A Physics-based Virtual Reality System Design and Evaluation by Simulating Human-Robot Collaboration

Suman Chowdhury [1], Syed T. Mubarrat [1], and Antonio Fernandes [1]

[1]Affiliation not available

April 16, 2024

Abstract

Recent advancements in VR technology facilitate tracking real-world objects and users' movements in the virtual environment (VE) and inspire researchers to develop a physics-based haptic system (i.e., real object haptics) instead of computer-generated haptic feedback. However, there is limited research on the efficacy of such VR systems in enhancing operators' sensorimotor learning for tasks that require high motor and physical demands. Therefore, this study aimed to design and evaluate the efficacy of a physics-based virtual reality (VR) system that provides users realistic cutaneous and kinesthetic haptic feedback. We designed a physics-based VR system, named PhyVirtual, and simulated human-robot collaborative (HRC) sequential pick-and-place lifting tasks in the VE. Participants performed the same tasks in the real environment (RE) with human-human collaboration instead of human-robot collaboration. We used a custom-designed questionnaire, the NASA-TLX, and electromyography activities from biceps, middle and anterior deltoid muscles to determine user experience, workload, and neuromuscular dynamics, respectively. Overall, the majority of responses (>65%) demonstrated that the system is easy-to-use, easy-to-learn, and effective in improving motor skill performance. While compared to tasks performed in the RE, the PhyVirtual system placed significantly lower physical demand (124.90%; $p < 0.05$) on the user. The electromyography data exhibited similar trends ($p > 0.05$; $r > 0.89$) for both environments. These results show that the PhyVirtual system is an effective tool to simulate safe human-robot collaboration commonly seen in many modern warehousing settings. Moreover, it can be used as a viable replacement for live sensorimotor training in a wide range of fields.

A Physics-based Virtual Reality Haptic System Design and Evaluation by Simulating Human-Robot Collaboration

Syed T. Mubarrat, Antonio Fernandes, and Suman K. Chowdhury*

Abstract—Recent advancements in virtual reality (VR) technology facilitate tracking real-world objects and users' movements in the virtual environment (VE) and inspire researchers to develop a physics-based haptic system (i.e., real object haptics) instead of computer-generated haptic feedback. However, there is limited research on the efficacy of such VR systems in enhancing operators' sensorimotor learning for tasks that require high motor and physical demands. Therefore, this study aimed to design and evaluate the efficacy of a physics-based ~~virtual reality~~ VR system that provides users realistic cutaneous and kinesthetic haptic feedback. We designed a physics-based VR system, named PhyVirtual, and simulated human-robot collaborative (HRC) sequential pick-and-place lifting tasks in the VE. Participants performed the same tasks in the real environment (RE) with human-human collaboration instead of human-robot collaboration. We used a custom-designed questionnaire, the NASA-TLX, and electromyography activities from biceps, middle and anterior deltoid muscles to determine user experience, workload, and neuromuscular dynamics, respectively. Overall, the majority of responses (>65%) demonstrated that the system is easy-to-use, easy-to-learn, and effective in improving motor skill performance. While compared to tasks performed in the RE, no significant difference was observed in the overall workload for the PhyVirtual system. The electromyography data exhibited similar trends ($p > 0.05; r > 0.89$) for both environments. These results show that the PhyVirtual system is an effective tool to simulate safe human-robot collaboration commonly seen in many modern warehousing settings. Moreover, it can be used as a viable replacement for live sensorimotor training in a wide range of fields.

Index Terms—Sensorimotor training and skill acquisition, virtual reality, neuromuscular motor learning, simulation fidelity, human-robot collaboration.

I. INTRODUCTION

THE use of virtual reality (VR) system as an emerging Industry 4.0 technology continues to surge across various industrial domains to improve operators' performance through sensorimotor training [1], [2]. The main driving force behind this popularity of VR technology is that it can simulate any real environment safely and reliably with a fraction of the cost. Consequently, the scientific discovery of its' technological advancement has been evolved as a distinct field of research to provide primarily a realistic and compelling user experience, which mainly depends on usability and simulation fidelity of the VR system.

The usability of a VR system refers to many crucial factors such as ease-of-use, ease-of-learning, ease-of-interaction, and usefulness, whereas the fidelity of a VR system relates to the degree of realism the system provides in an experimental setup or a computer simulation [3], [4]. In general, researchers have used the terms "immersion" (the objective level of sensory fidelity a VR system provides, i.e., realism) [5], [6] and "presence" (the psychological sense of being in the virtual environment, i.e., engagement) [6] to define the fidelity of a VR system. Previous studies showed that the fidelity of a VR system is associated with the number of sensory modalities it stimulates [7]. For instances, many studies have demonstrated that multimodal sensory cues—haptic, auditory, and visual (HAV)—enhances experience, motor skill performance, and learning of the VR users [8-10] compared to bimodal (audio and visual—AV) or unimodal VR system and is as effective as live training for motor skill learning [11], [12]. Nevertheless, simulating haptic interactions is challenging. Different types of haptic devices—such as wearable electromechanical device [13] and haptic fabrics [14], finger-mounted 3-DoF haptic interfaces [15], and wand-based controllers [16] have been used to provide cutaneous and/or kinesthetic haptic feedback in a virtual environment (VE). However, all these aforementioned haptic devices are expensive and provide computer-generated (i.e., non-natural) limited force feedback. Consequently, they are not suitable for many applications (e.g., gross-motor tasks that require high force demands) due mainly to the complexity of simulating force-displacement relationships on the scale and precision of human force sensing [17]. Furthermore, their non-natural haptic feedback fails to provide a high VR interaction and display fidelity (i.e., a sense of engagement, haptic realism, and motor skills) similar to real-world performance [18].

Recent advancements in VR technology have led to step-up changes in designing a high-fidelity VR haptic system. The development of VR trackers or the integration of VR and motion capture systems can facilitate researchers/practitioners tracking the movement of real-world objects or users in the VE in a cost-effective manner. This technological advancement in haptic technology aids researchers/practitioners to develop high-fidelity virtual replications of real-world objects (i.e., physics-based modeling), in which users are provided with real
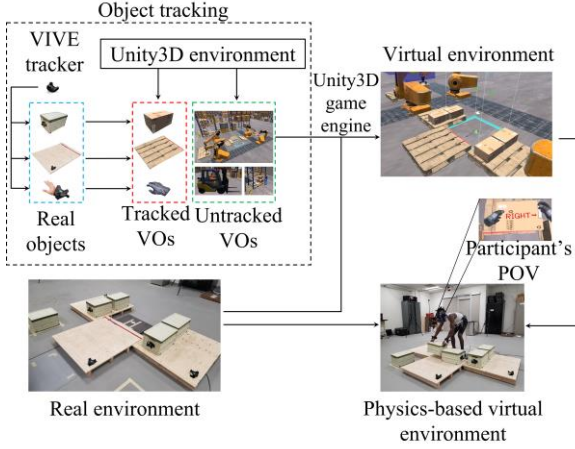
Fig. 1. Flowchart of the PhyVirtual system design process. Both 'Real objects' and 'Tracked virtual objects' existed in both real and virtual environments. Untracked virtual objects were interior components of the virtual environment that did not exist in the real environment. POV and VO denote the point-of-view of the participant in the virtual environment and virtual object, respectively.

cutaneous (object geometries) and kinesthetic (force/vibration) feedback of the interactive objects (i.e., physics-based haptics) at a much lower cost. A few recent studies have exploited the benefits of this physics-based haptic technology in their respective research [18], [19]. For instance, Cooper et al. [18] simulated a wheel change task on a virtual car by using the integration of VR and motion capture system. They used the physical prop of the wheel to provide the geometric and force feedback and two vibration motors attached to hand gloves to provide vibration feedback to the participants. Using the subjective measures, they observed that the physics-based haptics increased the user experience compared to bimodal (audio and video) and unimodal (audio or video only) sensory cues. To our knowledge, the usability of a physics-based VR system in terms of engagement and realism, motor skill (i.e., neuromuscular dynamics), and workload, and their interplays with regard to real-world tasks in a complex dynamic industrial environment—which are essential to ensure the degree of exactness and the transfer of learning skills from the VE to real environments [20], [21]—has yet to be explored.

Therefore, this study aimed to develop a high-fidelity physics-based VR system providing natural haptics to the users and the first-hand experience of a complex dynamic industrial environment and evaluate its usability. The significance of such virtual replications of a dynamic industrial environment was that their real-world staging requires high capital investments, specific geographic settings, or/and involves safety risks. For example, in recent times, human-robot collaboration (HRC) has been increasingly used to reduce human operators' risk of injuries and increase productivity in warehousing environments. Thus, an ergonomic training program using immersive technologies is also crucial to prepare new and old operators for the safe and efficient performance of the HRC tasks [22]. However, simulating the real-world training scenarios of HRC pick-and-place tasks for inexperienced/novice users in warehouses (or in advanced manufacturing or military settings) is impractical due mainly to high capital investment and the risk of injury to the users. On the contrary, a combination of some physics-based modeled objects (i.e., virtual counterparts of real objects) and high fidelity virtual industrial environment may not only address the aforementioned concerns at a lower cost but also mitigate the limitations of computer-generated haptics. In this study, we simulated HRC pick-and-place lifting tasks commonly seen in a warehousing environment to evaluate the usability of the developed VR system. In particular, we attempted to investigate how the system influences users' sense of engagement and realism, functional workload, and neuromuscular dynamics for tasks that require high force demand and gross-motor skills. We hypothesized that combining the physics-based haptic system with high-fidelity virtual objects would provide a same level of sense of engagement, functional workload and neuromuscular effort similar to tasks performed in the real-world context. However, we hypothesized that subjects would experience a reduced amount of workload as they get more familiar with the task complexity and high-fidelity virtual objects in the virtual environment. Our main objective was to find statistically insignificant differences between virtual environment tasks (VETs) and real environment tasks (RETs) regarding functional workload, neuromuscular effort, and task performance, which would validate the feasibility of our developed system.

II. METHODS

A. PhyVirtual (Physics-based Multimodal Virtual Reality) System Design

We named our developed system the physics-based multimodal VR system (PhyVirtual) since we derived the virtual replications of the interactive objects based on their known-physical geometries to provide users with actual cutaneous and kinesthetic haptic feedback of the objects while they manipulated their virtual counterparts in the virtual environment. We developed the PhyVirtual System using the HTC Vive Pro Eye Virtual Reality System (HTC Corporation, New Taipei, Taiwan and Valve Corporation, Bellevue, Washington, USA), consisting of a head-mounted display (HMD), two base stations, two controllers, a link box, wireless add-on, and multiple VIVE trackers. The system has dual AMOLED 3.5 in diagonal HMD of 2880 x 1600 pixels resolution with a 90 Hz refresh rate, embedded eye-tracking sensors with 110° field of view (up to 120 Hz sampling rate) for a fully immersive experience. To design and display the virtual environment on a participant's HMD, we used the Unity3D game engine (Unity Technologies, San Francisco, USA) with the SteamVR plugin installed from the Unity3D asset store. Muscle activity was measured using The Delsys Trigno™ Wireless EMG system (Delsys Inc., Natick, Massachusetts, USA). In addition, we used the 12-camera motion capture system (Motion Analysis Corporation, Rohnert Park, California, USA) to record the position, movement, and orientation of subjects and real objects. Other apparatus consisted of plastic boxes of 0.42 m × 0.28 m × 0.24 m
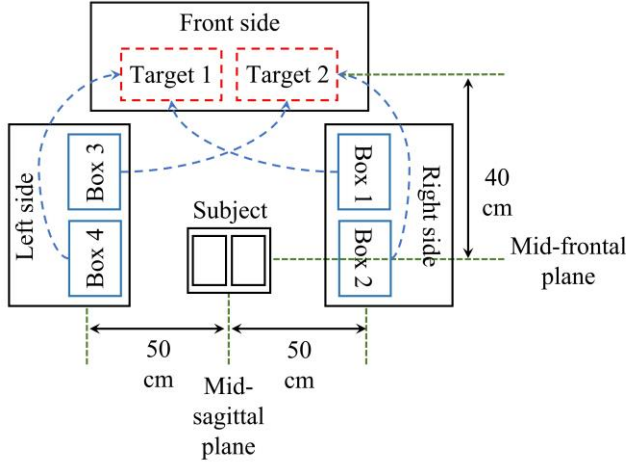
Fig. 2. The initial and target locations of the boxes and the subject. Box 1, Box 2, Box 3, and Box 4 denote locations at 30° right to the mid-sagittal plane, right mid-frontal plane, 30° left to the mid-sagittal plane, and left mid-frontal plane of the subjects, respectively. Target 1 and Target 2 denote locations at 10° left to the mid-sagittal plane and 10° right to the mid-sagittal plane of the subject, respectively.

dimensions and custom-made wooden pallets of 0.91 m × 0.91 m × 0.15 m dimensions.

The VE was designed to simulate the interior of a warehouse containing workers, forklifts, pallet racks, pallets, boxes, and four industrial collaborative robots (or cobots), which are some of the key aspects of an industrial HRC task scenario. Three real-world pallets (one in the front, and one each on the left and the right side of the participant), four boxes (weighing 20 lbs. with same dimensions), and the participants' hands (real objects in Fig. 1) were registered and connected to the VE via the attachment of one HTC VIVE tracker on each real-world object. We chose boxes weighing 20 lbs as the Lifting Index for our designed tasks with 20 lbs ranged from 1.0 to 1.5 (safe range) according to the NIOSH Lifting Equation [23]. Two base stations of the VR system located within the experimental setup area tracked the position, orientation, and movement of these trackers through SteamVR tracking technology. The data were then used to simulate the position and movement of real-world pallets, boxes, and participants' hands in the virtual environment (tracked virtual objects in Fig. 1), providing the cutaneous and kinesthetic haptic feedback to the users in the VE. We filled the virtual warehouse with forklifts, workers, pallet racks, and other interior components (untracked virtual objects in Fig. 1). These objects were added to augment the display fidelity (i.e., display realism) of the VE. We simulated a sequential HRC task, where cobots and users can perform sequential pick-and-place lifting tasks using four boxes (i.e., workpieces). Box 1, Box 2, Box 3, and Box 4 indicated initial picking locations of a box respectively placed at 30° right to the mid-sagittal plane, right mid-frontal plane, 30° left to the mid-sagittal plane, and left mid-frontal plane of the subjects at a horizontal distance of 50 cm (Fig. 2). The movement trajectories and destination locations of the boxes were controlled (Fig. 2); Target 1 and Target 2 denoted locations at 10° left to the mid-sagittal plane and 10° right to the mid-

sagittal plane, respectively. We designed four virtual industrial cobots and programmed them to deposit boxes and interact with the users in the VE (Fig. 1). Both real and virtual environments included a metronome and a varied level of workplace noise (85-90 dB), including sound effects of the robots' movements.

B. Experimental Design

Twelve healthy participants (all males; mass: 72.96 ± 6.90 kg, height: 1.74 ± 0.05 m, age: 25.92 ± 3.55 years; all university students; all right-hand dominants; two with and ten without VR experience; free from any musculoskeletal discomfort or abnormality at the time of participation) were recruited to participate in this study. The participants were recruited by distributing flyers and online advertisements across the university. All participants provided written informed consent, approved by the Institutional Review Board at Texas Tech University.

Participants performed six pick-and-place task trials—three in the RE and three in the VE. At the beginning of the task, the boxes were placed in their initial locations in the real world; however, they were not visible in the participants' HMD until the cobots deposited the boxes at their respective locations (Fig. 2). As soon as a cobot deposited a box on one of the four slots of the side pallets, the participant had to move the box from the side pallet (located at a horizontal distance of 40 cm) to one of the designated front pallet positions (Target 1 or Target 2) by twisting their upper torso in a squat lifting manner (Fig. 2). The pace of the lifting was controlled—8 seconds—using a virtual metronome (through visual and auditory feedback) in the VR environment. A 30-second rest period was provided using a virtual timer (through visual and auditory feedback) between the box pick-and-place tasks. During the rest period, another cobot deposited a box in one of the side pallets. The participant had to move the box by repeating the previous steps (Fig. 2). After two boxes were moved from side pallets to the front pallet, the experimenters removed the boxes from the experimental setup while the participant was resting. Similarly, two additional cobots deposited two other boxes, one by one, on the side pallet, and the participant moved them to the front pallet one by one. A VE task (VET) was completed when all four boxes (from both sides) were moved to the empty pallet, signifying one VET trial. Similar to VETs, participants performed the same pick-and-place sequential order in the real environment using the same background noise level (85 – 90 dB), metronome pace (8 seconds), and timer to guide both sequential tasks and rest period. The real environment task (RET) did not include industrial cobots and warehouse interiors as we conducted the study in a controlled laboratory setting with human-human collaboration (i.e., the experimenter instructed the participant through verbal guidance). The box deposition order was randomly selected, and a unique order was assigned to the individual participants throughout the session for both VETs and RETs. We also randomized the order of six

Questionnaire results (%)

⟵ Disagree ⎯ 0 ⎯ Agree ⟶

Perceived Ease-of-Use

1. Using it is effortless (it is simple to use)
| 20 | 14 | 27 | 23 | 17 |

2. Interacting with the VR display interface is clear and understandable to me
| 3 | 13 | 6 | 17 | 40 | 20 |

3. Interacting with the VR hardware is flexible and comfortable
| 10 | 26 | 13 | 30 | 20 |

4. I can use the system without written instructions
| 7 | 7 | 20 | 27 | 20 | 20 |

5. I don't notice any inconsistencies in the virtual environment as I use it
| 3 | 20 | 20 | 4 | 37 | 13 | 3 |

6. Both regular and occasional users would find the system easy to use
| 3 | 10 | 13 | 10 | 40 | 20 | 3 |

Ease-of-Learning

1. Learning it is effortless (It is easy to learn)
| 3 | 14 | 33 | 33 | 17 |

2. I find it easy to understand and learn the real-world scenarios
| 3 | 6 | 43 | 30 | 17 |

3. I quickly became skillful with it
| 3 | 10 | 27 | 33 | 27 |

4. I easily remember how to use it
| 10 | 4 | 17 | 33 | 37 |

5. I can learn the system without written instructions
| 3 | 10 | 24 | 13 | 23 | 27 |

6. Learning it is not time-consuming
| 13 | 10 | 27 | 27 | 23 |

Usefulness

1. Using this system as a training tool would enable a person to accomplish tasks more quickly
| 3 | 20 | 47 | 30 |

2. Using this system would improve motor skill performance
| 13 | 6 | 47 | 30 | 3 |

3. Using this system as a training tool would increase productivity in real-life
| 20 | 37 | 37 | 7 |

4. Using the system would enhance the effectiveness in real-life
| 10 | 20 | 30 | 30 | 10 |

5. Using the system would make a person more situation-aware in real-life
| 13 | 14 | 30 | 33 | 10 |

6. Using the system would make it easier to do the job in real-life
| 13 | 16 | 40 | 30 |

7. Using this system is useful to learn real-world scenarios
| 10 | 10 | 40 | 33 | 7 |

Comparison between VET and RET

1. I don't notice any inconsistencies between VE and RE as I use it
| 6 | 23 | 17 | 6 | 30 | 13 | 4 |

2. Performing tasks using this system is more comfortable than RETs
| 17 | 27 | 6 | 30 | 17 | 3 |

3. Performing tasks using this system is equally effective like RETs
| 37 | 34 | 23 | 7 |

4. Performing tasks using this system is more effective than RETs
| 3 | 20 | 27 | 16 | 20 | 10 | 3 |

☐ Strongly disagree　☐ Firmly disagree　☐ Disagree　☐ Neutral　☐ Agree　☐ Firmly agree　☐ Strongly agree

Fig. 3. Frequency percentage distribution of the questionnaire survey results, reflecting the participants' agreement-disagreement level for each question in the questionnaire. The percentages on the left of the mid-axis line indicate the disagreement percentages (maximum 100%), whereas those on the right signify the agreement percentages (maximum 100%). A seven-point psychometric Likert scale was used to rate each question in the questionnaire, where the ratings were as follows – '1'= 'Strongly disagree,' '2' = 'Firmly disagree,' '3' = 'Disagree,' '4' = 'Neutral,' '5' = 'Agree,' '6' = 'Firmly agree,' '7' = 'Strongly agree.'

trials, with at least 3 minutes of rest in between trials. No formal practice trials were instructed as we investigated the learning effect across all VETs and RETs.

C. Data Collection

Muscle activation was measured from biceps, anterior deltoid, and middle deltoid muscles of the right side for the complete task duration. Prior to experimental tasks, participants performed two maximum voluntary contractions (MVC) for each muscle. The interelectrode distance of the sensors was 1 cm, and data were collected at a sampling rate of 2148 Hz. We followed the procedures as discussed in a previous study for surface electrode placement [24]. The position, movement, and orientation of all boxes and the subjects were recorded for the

entire duration of each task using the motion capture system. Eight surface markers were placed on each box on the eight vertices, and multiple surface markers were placed on the subject according to the Plug-in-gait model [25]. Furthermore, four markers were placed on the HMD during the VETs. The marker data were collected continuously at a sampling frequency of 100 Hz.

As time to completion (TTC) has been widely used for measuring performance and motor skills [26-28], we used the TTC for lifting each box in seconds as the measure of performance and motor skills of the subjects for VETs and RETs. The TTC (s) was defined as the time taken for a subject to move from an initial upright standing position, lift the box, place it, and come back to the initial upright standing position. We used the motion capture data to measure the TTC (s) for lifting each box. Similarly, the EMG trends were considered as another measure of motor skill to understand the synergy of neuromuscular dynamics in performing the VE and RE tasks. The NASA-TLX [29] based on a weighted average of ratings on six subscales: 'mental demand,' 'physical demand,' 'temporal demand,' 'performance,' 'effort,' and 'frustration' was used to measure the overall workload (a score from 0 to 100) of the subjects for the successful completion of each task. Based on previous literature [30-32] and our previous work [33], we developed a novel subjective questionnaire to evaluate the perceived ease-of-use, usefulness, and ease-of-learning of our PhyVirtual system along with the comparison between VET and RET (Fig. 3). We asked these structured questionnaires after each VET. The questionnaire contained four main domains—(1) perceived ease-of-use, (2) ease-of-learning, (3) usefulness, and (4) comparison between VET and RET. A seven-point psychometric Likert scale was used to rate each question in the questionnaire. The "Comparison between VET and RET" served as the measure of the system's performance in simulating an engaging and realistic VE (i.e., the measure of "engagement" and overall "realism" provided by the system).

D. Data Analysis

The EMG signals were processed using a custom-built Matlab script in MATLAB (R2019b, The Math Works Inc.) program. The raw EMG signals for the complete task duration were separated into four sections (one for each box lifting exertion). Only the EMG signals for lifting exertions starting from the right-hand side of the participants (Box 1 and 2) were considered for the muscle activity analysis. We only considered the dominant side of the participants as previous studies showed that the dominant side effect is not significant for muscle activation patterns for healthy cohorts [34-37]. These EMG signals, including the MVC trials, were demeaned, notch filtered, and band-pass filtered at 10–400 Hz. The 60 Hz equipment noise and its harmonics were also attenuated. The filtered signals were full-wave rectified and smoothed using a moving average window of 25 frames. Then, the smoothed signals were normalized with respect to the maximum amplitude of the MVC signal to minimize between-subject or between-muscle errors. The N-MAV was then calculated for each trial using a 200 ms window with 100 ms overlap and
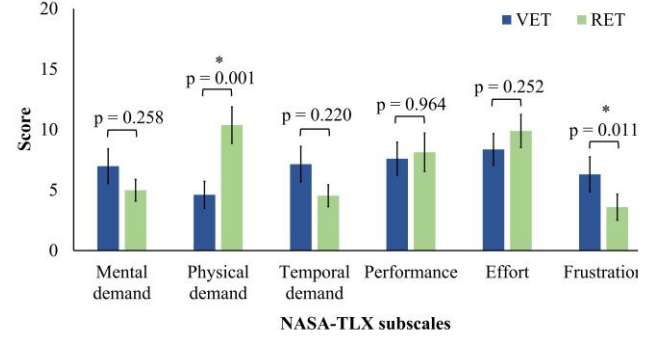

Fig. 4. Comparison of the mean of weighted ratings for different NASA-TLX subscales between VETs and RETs averaged across all subjects. Each error bar represents one standard error. VET and RET denote virtual environment task and real environment task, respectively. A p-value < 0.05 is denoted by asterisk (*).

resampled into 100 data points, with each data point representing 1% of the TTC of the box lifting exertion. Each exertion consisted of two peaks of neuromuscular dynamics, with Peak 1 representing the occurrence instance of maximum muscle torque during the acceleration phase and Peak 2 representing the occurrence instance of maximum deceleration phase of the lifting.

E. Statistical Analysis

We calculated the descriptive statistics (mean ± standard error) of the NASA-TLX scores (both individual subscales and overall) and TTCs for VETs and RETs separately by averaging across all subjects and all trials. We also estimated the N-MAVs and the percentile of TTC (%) at Peak 1 and Peak 2 by averaging across all subjects and all trials for the individual muscles and boxes (Box 1 and Box 2). To observe the trend of muscle activation for VETs and RETs separately, we also averaged the N-MAV values of every 1% between 0 to 100% task completion across all subjects and all trials and computed the Pearson's Correlation Coefficient (r) between them. For the subjective questionnaire, we determined the overall rating for each domain by averaging the ratings of the individual questions under that domain. Then, we averaged the ratings of the domains separately across all subjects. Furthermore, we analyzed the ratings of each question in the questionnaire separately by calculating their frequency percentage distributions.

We calculated the Intra-class Correlation Coefficient (ICC) (two-way mixed absolute agreement) and nonparametric Spearman correlation coefficient (ρ; with 95% significance level) between the subjects to measure the questionnaire's inter-rater reliability and validity, respectively. At first, we performed Shapiro-Wilk's test to check if the normal distribution assumption was met in the individual datasets. We employed independent samples T-tests and pairwise T-tests for the normally distributed datasets. If the normality assumption was violated, we employed non-parametric Wilcoxon signed-rank and two-sample Kolmogorov-Smirnov tests for pairwise and independent sample comparisons, respectively. We set the

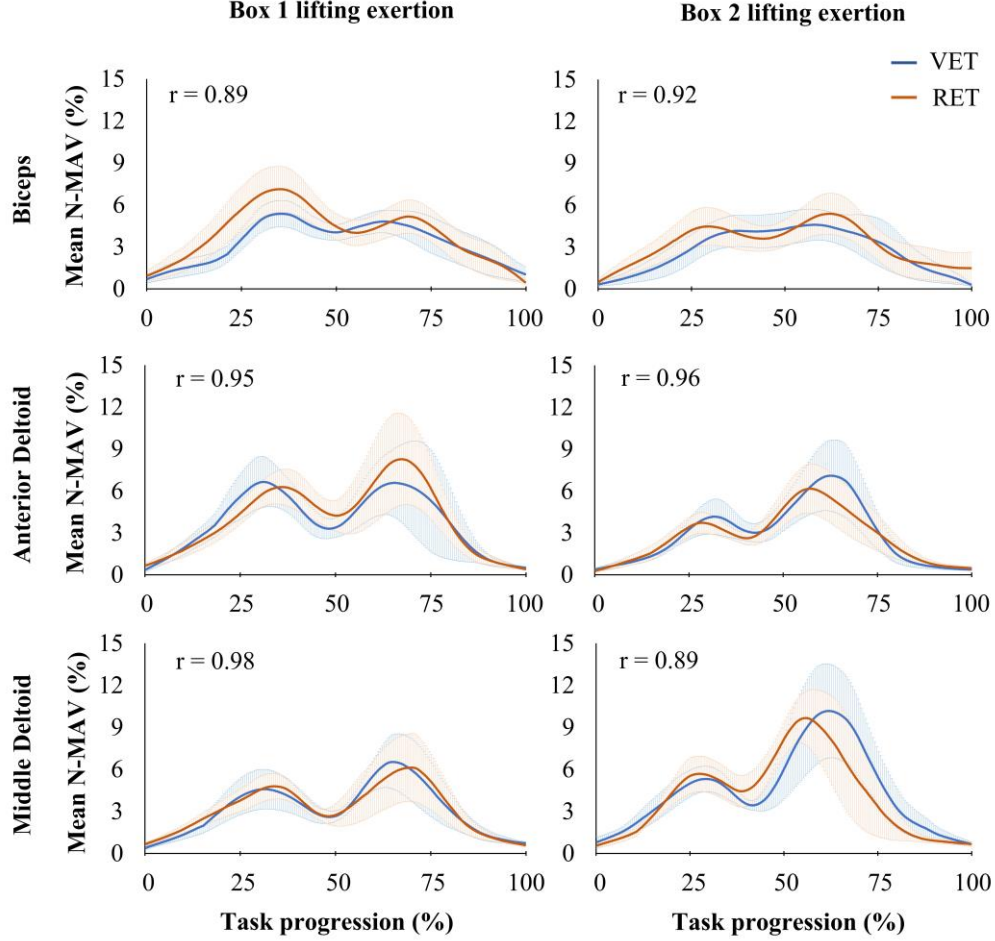**Box 1 lifting exertion** **Box 2 lifting exertion**

Fig. 5. Comparison between VETs and RETs in terms of the mean N-MAV of biceps, anterior deltoid, and middle deltoid muscles throughout the progression of the task cycle for two different box positions, averaged across all subjects and all trials with bands denoting the standard error. Box 1 and Box 2 lifting exertions denote moving a box from 30° right to mid-sagittal plane to 10° left to mid-sagittal plane and from right mid frontal plane to 10° right to mid-sagittal plane, respectively. VET and RET denote virtual environment task and real environment task, respectively. r denotes Pearson's correlation coefficient.

confidence interval at 95% ($\alpha = 0.05$). Except the inferential statistics of the questionnaire datasets, the task type (VET versus RET) was the independent variable, whereas the overall and individual scores of six NASA-TLX subscales, TTC (s), N-MAV, and TTC (%) (i.e., the occurrence) of Peak 1 and Peak 2 of all muscles were the dependent variables across all other datasets at their respective pairwise tests. We also performed pairwise T-tests to examine the effect of trial on the overall NASA-TLX workload score and TTC (s). Moreover, we implemented multiple independent samples T-tests to investigate if the mean rating score of each domain (ease-of-use, ease-of-learning, usefulness, and comparison between VET and RET) of the questionnaire datasets were significantly different from the mean score of 4, as it is the borderline score between disagree and agree rating. The Shapiro-Wilk's tests for normality showed that the individual scores of six NASA-TLX subscales, Peak 2 N-MAV of the anterior deltoid muscle for Box 1 lifting exertion, Peak 1 N-MAV and TTC (%) of the anterior deltoid muscle for Box 2 lifting exertion, and the mean ratings of the "Perceived ease-of-use" domain did not satisfy the normality assumption. Thus, we employed non-parametric

tests for these datasets.

III. RESULTS

A. User Experience

The inter-rater reliability, measured using the ICC, for all questions in our questionnaire was found to be 0.866, suggesting a high correlation between the ratings of 12 subjects. The validity measure also showed that all the questions in the questionnaire were valid, except the question "Using this system as a training tool would increase productivity in real-life" in the "usefulness" domain ($\rho = 0.060$; $p = 0.753$). As a result, we dropped this question from the statistical analysis.

The overall subjective responses for the "perceived ease-of-use" domain were 4.87 ± 0.11 (65.00% agreed, 13.33% neutral, and 21.67% disagreed; significantly higher than 4 with $z = 6.166$ and $p \ll 0.001$) (Fig. 3). Similarly, the overall Likert ratings were 5.52 ± 0.09 (81.11 % agreed, 11.11% neutral, and 7.78% disagreed; significantly higher than 4 with $t = 16.789$ and $p \ll 0.001$) and 5.06 ± 0.08 (64.29% agreed, 12.38% neutral, and 9.05% disagreed; significantly higher than 4 with $t = 13.623$

and p << 0.001) for the "ease-of-learning" and the "usefulness" domain, respectively (Fig. 3). On the contrary, the overall rating was 3.90 ± 0.13 (39.17% agreed, 15.83% neutral, and 45.00% disagreed; not significantly lower than 4 with t = -0.770 and p = 0.442) for the "comparison between the VE and the RE" domain. Nevertheless, the responses to Questions 1 and 2 in this domain revealed that almost half of the participants found no inconsistency (47% of total responses) between VET and RET and performing the VET in the PhyVirtual system to be more comfortable (50% of total responses) than RET (Fig. 3).

B. Functional Workload and User Performance

The NASA TLX data showed that the overall workload scores for VETs (41.20 ± 3.40) were not significantly (t-stat = 0.052, p = 0.959) different from RETs (41.05 ± 2.76). The TTC (s) for VETs (8.68 ± 0.28) was comparatively lower (t-stat = -1.445, p = 0.157) than RETs (9.17 ± 0.23) though the pace of all tasks was controlled using a metronome. The NASA-TLX scores of 'physical demand', 'performance', and 'effort' subscales were observed to be 124.90% (highly significant), 7.04%, and 18.33% higher for RETs compared to VETs, whereas 'mental demand', 'temporal demand', and 'frustration' subscales of the NASA-TLX showed respectively 39.88%, 57.45%, and 75.55% (significant) greater scores for VETs compared to RETs (Fig. 4). Nevertheless, the consecutive VET trials showed decreasing trends of both workload (47.25 ± 5.57, 39.05 ± 6.04, and 37.30 ± 6.14 for first, second, and third VET trials, respectively; t = 1.508 and p = 0.160 between first and second trials; t = 0.351 and p = 0.732 between second and third trials) and TTC (s) (9.09 ± 0.50 s, 8.57 ± 0.47 s, 8.38 ± 0.50 s for first, second, and third VET trials, respectively; t = 2.036 and p = 0.067 between first and second trials; t = 1.940 and p = 0.078 between second and third trials). In contrast, both workload and TTC (s) data did not show any trend across the three consecutive RET trials (workload: 40.86 ± 4.80, 43.41 ± 5.15, and 38.88 ± 4.72 for first, second, and third VET trials, respectively; t = -0.792 and p = 0.445 between first and second trials; t = 1.133 and p = 0.281 between second and third trials and TTC (s): 9.01 ± 0.28 s, 9.56 ± 0.48 s, and 8.95 ± 0.40 s for first, second, and third RET trials, respectively; t = -1.625 and p = 0.133 between first and second trials; t = 2.995 and p = 0.012 between second and third trials).

C. Neuromuscular Effort

The N-MAV results of all muscles showed similar neuromuscular activation trajectories for both VETs and RETs with high values of r, revealing consistencies in the neuromuscular dynamics, i.e., evincing similar movement patterns for both tasks (Fig. 5). The N-MAV and TTC (%) data of Peak 1 and Peak 2 instances showed statistically insignificant (p > 0.05) differences across all muscles except the TTC (%) data for the middle deltoid muscle for the Box 2 lifting task (p = 0.03) (Table I). In general, the occurrences of Peak 1 and Peak 2 were approximately at 2.5 s – 3 s and 5.5 s – 6 s of the lifting exertions, respectively. N-MAV values of biceps muscle exhibited comparatively higher (0.06% to 3.16%) activation levels consistently for RETs than VETs for both Box 1 and Box 2 lifting tasks (Table I and Fig. 5). Interestingly, N-MAV values of anterior and middle deltoid muscles were found to be relatively lower during the decelerating phase (Peak 2) but comparatively higher during the accelerating phase (Peak 1) for RETs than VETs (Table I and Fig. 5). Overall, the N-MAV magnitudes and trends revealed similar neuromuscular effort for VETs and RETs.

IV. DISCUSSION

This study focused on designing and evaluating the efficacy of a physics-based VR (PhyVirtual) system that can provide the users realistic cutaneous and kinesthetic haptic feedback through the use of physical props. Participants performed the same asymmetric pick-and-place lifting tasks in both virtual and real-world environments, while the virtual environment simulated sequential human-robot collaborative (HRC) task scenarios that are seen in real-world industrial settings. Overall, the results showed that performing the tasks using the PhyVirtual system exerted workload similar to real-world tasks and facilitated increased user performance with task familiarity compared to real-world pick-and-place tasks. Moreover, the physical interaction and high-fidelity objects in the virtual environment provided the users with a sense of engagement and realism similar to the real world and thus augmented the effectiveness of the virtual sensorimotor training.

The results of the subjective questionnaire demonstrated that the system provided users a sense of engagement and realism that were similar to a real-world industrial workplace environment. Two key features—"presence" and "realism"—of a VR system play vital roles in providing users a sense of actual presence in and control over the simulated virtual environment. For instance, Pedroli et al. [38] evaluated the usability of a VR system consisting of a cycle-ergometer, which was integrated with an immersive virtual reality system. In that study, participants provided two crucial feedbacks: 1) they did not feel active in the VE while performing the cycling task, and 2) the simulated VE did not feel real. Although the authors simulated cutaneous haptic feedback by using a real control interface, i.e., the cycle-ergometer, there was no kinesthetic or proprioceptive feedback as no movement was induced in the ergometer while simulating the cycling task. In our developed system, the subjective rating of the "Comparison between VET and RET" were not significantly different from neutral, suggesting that that the system provided the users with a sense of engagement and realism similar to the real world. This implies that the presence of kinesthetic haptic feedback increased the realism and a sense of engagement compared to non-natural haptic feedback. Additionally, one study that has used passive kinesthetic haptic feedback to improve interaction and learning in VR environments [18] evaluated the user experience (including comfort and enjoyment) of the system using a 10-point scale (0: strongly disagree, 10: strongly agree). The mean score of 7.51 ± 0.37 for their system containing auditory, visual, and haptic feedback is consistent with the high scores for the "perceived ease-of-use" and "usefulness" domains of our developed system.

Out of six subscales of the NASA-TLX, data from four subscales: temporal demand, mental demand, performance, and frustration, exhibited a higher workload for the VETs than RETs. The participants experienced higher temporal demand (Fig. 4) for the VETs mainly due to the temporal requirement to match their reaction times with the movement time of the cobots in the VE [29]. On the contrary, this modulation of the individual temporal components of the lifting task by the cobots has yielded a lesser TTC compared to RETs as the task commands were modulated by the experimenter instead of cobots in the RE. This finding suggests that computer-generated high-fidelity audiovisual stimuli (e.g., digital metronome, cobots, etc.) can not only enhance the user performance (reduced task completion time) but also regulate users to gain the required motor skills. Previous VR studies have shown that virtual environments that can provide a high sense of "immersion" can help users to improve task performance [39]. In this study, we designed high-fidelity static and dynamic visual stimuli to provide a high sense of "immersion" as well as to represent real-world HRC task scenarios. However, an increase in visual stimuli and task complexity in the VE has increased the users' mental demand compared to tasks performed in the RE. Moreover, the majority of the participants (10 out of 12) did not have any prior experience of using a VR system. As a result, the first-hand familiarity of these audiovisual stimuli in the VE led to a higher degree of 'frustration' and a lower degree of 'performance' in the NASA-TLX scale. Nevertheless, as the participants became more familiar with the visual stimuli and task complexity in the VE, the overall workload score decreased across the consecutive VET trials, which is consistent with the findings of a previous study on VR systems with haptic feedback [40]. The workload trend for the RETs did not show any such increasing or decreasing trend, which demonstrated the learning effect of the physics-based VR system.

In addition to facilitating a sense of actual presence in the virtual environment, another cardinal feature of a VR training system is the precision in movement kinematics. The neuromuscular movement of a task made in virtual environments must be similar to movements in the real environment to ensure the transfer of training benefits to the real-life situation [21]. In this respect, our developed PhyVirtual system showed neuromuscular dynamics similar to real-world tasks (i.e., similar kinematic patterns and insignificant differences between VET and RET), which showed the effectiveness of the system for real-world sensorimotor rehabilitation and training applications. The aberrant finding that the occurrence of peak activation (N-MAV) of only the middle deltoid muscle at Peak 2 instance of the lifting task was significantly different between VETs and RETs (Table I) can be explained by the fact that although both lifting load and technique were same and the dynamics (pace and trajectory) of the lifting were controlled, some differences existed due to subject-specific differences in the lifting dynamics, namely individual differences in movement trajectories (i.e., movement kinematics).

Surprisingly, even though the neuromuscular effort did not show any statistically significant differences in the muscle activation between VET and RET, the subjective workload data showed that the physical demand was significantly lesser for VETs compared to RETs. As a result, the overall workload score was found to be slightly lower for the VETs compared to RETs, which is not in line with previous VR studies based on haptic [12] and no-haptic feedback [41], [42]. For example, roller-based and joystick-based control interfaces for multiple VR driving scenarios [41] and the Phantom Omni® Haptic Device for a psychomotor test in a VE [12] have been used to provide haptic feedback to users. The relatively higher physical demand and effort scores reported in those studies for VETs could have been due to inconsistencies in the geometries and weights of the haptic control interfaces between virtual and real-world environments. On the other hand, we used real objects as the control interface for the virtual objects, which could have felt more natural to the users, leading to lower physical demand in the VETs compared to the previous studies. The implication of this interesting finding is that while designing a real-world work system or product based on VR-based sensorimotor training performance, human factors practitioners should consider a higher margin of workload (i.e., safety margin) for the real-world counterpart. Nevertheless, this finding also warrants a large-scale biomechanical study (e.g., a gender-balanced large sample size with an equal proportion of experienced and inexperienced study subjects) to investigate any differences in the physical workload experience as the EMG trends for VETs were not significantly different (albeit slightly lower) than RETs.

There are a few limitations of this study that need to be acknowledged. Firstly, our sample size was small and did not include any female subjects. This might have influenced the statistical power and the Type II error of the tests, i.e., some of the non-significant results might have been observed due to the small sample size and effect size. In future studies, we aspire to evaluate this system diverse age and sex groups to get better insights into its' usability and investigate whether males and females show similar performance patterns. Additionally, we seek to evaluate the construct validity of the survey questionnaire for a larger and diverse participant pool by implementing standard statistical analyses, such as factor analysis. Secondly, the simulation of human-robot collaborative tasks with the presence of actual robots in the real environment could provide a fairer comparison between VETs and RETs and more insights (e.g., task complexity, safety issues, functional workload etc.) about human-robot physical interaction. Lastly, our future studies aspire to focus on analyzing the movement kinematics of the participants as they perform VETs and RETs to further validate the training effectiveness of the PhyVirtual system.

V. CONCLUSION

In this study, we presented the design of a novel physics-based VR system and evaluated the system by simulating a pick-and-place task scenario. The user evaluation and

neuromuscular dynamics results showed that, while compared to similar tasks in the real environment, the PhyVirtual system is highly effective in simulating tasks that require high gross-motor demand, with a high sense of engagement and realism. We believe that our developed system can be used as a viable replacement for live sensorimotor training in a wide range of fields. Moreover, it can inspire a new generation of other researchers to adopt physics-based VR haptic system, which is affordable compared to haptic gloves, as a valuable tool to explore how humans learn and transfer motor skills in various real-world contexts.

REFERENCES

[1]   K. K. Yuen, S. H. Choi, and X. B. Yang, "A Full-immersive CAVE-based VR Simulation System of Forklift Truck Operations for Safety Training," *Computer-Aided Design and Applications,* vol. 7, no. 2, pp. 235-245, 2010/01/01 2010.

[2]   P. Zawadzki, P. Buń, and F. Górski, "Virtual reality training of practical skills in industry on example of forklift operation," in *International Conference on Innovation, Engineering and Entrepreneurship*, J. Machado, F. Soares, and G. Veiga Eds.: Springer, 2018, pp. 46-52.

[3]   S. J. Gerathewohl, *Fidelity of simulation and transfer of training: a review of the problem*. Federal Aviation Administration Office of Aviation Medicine Report, 1969.

[4]   R. T. Hays and M. J. Singer, "Simulation fidelity as an organizing concept," in *Simulation fidelity in training system design*: Springer, 1989, pp. 47-75.

[5]   D. A. Bowman and R. P. McMahan, "Virtual reality: how much immersion is enough?," *Computer,* vol. 40, no. 7, pp. 36-43, 2007.

[6]   M. Slater, M. Usoh, and A. Steed, "Taking steps: the influence of a walking technique on presence in virtual reality," *ACM Transactions on Computer-Human Interaction (TOCHI),* vol. 2, no. 3, pp. 201-219, 1995.

[7]   K. Rangarajan, H. Davis, and P. H. Pucher, "Systematic Review of Virtual Haptics in Surgical Simulation: A Valid Educational Tool?," *Journal of surgical education,* vol. 77, no. 2, pp. 337-347, 2020.

[8]   B. L. Grant, P. C. Yielder, T. A. Patrick, B. Kapralos, M. Williams-Bell, and B. A. Murphy, "Audiohaptic Feedback Enhances Motor Performance in a Low-Fidelity Simulated Drilling Task," *Brain sciences,* vol. 10, no. 1, p. 21, 2020.

[9]   C. Harvey, E. Selmanović, J. O'Connor, and M. Chahin, "A comparison between expert and beginner learning for motor skill development in a virtual reality serious game," *The Visual Computer,* pp. 1-15, 2019.

[10] F. Škola, S. Tinková, and F. Liarokapis, "Progressive training for motor imagery brain-computer interfaces using gamification and virtual reality embodiment," *Frontiers in Human Neuroscience,* vol. 13, p. 329, 2019.

[11] G. Ameri, J. S. H. Baxter, D. Bainbridge, T. M. Peters, and E. C. S. Chen, "Mixed reality ultrasound guidance system: a case study in system development and a cautionary tale," *International Journal of Computer Assisted Radiology and Surgery,* vol. 13, no. 4, pp. 495-505, 2018/04/01 2018.

[12] W. Zhang, W. Ma, M. Brandao, D. B. Kaber, P. Bloomfield, and M. Swangnetr, "Biometric validation of a virtual reality-based psychomotor test for motor skill training," *Assistive Technology,* vol. 28, no. 4, pp. 233-241, 2016/10/01 2016.

[13] I. Choi, H. Culbertson, M. R. Miller, A. Olwal, and S. Follmer, "Grabity: A wearable haptic interface for simulating weight and grasping in virtual reality," presented at the 30th Annual ACM Symposium on User Interface Software and Technology, Quebec City, QC, Canada, October 22-25, 2017, Paper presentation.

[14] S. Condino *et al.*, "Tactile augmented reality for arteries palpation in open surgery training," in *Medical Imaging and Augmented Reality*, G. Zheng, H. Liao, P. Jannin, P. Cattin, and S.-L. Lee Eds.: Springer, 2016, pp. 186-197.

[15] C. Pacchierotti, L. Meli, F. Chinello, M. Malvezzi, and D. Prattichizzo, "Cutaneous haptic feedback to ensure the stability of robotic teleoperation systems," *The International Journal of Robotics Research,* vol. 34, no. 14, pp. 1773-1787, 2015.

[16] E. Strasnick, C. Holz, E. Ofek, M. Sinclair, and H. Benko, "Haptic links: Bimanual haptics for virtual reality using variable stiffness actuation,"

[17] presented at the CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, April 21-26, 2018.

[17] H. Culbertson, S. B. Schorr, and A. M. Okamura, "Haptics: The present and future of artificial touch sensation," *Annual Review of Control, Robotics, and Autonomous Systems,* vol. 1, pp. 385-409, 2018.

[18] N. Cooper, F. Milella, C. Pinto, I. Cant, M. White, and G. Meyer, "The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment," *PLOS ONE,* vol. 13, no. 2, p. e0191846, 2018.

[19] S. Hasanzadeh and M. Jesus, "Understanding Roofer's Risk Compensatory Behavior through Passive Haptics Mixed-Reality System," in *Computing in Civil Engineering 2019: Visualization, Information Modeling, and Simulation*, Y. K. Cho, F. Leite, A. Behzadan, and C. Wang Eds.: American Society of Civil Engineers Reston, VA, 2019, pp. 137-145.

[20] M. S. Cameirão, S. B. i Badia, E. D. Oller, and P. F. Verschure, "Neurorehabilitation using the virtual reality based Rehabilitation Gaming System: methodology, design, psychometrics, usability and validation," *Journal of neuroengineering and rehabilitation,* vol. 7, no. 1, pp. 1-14, 2010.

[21] J. H. Chien, I. H. Suh, S.-H. Park, M. Mukherjee, D. Oleynikov, and K.-C. Siu, "Enhancing fundamental robot-assisted surgical proficiency by using a portable virtual simulator," *Surgical innovation,* vol. 20, no. 2, pp. 198-203, 2013.

[22] A. Realyvásquez-Vargas, K. C. Arredondo-Soto, J. L. García-Alcaraz, B. Y. Márquez-Lobato, and J. Cruz-García, "Introduction and configuration of a collaborative robot in an assembly task as a means to decrease occupational risks and increase efficiency in a manufacturing company," *Robotics and Computer-Integrated Manufacturing,* vol. 57, pp. 315-328, 2019.

[23] T. R. Waters, V. Putz-Anderson, and A. Garg, *Applications manual for the revised NIOSH lifting equation*. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH), 1994.

[24] S. K. Chowdhury, A. D. Nimbarte, H. Hsiao, B. Gopalakrishnan, and M. Jaridi, "A biomechanical shoulder strain index based on stabilizing demand of shoulder joint," *Ergonomics,* vol. 61, no. 12, pp. 1657-1670, 2018.

[25] Vicon®, "Plug-in-gait modelling instructions," *Vicon Manual, Vicon 612 Motion Systems. Oxford Metrics Ltd., Oxford,* p. UK, 2002.

[26] W. Chinthammit *et al.*, "Ghostman: Augmented reality application for Telerehabilitation and remote instruction of a novel motor skill," (in English), *BioMed Res. Int.,* Article vol. 2014, 2014, Art no. 646347, doi: 10.1155/2014/646347.

[27] S. Choussein *et al.*, "Robotic Assistance Confers Ambidexterity to Laparoscopic Surgeons," (in English), *J. Minimally Invasive Gynecol.,* Article vol. 25, no. 1, pp. 76-83, 2018, doi: 10.1016/j.jmig.2017.07.010.

[28] S. Harenberg *et al.*, "Can Multiple Object Tracking Predict Laparoscopic Surgical Skills?," (in English), *Journal of Surgical Education,* Article vol. 73, no. 3, pp. 386-390, 2016, doi: 10.1016/j.jsurg.2015.11.013.

[29] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," in *Advances in psychology*, vol. 52, P. A. Hancock and N. Meshkati Eds.: Elsevier, 1988, pp. 139-183.

[30] A. Arora *et al.*, "Face and content validation of a virtual reality temporal bone simulator," (in English), *Otolaryngol. Head Neck Surg.,* Article vol. 146, no. 3, pp. 497-503, 2012, doi: 10.1177/0194599811427385.

[31] P. Weinstock, R. Rehder, S. P. Prabhu, P. W. Forbes, C. J. Roussin, and A. R. Cohen, "Creation of a novel simulator for minimally invasive neurosurgery: Fusion of 3D printing and special effects," (in English), *J. Neurosurg. Pediatr.,* Article vol. 20, no. 1, pp. 1-9, 2017, doi: 10.3171/2017.1.PEDS16568.

[32] W. Xu, H. N. Liang, Z. Zhang, and N. Baghaei, "Studying the Effect of Display Type and Viewing Perspective on User Experience in Virtual Reality Exergames," (in English), *Games Health J.,* Article vol. 9, no. 6, pp. 405-414, 2020, doi: 10.1089/g4h.2019.0102.

[33] S. T. Mubarrat, O. Opafunso, and S. K. Chowdhury, "The Evaluation of User Experience and Functional Workload of a Physically Inter-active Virtual Reality System," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2020*, Chicago, USA, October 5-9 2020, vol. 64, no. 1: SAGE Publications Sage CA: Los Angeles, CA, pp. 2084-2086.

[34] N. Duthilleul, E. Pirondini, M. Coscia, and S. Micera, "Effect of handedness on muscle synergies during upper limb planar movements," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015: IEEE, pp. 3452-3455.

[35] D. Liang *et al.*, "Synergistic Activation Patterns of Hand Muscles in Left- and Right-Hand Dominant Individuals," *Journal of Human Kinetics,* vol. 76, no. 1, pp. 89-100, 2021.

[36] L. Pellegrino, M. Coscia, and M. Casadio, "Muscle activities in similar arms performing identical tasks reveal the neural basis of muscle synergies," *Experimental Brain Research,* vol. 238, no. 1, pp. 121-138, 2020.

[37] J. Roh, W. Z. Rymer, E. J. Perreault, S. B. Yoo, and R. F. Beer, "Alterations in upper limb muscle synergy structure in chronic stroke survivors," *Journal of neurophysiology,* vol. 109, no. 3, pp. 768-781, 2013.

[38] E. Pedroli *et al.*, "Characteristics, usability, and users experience of a system combining cognitive and physical therapy in a virtual environment: positive bike," *Sensors,* vol. 18, no. 7, p. 2343, 2018.

[39] M. F. Montoya, J. E. Muñoz, and O. A. Henao, "Enhancing Virtual Rehabilitation in Upper Limbs With Biocybernetic Adaptation: The Effects of Virtual Reality on Perceived Muscle Fatigue, Game Performance and User Experience," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 28, no. 3, pp. 740-747, 2020.

[40] S. Pareek, P. Chembrammel, and T. Kesavadas, "Development and evaluation of haptics-based rehabilitation system," presented at the 2018 International Symposium on Medical Robotics (ISMR), Atlanta, GA, USA, March 1-3, 2018.

[41] D. C. Kamaraj, B. E. Dicianno, H. P. Mahajan, A. M. Buhari, and R. A. Cooper, "Stability and workload of the virtual reality–based simulator-2," *Archives of physical medicine and rehabilitation,* vol. 97, no. 7, pp. 1085-1092. e1, 2016.

[42] V. Mouraviev *et al.*, "Urology residents experience comparable workload profiles when performing live porcine nephrectomies and robotic surgery virtual reality training modules," *Journal of robotic surgery,* vol. 10, no. 1, pp. 49-56, 2016.