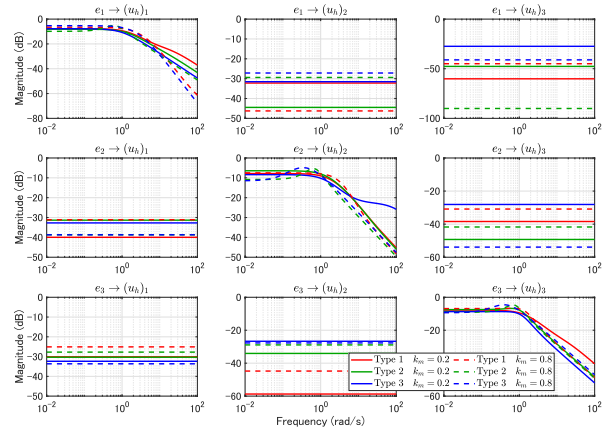


FIGURE 19. Bode diagrams of the human operator models for the three networks and two k_m .

We finally present interesting analysis results, not necessarily related to the main argument, below. Bode diagrams of the identified models are presented in Fig. 19. We see from these figures that the human characteristics differ in the value of k_m and the network type. In the frequency domain lower than 1 rd/s, the diagonal elements of the models for $k_m = 0.2$ have almost flat gains while those for $k_m = 0.8$ have gain peaks except for Type 1. The frequencies of the peaks almost correspond to the lags of the gains in Fig. 16 for each network. This is compatible with the claim of the crossover model presented in [43], and it is also highlighted in [23]. Specifically, the operator tries to shape the open-loop transfer function so that it gets closer to the single integrator while learning the inverse model of the system to be operated. This also explains the validity of the absence of peaks for $k_m = 0.2$ and network Type 1.

B. HUMAN PASSIVITY ANALYSIS

In this subsection, we examine human passivity for the passivity-shortage-based control architecture. We begin by noting that [24] showed that the cascade system needs to be passivity short from $r_p - y_h(t)$ to $q_m(t) - r_p$ with an impact coefficient $\bar{\nu}$ greater than -1 , thus enabling (3) for any network. This means that the passivity index ν for the system $\frac{H(s)}{s}$ with the operator model $H(s)$ should be greater than -1 over the entire frequency domain.

We also note that the index ν for $\frac{H(s)}{s}$ takes extremely small values over the domain lower than 10^{-2} . This is because the model does not correctly identify the human property over such a low frequency domain since we use only data over 150 s. Meanwhile, passivity over the low frequency domain does not matter in practice, because the gain crossover frequency lies between 10^{-1} rad/s and 1 rd/s for all settings. This is why we check the index for $\frac{100H(s)}{100s+1}$, which has almost the same frequency responses as $\frac{H(s)}{s}$, at least over the domain greater than 10^{-1} rad/s.

The passivity indices for the three networks and the two k_m are given in Fig. 20. We see from these figures that $\bar{\nu} > -1$

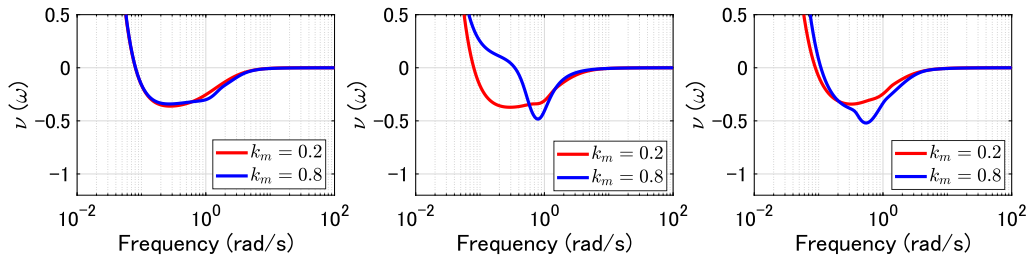


FIGURE 20. Passivity index ν for the passivity-shortage-based control architecture (left: Type 1, middle: Type 2, right: Type 3).

is satisfied for all networks and k_m . The human operator is thus expected to behave as a system meeting the requirement in [24].

Remark 1: The appropriate selection of k_m was discussed in [24], [25]. In these papers, NASA TLX studies showed that a large k_m tends to increase the workload of the operator due to the complexity of the dynamics of the system controlled by the operator. On the other hand, a small k_m makes the real robots invisible from the human, which degrades transparency for the operator. It is thus expected to manage the trade-off between transparency and workload online by appropriately tuning k_m based on the human state like fatigue and skill level. We leave this issue to the future work.

VII. CONCLUSION

In this article, we studied human-enabled multi-robot navigation in three dimensions designed based on passivity. The high dimensionality posed new challenges not in operations up to two dimensions. Namely, we had to consider how to enable humans to understand the robots' 3-D information and to stably manipulate the 3-D robots' motion. To address the issues, we prepared two pairs of command interfaces, joystick and VR controller, and feedback interfaces, 2-D display and HMD, and we conducted user studies to acquire the operation data for three interface candidates. Through these user studies, we have obtained the following four findings regarding the benefits of the VR interfaces.

- VR interfaces improve the accuracy of the human dynamic model, and thus can ease the design of human-in-the-loop systems,
- VR interfaces enhance human passivity, which contributes to enabling stable interactions between the human and robots.
- VR interfaces improve the tracking performance due to the depth information given by the HMD.
- Operators are likely to fail in achieving passivity for the 3-D operation, even with the VR interface.

We also conducted user studies on the passivity-shortage-based control architecture specifically to address this third issue. We showed that operators tend to meet the degree of passivity shortage needed to prove closed-loop stability.

There are open questions to be addressed in the future. First, we need to address whether the above insights are applied to velocity navigation. An appropriate design of the virtual robot dynamics may contribute to all aspects including human

TABLE 5. Average model fit ratios among three elements of u_h for the verification data of the 2nd-order non-diagonal blocks.

	Type 1	Type 2	Type 3
Interface 1	32.45%	24.00%	19.68%
Interface 2	34.79%	28.59%	28.72%
Interface 3	71.19%	54.50%	66.42%

model accuracy, passivity, performance and workload. How the skill level of the operator affects the model accuracy and passivity is also open. How having the operator feed back the real scenes through AR affects the human modeling, passivity, workload, and task performance is also left as future work. Linking the human properties with personal data like gender, age, and nationality should be also addressed in the future.

Finally, all user studies in this article were conducted under the permission of Administrative office of Human Subjects Research Ethics Review Committee in Tokyo Institute of Technology (Permit No. 2023109).

APPENDIX

This Appendix covers how the orders of the operator model were determined for this article. In the conference version [40], various model orders were examined under the restriction that all elements had to have the same number of poles and zeros, and we concluded that having 2 poles and 1 zero is the best for all of three interface selections and all network types, based on the fit ratio for the verification data. The fit ratios for the model in [40] are summarized in Table 5. As pointed out in [40], the model tends to have large resonance peaks in the non-diagonal elements, as can be seen in the blue curves in Fig. 21. The same applies to the human model for the passivity-shortage-based architecture in Fig. 22. Given that this is a human model, it is reasonable to assume that these resonance peaks are the result of overfitting rather than correct identification. To avoid overfitting, we applied regularization techniques prepared by the System Identification Toolbox, but the resulting model accuracy was far worse than those in Table 5. We thus set the order of the non-diagonal elements to 0, which produces the models illustrated by the red curves in Figs. 21 and 22. Comparing Tables 2 and 5, we can see that this model not only eliminates the peaks but also achieves slightly better model accuracy. This is why we took the present model in the discussion and subsequent human passivity analysis.

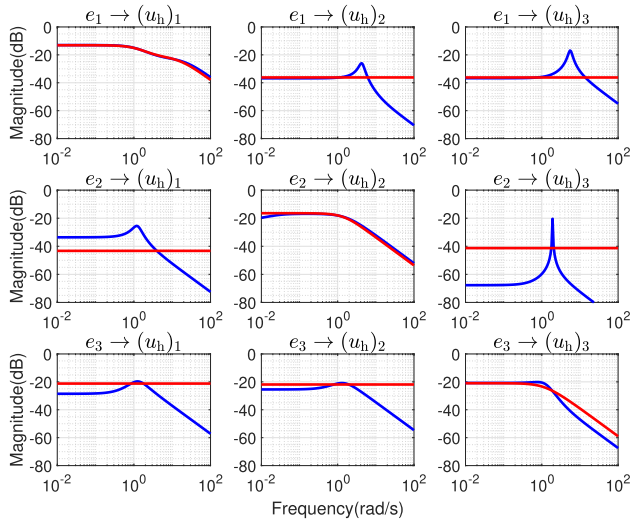


FIGURE 21. Bode diagrams of a human operator model for the passivity-based architecture with interface # 3 and the Type 1 network. The blue line represents the model with 2 poles and 1 zero for all elements, and the red represents that with 2 poles and 1 zero only for the diagonal elements and constants for non-diagonal elements.

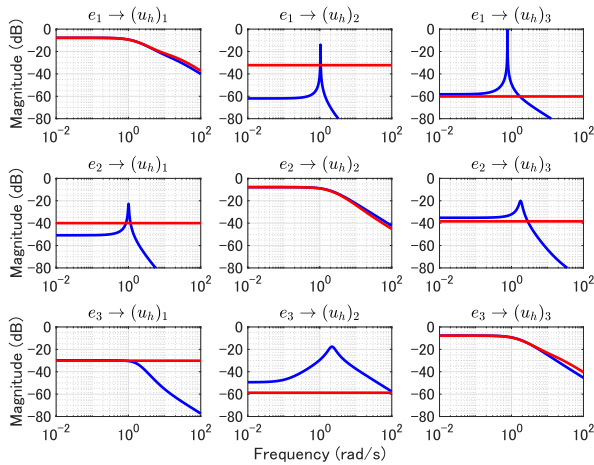


FIGURE 22. Bode diagrams of a human operator model for the passivity-shortage-based architecture with interface # 3, the Type 1 network, and $k_m = 0.2$. The blue line represents the model with 2 poles and 1 zero for all elements, and the red represents that with 2 poles and 1 zero only for the diagonal elements and constants for non-diagonal elements.

REFERENCES

- [1] S. Musić and S. Hirche, "Control sharing in human-robot team interaction," *Annu. Rev. Control*, vol. 44, pp. 342–354, 2017.
- [2] T. Sheridan and W. Verplank, "Human and computer control of undersea teleoperators," DTIC Document ADA057655, 1978.
- [3] J. Chen and M. Barnes, "Human-agent teaming for multirobot control: A review of human factors issues," *IEEE Trans. Human-Mach. Syst.*, vol. 44, no. 1, pp. 13–29, Feb. 2014.
- [4] P. Hokayem and M. Spong, "Bilateral teleoperation: An historical survey," *Automatica*, vol. 42, no. 12, pp. 2035–2057, 2006.
- [5] E. Nuño, L. Basañez, and R. Ortega, "Passivity-based control for bilateral teleoperation: A tutorial," *Automatica*, vol. 47, no. 3, pp. 485–495, 2011.
- [6] T. Hatanaka, N. Chopra, M. Fujita, and M. Spong, *Passivity-Based Control and Estimation in Networked Robotics*. Berlin, Germany: Springer, 2015.
- [7] D. Lee and M. Spong, "Bilateral teleoperation of multiple cooperative robots over delayed communication network: Theory," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2005, pp. 360–365.
- [8] G. Gioioso, A. Franchi, G. Salvietti, S. Scheggi, and D. Prattichizzo, "The flying hand: A formation of UAVs for cooperative aerial tele-manipulation," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2014, pp. 4335–4341.
- [9] M. Mohammadi, A. Franchi, D. Barcelli, and D. Prattichizzo, "Cooperative aerial tele-manipulation with haptic feedback," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2016, pp. 5092–5098.
- [10] N. Staub et al., "The Tele-MAGMaS: An aerial-ground comanipulator system," *IEEE Robot. Automat. Mag.*, vol. 25, no. 4, pp. 66–75, Dec. 2018.
- [11] E. Rodríguez-Seda, J. Troy, C. Erignac, P. Murray, D. Stipanovic, and M. Spong, "Bilateral teleoperation of multiple mobile agents: Coordinated motion and collision avoidance," *IEEE Trans. Control Syst. Technol.*, vol. 18, no. 4, pp. 984–992, Jul. 2010.
- [12] A. Franchi, C. Secchi, H. Son, H. Bulthoff, and P. Giordano, "Bilateral teleoperation of groups of mobile robots with time-varying topology," *IEEE Trans. Robot.*, vol. 28, no. 5, pp. 1019–1033, Oct. 2012.
- [13] L. Sabattini, B. Capelli, C. Fantuzzi, and C. Secchi, "Teleoperation of multi-robot systems to relax topological constraints," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 4558–4564.
- [14] W.-T. Li and Y.-C. Liu, "Human-swarm collaboration with coverage control under nonidentical and limited sensory range," *J. Franklin Inst.*, vol. 356, no. 16, pp. 9122–9151, 2019.
- [15] Y. Yang, D. Constantinescu, and Y. Shi, "Proportional and reachable cluster teleoperation of a distributed multi-robot system," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 8984–8990.
- [16] M. Dyck, A. Jazayeri, and M. Tavakoli, "Is the human operator in a teleoperation system passive?," in *Proc. IEEE World Haptics Conf.*, 2013, pp. 683–688.
- [17] A. Jazayeri and M. Tavakoli, "Bilateral teleoperation system stability with non-passive and strictly passive operator or environment," *Control Eng. Pract.*, vol. 40, pp. 45–60, 2015.
- [18] D. Venkateswaran and Z. Qu, "A passivity-shortage based control design for teleoperation with time-varying delays," *IEEE Robot. Automat. Lett.*, vol. 5, no. 3, pp. 4070–4077, Jul. 2020.
- [19] I. Polat and C. W. Scherer, "Stability analysis for bilateral teleoperation: An IQC formulation," *IEEE Trans. Robot.*, vol. 28, no. 6, pp. 1294–1308, Dec. 2012.
- [20] C. L. Martinez, R. van de Molengraft, S. Weiland, and M. Steinbuch, "Switching robust control for bilateral teleoperation," *IEEE Trans. Control Syst. Technol.*, vol. 24, no. 1, pp. 172–188, Jan. 2016.
- [21] H. Tugal, J. Carrasco, P. Falcon, and A. Barreiro, "Stability analysis of bilateral teleoperation with bounded and monotone environments via Zames-Falb multipliers," *IEEE Trans. Control Syst. Technol.*, vol. 25, no. 4, pp. 1331–1344, Jul. 2017.
- [22] I. Polushin, A. Tayebi, and H. Marquez, "Control schemes for stable teleoperation with communication delay based on IOS small gain theorem," *Automatica*, vol. 42, no. 6, pp. 905–915, 2006.
- [23] T. Hatanaka, N. Chopra, J. Yamauchi, and M. Fujita, "A passivity-based approach to human-swarm collaborations and passivity analysis of human operators," in *Trends in Control and Decision-Making for Human-Robot Collaboration Systems*, Y. Wang and F. Zhang, Eds. Cham, Switzerland: Springer, 2017, pp. 325–355.
- [24] M. Atman, K. Noda, J. Yamauchi, T. Hatanaka, and M. Fujita, "On passivity-shortage of human operators for a class of semi-autonomous robotic swarms," in *Proc. 2nd IFAC Conf. Cyber- Phys. Hum. Syst.*, 2018, pp. 21–27.
- [25] T. Hatanaka, J. Yamauchi, M. Fujita, and H. Handa, "Contemporary issues and advances in human-robot collaborations," in *Cyber-Physical-Human Systems: Fundamentals and Applications*, A. Annaswamy, P. Khargonekar, F. Lamnabhi-Lagarrigue, and S. K. Spurgeon, Eds. Hoboken, NJ, USA: Wiley, 2023, pp. 365–399.
- [26] Z. Qu and M. Simaan, "Modularized design for cooperative control and plug-and-play operation of networked heterogeneous systems," *Automatica*, vol. 50, no. 9, pp. 2405–2414, 2014.
- [27] R. Chadalavada, H. Andreasson, R. Krug, and A. Lilienthal, "That's on my mind! Robot to human intention communication through on-board projection on shared floor space," in *Proc. IEEE Eur. Conf. Mobile Robots*, 2015, pp. 1–6.
- [28] M. Moniri, F. Valcarcel, D. Merkel, and D. Sonntag, "Human gaze and focus-of-attention in dual reality human-robot collaboration," in *Proc. 12th IEEE Int. Conf. Intell. Environ.*, 2016, pp. 238–241.

- [29] G. Bolano, C. Juelg, A. Roennau, and R. Dillmann, "Transparent robot behavior using augmented reality in close human-robot interaction," in *Proc. 28th IEEE Int. Conf. Robot Hum. Interactive Commun.*, 2019, pp. 1–7.
- [30] C. Vogel, C. Walter, and N. Elkmann, "A projection-based sensor system for safe physical human-robot collaboration," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, 2013, pp. 5359–5364.
- [31] A. De Giorgio, M. Romero, M. Onori, and L. Wang, "Human-machine collaboration in virtual reality for adaptive production engineering," *Procedia Manuf.*, vol. 11, pp. 1279–1287, 2017.
- [32] D. Huy, I. Vietcheslav, and S. Gerald, "See-through and spatial augmented reality—A novel framework for human-robot interaction," in *Proc. 3rd IEEE Int. Conf. Control, Automat. Robot.*, 2017, pp. 719–726.
- [33] A. Hietanen, R. Halme, J. Latokartano, R. Pieters, M. Lanz, and J. Kämäräinen, "Depth-sensor–projector safety model for human-robot collaboration," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. Workshop Robotic Co-Workers 4.0*, 2018.
- [34] A. Hietanen, R. Lanz, and J. L. J. Kämäräinen, "AR-based interaction for human-robot collaborative manufacturing," *Robot. Comput. Integr. Manuf.*, vol. 63, 2020, Art. no. 101891.
- [35] S. Green, J. Chase, X. Chen, and M. Billingham, "Evaluating the augmented reality human-robot collaboration system," *Int. J. Intell. Syst. Technol. Appl.*, vol. 8, no. 1-4, 2010, Art. no. 13900143.
- [36] D. Bambüsek, Z. Materna, M. Kapinus, V. Beran, and P. Smrž, "Combining interactive spatial augmented reality with head-mounted display for end-user collaborative robot programming," in *Proc. IEEE 28th Int. Conf. Robot Hum. Interactive Commun.*, 2019, pp. 1–8.
- [37] R. Hetrick, N. Amerson, B. Kim, E. Rosen, E. de Visser, and E. Phillips, "Comparing virtual reality interfaces for the teleoperation of robots," in *Proc. Syst. Inf. Eng. Des. Symp.*, 2020, pp. 1–7.
- [38] D. Whitney, E. Rosen, E. Phillips, G. Konidaris, and S. Tellex, "Comparing robot grasping teleoperation across desktop and virtual reality with ROS reality," in *Proc. 18th Int. Symp. ISRR Robot. Res.*, 2020, pp. 335–350.
- [39] M. Dianafar, J. Latokartano, and M. Lanz, "Review on existing VR/AR solutions in human–robot collaboration," *Procedia CIRP*, vol. 97, pp. 407–411, 2021.
- [40] T. Hatanaka, T. Mochizuki, J. Maestre, and N. Chopra, "Impact of VR technology on a human in semi-autonomous multi-robot navigation: Control theoretic perspective," in *Proc. 4th IFAC Workshop Cyber-Phys. Hum. Syst.*, 2022, pp. 38–43.
- [41] T. Katayama, *Subspace Methods for System Identification*. Berlin, Germany: Springer, 2010.
- [42] S. Hart and L. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," in *Human Mental Workload* (Advances in Psychology Series), vol. 52, P. Hancock and N. Meshkati, Eds. Amsterdam, The Netherlands: North-Holland, 1988, pp. 139–183.
- [43] D. McRuer, "Human dynamics in man-machine systems," *Automatica*, vol. 16, no. 3, pp. 237–253, 1980.



TAKESHI HATANAKA (Senior Member, IEEE) received the Ph.D. degree in applied mathematics and physics from Kyoto University, Kyoto, Japan, in 2007. He then held faculty positions with the Tokyo Institute of Technology, Tokyo, Japan and Osaka University, Suita, Japan. Since April 2020, he has been an Associate Professor with the Tokyo Institute of Technology. He is coauthor of *Passivity-Based Control and Estimation in Networked Robotics* (Springer, 2015) and co-editor of *Economically-enabled Energy Management* (Springer Nature, 2020). His research interests include cyber-physical-human systems and networked robotics. He was the recipient of the Kimura Award (2017), Pioneer Award (2014), Outstanding Book Award (2016), Control Division Conference Award (2018), Takeda Prize (2020), and Outstanding Paper Awards (2009, 2015, 2020, 2021, 2023) all from The Society of Instrumental and Control Engineers (SICE). He is serving/served as an SE for IEEE TSCT, an editorial board member of Annual Reviews in Control, and an AE for IEEE TSCT, *IFAC Journal of Mechatronics, Advanced Robotics* and *SICE Journal of Control, Measurement, and System Integration*, and an IEEE CSS Conference Editorial Board member.



TAKAHIRO MOCHIZUKI is currently a Master Student with the Department of Systems and Control Engineering, the School of Engineering, Tokyo Institute of Technology, Tokyo, Japan. His research interests include human–robot collaborations and networked robotics.



TAKUMI SUMINO is currently a Master Student with the Department of Systems and Control Engineering, the School of Engineering, Tokyo Institute of Technology, Tokyo, Japan. His research interests include human–robot collaborations and networked robotics.



JOSÉ M. MAESTRE (Senior Member, IEEE) received the Ph.D. degree from the University of Seville, Seville, Spain. He is currently a Full Professor with the University of Seville. He has held various positions with universities such as TU Delft, Delft, The Netherlands, University of Pavia, Pavia, Italy, Tokyo Institute of Technology, Tokyo, Japan, and Kyoto University, Kyoto, Japan. His research focuses on the control of distributed cyber-physical systems, with a special emphasis on the integration of heterogeneous agents in the control loop. He has published more than 200 journal and conference papers, co-edited several books, and led several research projects. Finally, his achievements have been recognized through several awards and honors, including the Spanish Royal Academy of Engineering's medal for his contributions to the predictive control of large-scale systems.



NIKHIL CHOPRA (Member, IEEE) received the Bachelor of Technology (Honors) degree in mechanical engineering from the Indian Institute of Technology, Kharagpur, India, in 2001, the M.S. degree in general engineering in 2003, and the Ph.D. degree in systems and entrepreneurial engineering in 2006 from the University of Illinois at Urbana-Champaign, Urbana-Champaign, IL, USA. He is currently a Professor with the Department of Mechanical Engineering, the University of Maryland, College Park, MD, USA, with an affiliate appointment, the Department of Electrical and Computer Engineering. Before joining the University of Maryland, he was a Postdoctoral Research Associate with the Coordinated Science Laboratory, the University of Illinois at Urbana-Champaign from 2006 to 2007. His research focuses on the area of systems and control. His current research interests include control and estimation in robotic systems, soft robotics, and machine learning. He is the coauthor of the book *Passivity-Based Control and Estimation in Networked Robotics*. He was previously the Co-Chair of the IEEE Technical Committee on Telerobotics. He was previously an Associate Editor for *Automatica*, IEEE TRANSACTIONS ON CONTROL OF NETWORK SYSTEMS, and IEEE TRANSACTIONS ON AUTOMATIC CONTROL.