Eye-Tracking in Physical Human–Robot Interaction: Mental Workload and Performance Prediction

Satyajit Upasani, Virginia Tech, Blacksburg, VA, USA, **Divya Srinivasan**, Clemson University, Clemson, SC, USA, **Qi Zhu**, National Institute of Standards and Technology, Boulder, CO, USA, **Jing Du**, University of Florida, Gainesville, FL, USA, and **Alexander Leonessa**, Virginia Tech, Blacksburg, VA, USA

Background: In Physical Human–Robot Interaction (pHRI), the need to learn the robot's motor-control dynamics is associated with increased cognitive load. Eye-tracking metrics can help understand the dynamics of fluctuating mental workload over the course of learning.

Objective: The aim of this study was to test eye-tracking measures' sensitivity and reliability to variations in task difficulty, as well as their performance-prediction capability, in physical human–robot collaboration tasks involving an industrial robot for object comanipulation.

Methods: Participants (9M, 9F) learned to coperform a virtual pick-and-place task with a bimanual robot over multiple trials. Joint stiffness of the robot was manipulated to increase motor-coordination demands. The psychometric properties of eye-tracking measures and their ability to predict performance was investigated.

Results: Stationary Gaze Entropy and pupil diameter were the most reliable and sensitive measures of workload associated with changes in task difficulty and learning. Increased task difficulty was more likely to result in a robot-monitoring strategy. Eye-tracking measures were able to predict the occurrence of success or failure in each trial with 70% sensitivity and 71% accuracy.

Conclusion: The sensitivity and reliability of eye-tracking measures was acceptable, although values were lower than those observed in cognitive domains. Measures of gaze behaviors indicative of visual monitoring strategies were most sensitive to task difficulty manipulations, and should be explored further for the pHRI domain where motor-control and internal-model formation will likely be strong contributors to workload.

Application: Future collaborative robots can adapt to human cognitive state and skill-level measured using eye-tracking measures of workload and visual attention.

Keywords: strategies, reliability, virtual environments, motor learning, psychometrics

Address correspondence to Divya Srinivasan, Department of Industrial Engineering, Clemson University, Freeman Hall, Clemson, SC 29634, USA; e-mail: sriniv5@clemson.edu

HUMAN FACTORS
2024, Vol. 66(8) 2104–2119
DOI:10.1177/00187208231204704
Article reuse guidelines: sagepub.com/journals-permissions

INTRODUCTION

Collaborative robots (cobots) are becoming more usable, versatile, and increasingly safe to operate in close proximity with humans (Haddadin & Croft, 2016). In Physical Human–Robot Interaction (pHRI), although cobots have achieved higher standards of safety and compliance in recent years, they can still impose a significant workload on the user's attentional and cognitive-motor resources (Marchand et al., 2021; Stirling et al., 2020), and may require time and effort to learn (Aronson et al., 2018; Cornwall, 2015). Hence, it is important to understand the cognitive challenges involved in controlling these complex devices and be able to predict the consequences of such cognitive challenges on performance.

Past research has shown that performing joint actions with a robot, for example, using wearable robots such as myoelectric prostheses and powered exoskeletons or using joystick-operated robotic arms, tends to pose cognitive challenges because the user cannot easily predict the device's control dynamics (Aronson et al., 2018; Chadwell et al., 2016; Cornwall, 2015; Kao, 2009). In other words, users may find it difficult to develop an internal mental model of the cobot and hence not be able to anticipate the consequences of a joint action with the robot, thereby exhibiting an increased reliance on vision to monitor robot behavior. Human motor control literature has theorized that learning to use novel or complex tools such as cobots requires the formation of new internal models for tool behaviors, as well as updating of the internal mental models of the limb controlling the tool (Wolpert et al., 2011). Additionally, using cobots for comanipulation tasks (i.e., tasks where the human and robot cooperate to manipulate shared objects) may also be challenging because of the need to monitor the manipulated object as well as the surrounding environment for hazards or potential collisions (Steinfeld et al., 2006). Lastly, comanipulation tasks can be intrinsically difficult, due to the need to remember multiple task steps with conditional relationships, and needing to perform complex spatial transformations mentally (Van Acker et al., 2020) or deal with time pressures (Bommer & Fendley, 2018).

Thus, it is expected that there is a high mental workload present during initial stages of learning to perform joint actions with a cobot, as the user attempts to build an internal model of the device and movement (Sailer, 2005). Over the course of practice, mental workload is expected to attenuate due to refinement of neural processes and increasing automaticity in the task (Sailer, 2005; O. White & French, 2017). Understanding the dynamics of mental workload over the course of motor learning and continuous measurement of these constructs can help in designing learning/training protocols and help minimize workload for users of cobots. Eye-tracking is a promising technique for measuring mental workload and predicting performance, since it can provide both physiological measures, (e.g., pupil dilation) that correlate with the involuntary neural response to mental workload (Just et al., 2003), as well as eye-movement measures (e.g., fixation rate), which reflect voluntary gaze behavior and strategies for maximizing performance (Land, 2009; Srinivasan & Martin, 2010). Eye-tracking measures have also been shown to change over the course of learning and to be able to classify distinct stages of motor-skill acquisition (Sailer, 2005). In surgical tasks, the ability to focus on the most informative anatomical regions is an important determinant of expertise (Law et al., 2004; Zheng et al., 2021), which suggests that gaze patterns may correlate with, and predict task performance. The versatility of eye-tracking, coupled with its increasing wearability and ubiquity, make eye-tracking a viable technology to implement in dynamic, real-world environments (Cognolato et al., 2018).

Previous research on the psychometric properties of eye-tracking measures has largely focused on cognitive tasks involving change-detection or working memory (Matthews et al., 2015; Zargari Marandi et al., 2018). In the pHRI domain involving tasks requiring motor skills, although some studies have been conducted on changes in specific eye-tracking measures, the psychometric properties of eye tracking metrics are yet to be established as there have been no systematic investigations of the sensitivity and reliability of eye tracking in predicting changes in mental workload and/or performance. Studies in the domains of assistive and surgical robotics have found that higher mental workload increased pupil diameter (Aronson et al., 2018; M. M. White et al., 2017) and stationary gaze entropy (SGE) (Wu et al., 2019), and reduced the fixation rate (Novak et al., 2015). Additionally, SGE and the gaze transitions between areas of interest have been found to reduce over the course of learning (Sobuh et al., 2014; Wu et al., 2021). Recent work has also begun to explore the use of eye-tracking measures to predict performance in the pHRI domain (Aronson et al., 2018; Wu et al., 2021). Using machine-learning techniques, Aronson and colleagues were able to associate distinct scanning behaviors with different control modes of an assistive robot (Aronson et al., 2018). Other work by the same authors discussed the potential ability of gaze behavior to predict unexpected performance conditions or errors, based on which a cobot could take corrective actions (Aronson & Admoni, 2018). In robotic surgery, which requires efficient visual scanning to identify key anatomical features, stationary gaze entropy was found to predict task performance improvements (Wu et al., 2021). This finding supports earlier research that found that gaze behavior could discriminate between expert and novice surgical performance (Law et al., 2004; Wilson et al., 2010; Zheng et al., 2021).

A better and more systematic understanding of the reliability of these different eye tracking metrics, and their sensitivity to variations in task difficulty and learning, can help establish the potential utility of eye-tracking metrics as a continuous measure of human cognitive state in human–cobot interaction, thereby guiding online/adaptive control of robotic systems to be responsive to human state. A combination of pupillometric and gaze-behavior measures may also be able to effectively predict task performance in pHRI. If imminent failure can be predicted by eye-tracking metrics, the adaptability of robotic systems can be improved by enabling the design of anticipatory safety mechanisms.

Thus, this study aimed to quantify the sensitivity and reliability of eye-tracking measures of workload to variations in task difficulty and learning in a pHRI task. The ability of gaze-behavior measures to distinguish different visual strategies across task difficulty and learning, and predict task performance, were also studied. We designed a task that required participants to control a bimanual cobot to pick and place virtual objects at different target locations, while avoiding collisions with virtual objects. The task was timed, and participants performed it under two different levels of task difficulty. Several eye-tracking metrics such as pupil dilation, fixation count, fixations in different areas of interest (AOI), and gaze entropies were computed. Based on the literature reviewed, we expected that these metrics would provide different types of information related to mental workload. Specifically, pupil dilation and fixation count indicate overall workload and visual monitoring demand respectively. Stationary gaze entropy (SGE), accounting for the fixation distribution across different regions of the environment, indicates the need for visual monitoring of the task environment (to detect/avoid collisions in our task). Gaze transition entropy (GTE) can account for the transitions between different areas of the environment, and hence indicate the potential emergence of repetitive scanning patterns over the course of learning. Fixations in the different AOI extracted from eye tracking can be used to infer internal model formation. Since fixations are typically directed towards the targets in goal-directed movements, increased fixations on the manipulated cobot (vs. targets) would indicate an increased reliance on vision for controlling the cobot.

Considering these aspects of eye-tracking measures, we proposed the following aims and hypotheses. As our first aim, we quantified the sensitivity of eye-tracking metrics in a human–robot comanipulation task, by studying how human participants learned to use a cobot under two different levels of task difficulty. For the second aim, we estimated the reliability of the different eye-tracking metrics, to provide a more detailed characterization of their effectiveness as workload measures in pHRI. As a third aim, we explored the changes in visual strategies (in terms of the relative visual focus on different AOI in the environment) over the course of learning to use the robot. As the fourth aim, we explored the extent to which eye-tracking metrics during could predict pHRI task performance on a trial-to-trial basis. While Aim 4 was exploratory in nature, the hypotheses for Aims 1–3 are described below.

Hypotheses

Sensitivity and Reliability of Eye-Tracking Metrics (Aims 1 and 2). We expected SGE, GTE, pupil dilation, and fixation count to be higher in the more difficult task condition, and that these metrics would decrease over time (with learning) in both tasks, although the rate of change would depend on task difficulty. Based on the psychometric properties of eye-tracking metrics reported in cognitive domain (e.g., Matthews et al., 2015), we expected pupil dilation, fixation count, and entropy metrics to have good reliability in pHRI tasks.

Changes in Visual Strategy (Aim 3). Fixations on the manipulated cobot were expected to be higher in the high-difficulty task condition and reduce over the course of learning in both conditions.

MATERIALS AND METHODS

Experimental Setup

The inspiration for our task environment is a potential industrial scenario in which a human operator uses a physically coupled robot to lift heavy objects (perhaps oddly shaped ones) and place them in assigned target locations. The task was simulated in virtual reality (VR), since VR affords high freedom over experimental manipulations and speed in the presentation of stimuli, particularly in tasks involving physical-object interaction. VR also enables flexible measurement of motor performance and eye-movement behavior, since the position and orientation of all VR objects relative to the user is precisely known (Clay et al., 2019). Lastly, VR headsets enable the precise control of light incident on the eyes, thus helping reduce the undesirable effects of environmental illumination on pupil dilation. Participants performed the task using the Rethink Robotics Baxter Robot, which has two 7-DOF arms that can be manipulated freely by grasping their wrists. Custom handles were 3-D printed and attached to the robot wrists to enable a secure and comfortable grasp. We visualized a virtual model of Baxter inside a Unity VR environment (Unity version 2020.1.9f1) by sending Baxter's real-time joint positions to Unity at 140 Hz using the ROS# package (Zhou et al., 2020, schematic in Figure 1(a)). The HTC Vive Pro Eye VR headset was used for rendering the virtual task and environment in this study. Participants stood close to, and facing, the Baxter robot while wearing the VR headset, to coperform virtual object manipulations without needing any large head movements to scan their environment (Figure 1(b)). Participants experienced a first-person view of the task (VR visualization shown in Figure 1(c)). They manipulated a virtual plate with a ball on it, which

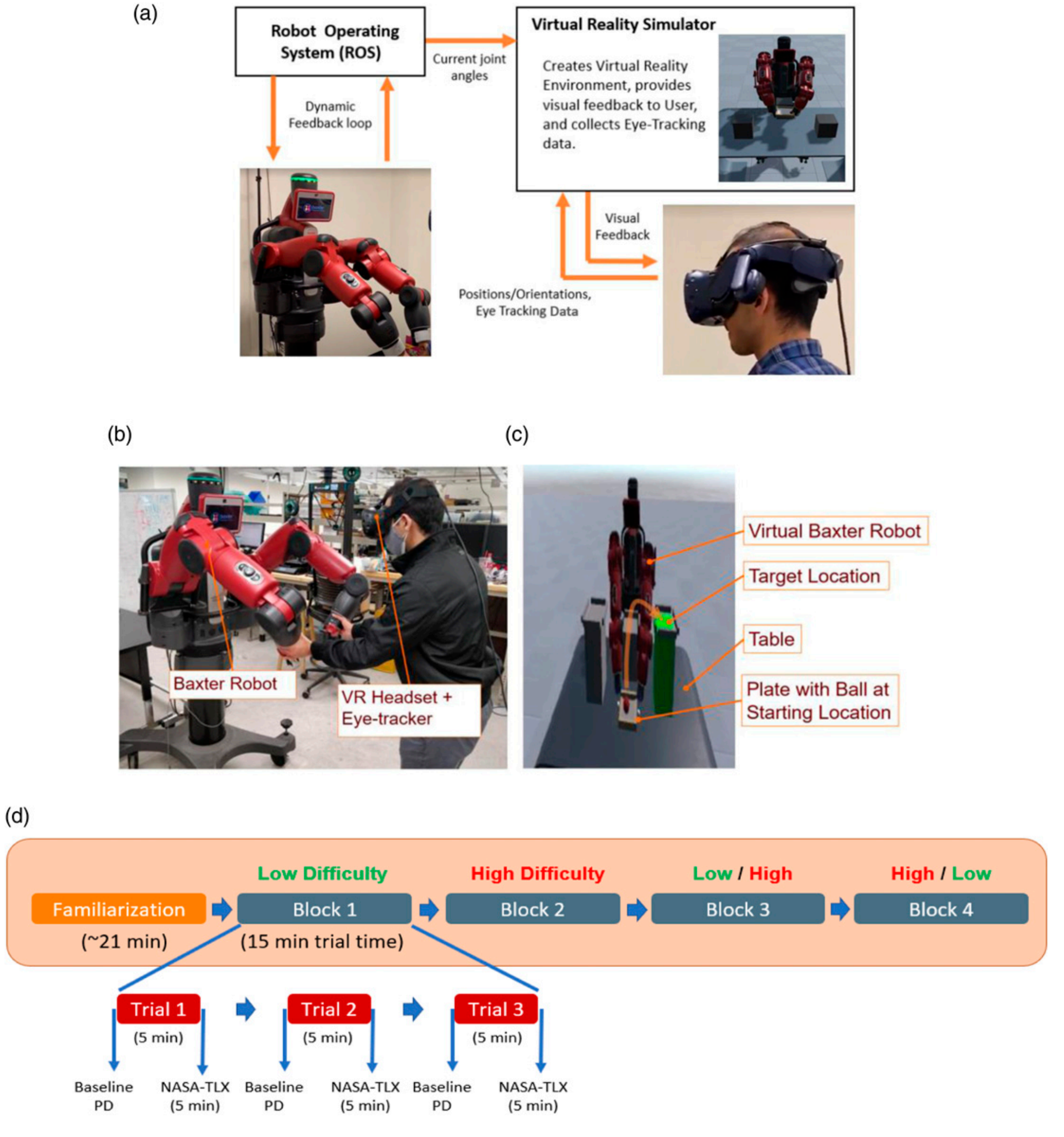

Figure 1. (a) Schematic of custom-built communication module between Baxter and VR; (b) Participant wearing the VR headset and operating the robot; (c) Participant's view in VR, with robot, plate, and target locations; (d) Experimental protocol showing familiarization followed by 4 blocks of 12 experimental trials. Baseline PD was obtained before, and NASA-TLX after, each trial.

simulated physical properties of an actual ball such as rolling around inside the plate and falling out if tilted. When cued by the experimenter, participants had to pick and transfer the object (plate with ball) on to a selected target location (any one of three target towers chosen randomly), without colliding with any virtual objects, within 10 seconds. In the case of any errors (collision, ball drop, or excessive time), the trial would be marked incomplete, and the object would be reset at the starting location.

Participants

A convenience sample of 24 participants was recruited from the local community. Data from 6 participants were lost or excluded due to the following factors—excessive head-jerking ($n = 1$), eye-strain ($n = 1$), physical exhaustion ($n = 1$), equipment failure ($n = 2$), and incomplete data recording ($n = 1$), resulting in data from a total of 18 participants (9M, 9F) for further analysis. Their mean age was 26.8 years (SD 4 years). Participants were included if they could read at arm's length without the use of corrective lenses, and were free of any recent history of musculoskeletal disorders (past 12 months). Individuals with a history of migraine, vertigo, and epilepsy were excluded, since these conditions can increase the susceptibility to VR sickness. Participants signed a written informed consent, and the research was approved by the Virginia Tech Institutional Review Board (#21-203).

Experiment Design and Protocol

The two independent variables in the experiment were task difficulty and trial number. Task difficulty was manipulated by changing the degree of "match" between the joint impedances of the two arms of the robot, with two levels, low difficulty (LD; matched impedances) and high difficulty (HD; mismatched impedances). The stiffness and damping parameters of the Baxter robot were used to vary joint impedances. The mismatched HD condition was created by stiffening some degrees of freedom on one Baxter arm, thereby limiting the arm's range of motion (ROM). This made the control of the robot less intuitive and increased motor coordination demands needed to bimanually balance the plate and ball. In a first familiarization session (Figure 1(d)), participants learnt to control the Baxter robot, put on the VR headset and practice the individual components of the main task. Familiarization was followed by the experimental session that included four blocks of trials with three trials of 5 min each. The first and second blocks were always performed in the order of LD followed by HD, to avoid transfer of learning effects. The order of difficulty (LD vs. HD) was randomized for the third and the fourth blocks (counterbalanced across participants).

Data Collection and Processing

A custom Unity script recorded pupil size and 3-D coordinates of the gaze point at 90 Hz from the eye-tracker embedded within the VR headset. When the participant looked at any VR object (also referred to as an area of interest or AOI), the gaze point intersected with the surface of the object, thus registering as a "hit" on the AOI. Each eye-tracking sample was associated with an AOI-hit. The script also recorded the 3-D coordinates of the virtual plate, ball, and the virtual Baxter's end-effectors, as well as the instantaneous timer value corresponding to transfer completion. The proportion of successful transfers in each trial was used as a measure of task performance. Task performance, fixation count, median pupil dilation (PD), stationary gaze entropy (SGE), and gaze transition entropy (GTE) were computed for each 30 s interval of each trial, using procedures described in Appendix A. The ratios of fixations on specific AOIs, that is, the plate (which included the ball), the robot-arms, and the top surfaces of the targets, per each 30-second interval were computed, to understand the relative attentional focus directed towards these AOI.

Statistical Analysis

Effect of Task Difficulty and Learning on Performance and Workload. A mixed factor ANOVA was used to test for the effects of difficulty Condition (LD, HD), Gender (male, female), Trial (1–6 for each condition), and their interactions on performance, eye-tracking metrics, and NASA-TLX mental workload ratings.

Significant effects were followed by post hoc pairwise comparisons using Tukey's HSD test. Sensitivity of eye-tracking metrics was estimated using effect size (partial eta-squared η_p^2) for main and interaction effects (Lakens, 2013). Eta-squared values were interpreted as .01 = small, .06 = medium, .14 = large (Cohen, 1988).

Reliability Analysis. The standard error of measurement (SEM), as an index of absolute reliability, and the intraclass correlation coefficient (ICC), as an index of relative reliability, were computed using procedures described in Appendix B.

Effect of Task Difficulty and Learning on AOI-Measures. To understand differences in visual behavior and strategies, the same mixed factor ANOVA described above used to test for the effects of our independent variables on AOI-measures.

Performance Prediction Using Eye-Tracking Metrics. Two logistic regression models were generated to predict the occurrence of a success (0) or failure (1)—the first model included only the independent variables of the experiment (condition, Trial, and their interaction) as predictor variables (for comparison purposes), whereas the second model only included eye-tracking metrics (pupil dilation, fixation metrics, entropy, and AOI-based metrics) as predictor variables. Confusion matrices were generated for both models, along with measures of classifier performance (accuracy, precision, recall, specificity, negative predictive value (NPV), and F1-score) (Webb, 2010).

For all analyses, the significance level was set at α = .05, and all statistical analyses were performed in JMP Pro (version 16.0.0, SAS Institute Inc., USA).

RESULTS

Performance and Perceived Workload

The proportion of successes was significantly lower in the HD condition ($p < .0001$, as seen in Figure 2(a). Additionally, successes increased significantly with time (trials) ($p < .0001$), indicating that participants learnt to better perform the task. There was no significant interaction effect of condition and trial on performance, suggesting that participants improved at a similar rate in both conditions. Females exhibited significantly lower performance than males (p = .0014). NASA-TLX mental workload ratings were significantly higher in the HD condition compared to the LD condition, ($p < .0001$), and ratings for both conditions reduced over the course of learning ($p < .0001$) accompanied by a significant Condition × Trial interaction effect ($p < .0001$). Ratings were not significantly different across genders.

Sensitivity of Eye-Tracking Metrics to Changes in Task Difficulty and Learning

All eye tracking metrics showed significant differences across condition (Figure 2). PD, FC, and SGE increased in the HD condition compared to the LD condition, and GTE was reduced in the HD condition compared to the LD condition. The effect of Trial was also significant on PD, SGE, and FC. PD and SGE reduced, and FC increased across successive trials, although the change in PD was nonmonotonic (Figure 2(c)). The Condition × Trial interaction effect was significant on PD and FC. There was no effect of Gender on eye-tracking workload metrics. The p-values and effect sizes for these statistical comparisons are shown in Table 1.

Reliability of Eye-Tracking Metrics

Table 1 also shows the participant variance components and ICC values for each metric. Based on the criteria stated in (Becser et al., 1998; Ettinger et al., 2003; Zargari Marandi et al., 2018), all eye-tracking metrics showed good reliability (>.4) except GTE, which showed poor reliability. Pupil dilation showed the highest reliability (.68), followed by SGE (.44), fixation count (.41), and GTE (.17).

Effect of Task Difficulty and Learning on AOI-measures

Considering AOI-based measures, plate-fixations significantly reduced, and robot-arm fixations significantly increased, in the HD condition compared to the LD condition (Figure 3(a) and (c)). Target-fixations were not

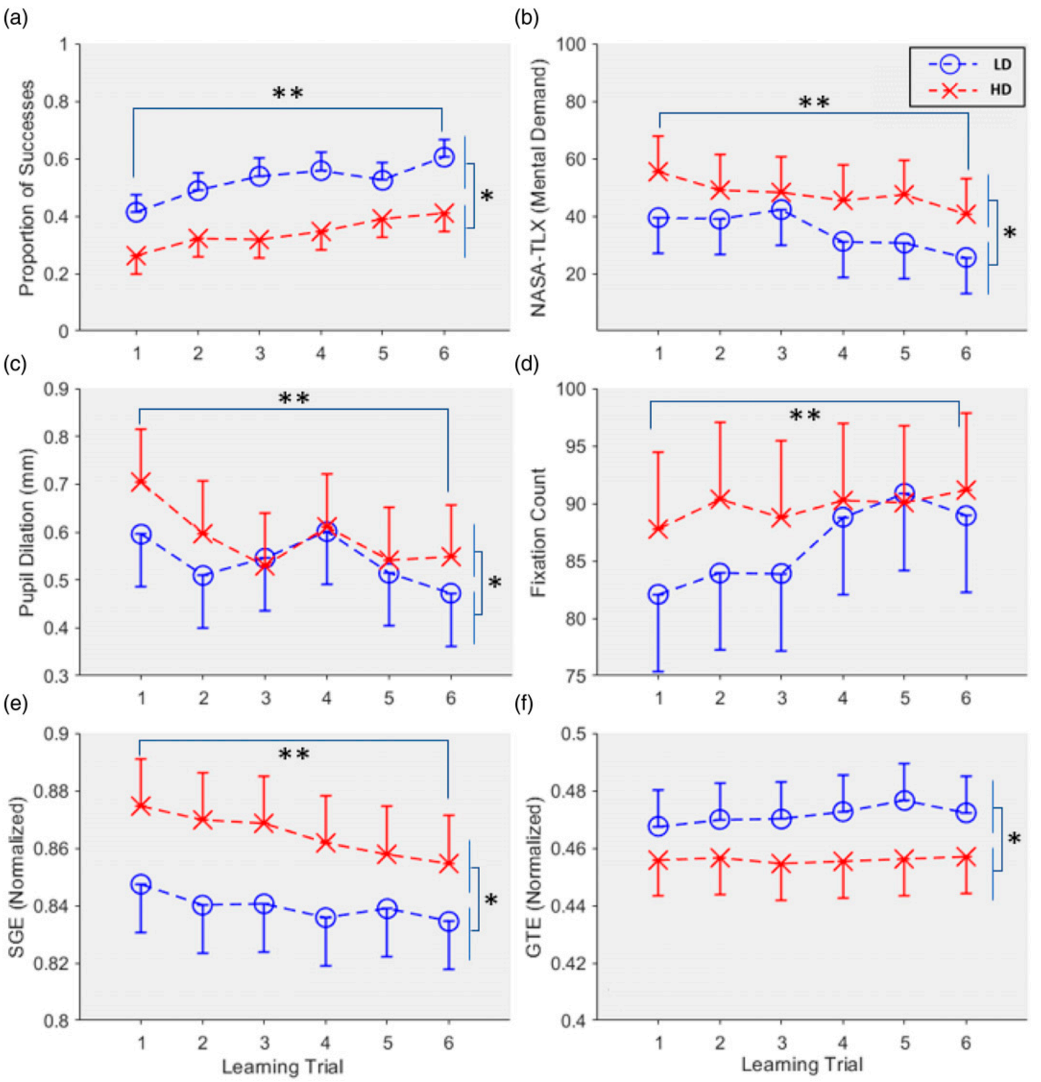

Figure 2. (a) Performance (b) NASA-TLX ratings on mental demand, and (c-f) eye-tracking workload metrics over the course of six learning trials for each task condition (low difficulty LD, high difficulty HD). Individual data points represent least squares means, and the error bars represent 1 standard error. Note: * denotes a significant main effect of condition, and ** denotes a significant main effect of Trial. Significant main effects of condition and trial were seen in performance (a) and NASA-TLX ratings (b). Among the eye-tracking metrics, pupil dilation, SGE and GTE (c,e,f) showed significant main effects of Condition, while pupil dilation, fixation count and SGE (c–e) showed significant main effects of Trial.

significantly different across conditions. The main effect of Trial on plate-fixations was significant, corresponding to an increase over time, whereas robot-arm fixations decreased over time. Significant Condition × Trial interaction effects (effect sizes and p-values in Table 2) and posthoc tests indicated that these changes over time occurred only in the HD condition. Target-fixations did not change significantly over trials. Males exhibited significantly more plate-fixations, compared to females, who exhibited significantly more robot-fixations.

TABLE 1: Sensitivity, Variance Components, and Reliability (ICC and SEM) of Eye Tracking Metrics

	Effect Size (η_p^2) (*p*-value)			Participant Variance Components		Reliability Metrics	
Workload Measures	Condition	Trial	Condition x Trial	Between (S^2_{Bs}) [LCI-UCI]	Within (S^2_{Ws}) [LCI-UCI]	ICC	SEM
Pupil dilation (mm)	**.019 (<.0001)**	**.085 (<.0001)**	**.019 (<.0001)**	.054 [.013, .07]	.026 [.024, .03]	.68	.16
Fixation count	**.005 (.0009)**	**.016 (<.0001)**	**.006 (.02)**	184 [58.8, 309.1]	263.7 [248.5, 280.3]	.41	16.24
SGE (normalized)	**.021 (<.0001)**	**.02 (<.0001)**	.0027 (.33)	.001 [.0004, .002]	.0015 [.0014, .0015]	.44	.04
GTE (normalized)	**.002 (.03)**	.0009 (.85)	.0008 (.86)	.0005 [.0002, .0009]	.0025 [.0023, .0026]	.17	.05

Sensitivity is quantified using partial eta-squared (η_p^2) as a measure of effect size. Bold fonts indicate significant effects ($p < .05$). The 95% lower and 95% upper confidence limits are shown in square brackets.

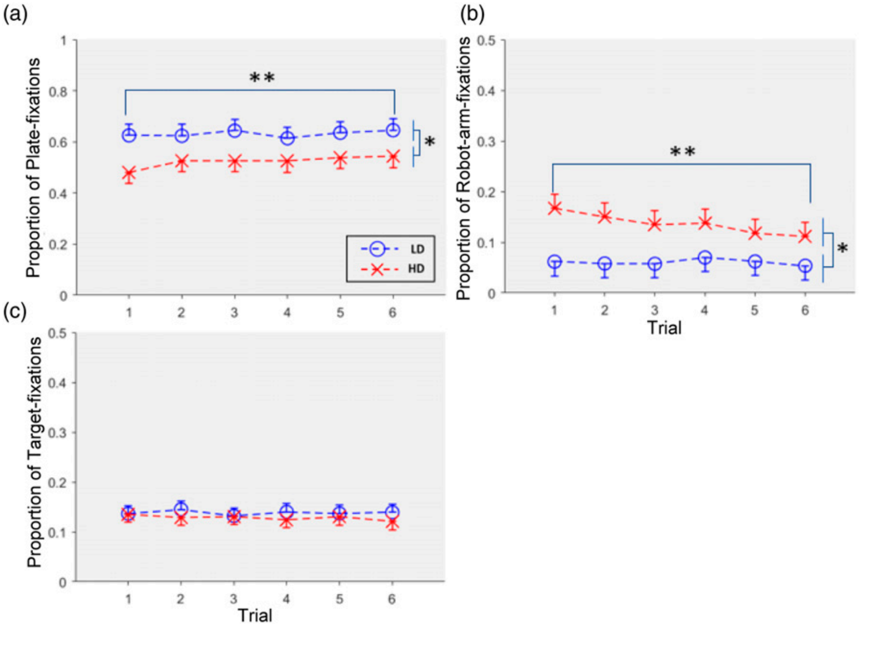

Figure 3. Proportion of fixations on (a) the plate, (b) the robot-arms, and (c) the targets over the course of six learning trials for each condition (LD, HD). Individual data points represent least squares means, and the error bars represent 1 standard error. Note: * denotes a significant main effect of Condition, and ** denotes a significant main effect of Trial. Significant main effects of Condition and Trial were seen in plate-fixations and robot-arm fixations.

TABLE 2: Effect Sizes and P-Values for AOI-Measures

AOI-Measures	Condition	Trial	Condition x Trial
Plate-fixations	**.08 (<.0001)**	**.018 (<.0001)**	**.008 (.004)**
Robot-fixations	**.08 (<.0001)**	**.0195 (<.0001)**	**.014 (<.0001)**
Target-fixations	.00001 (<.0001)	.0024 (<.0001)	.004 (<.0001)

Bold fonts indicate significant effects ($p < .05$).

Performance Prediction Using Eye-Tracking Metrics

The confusion matrices for the two logistic regression models are represented together in Table 3, along with the model performance metrics. Overall, the model with eye-tracking measures as predictors (Model 2) was more sensitive (recall of 70%), specific (71%), and accurate (71%) than the model with performance measures only (Model 1—recall 52%, specificity 67%, and accuracy 60%).

DISCUSSION

This study explored the potential of eye-tracking as a continuous and nonintrusive technique for measuring the mental workload associated with learning to use a bimanual robot. While the psychometric properties of eye-tracking metrics have been explored in cognitive tasks (Matthews et al., 2015), the sensitivity and reliability of eye-tracking metrics have not been systematically investigated in pHRI tasks. Further, no work in the human–robot interaction domain has explored the concurrent effects of task difficulty and time/learning on mental workload, as done here. Our overall results indicated that performance and perceived mental workload were affected by task difficulty but improved over the course of multiple trials (as intended). These changes were similar across males and females; however, females exhibited lower overall performance and reported higher workload compared to males. This performance advantage for males in manipulating the robot may be explained by their average physical strength and dexterity being higher than females (Thomas & French, 1985).

Sensitivity of Eye-Tracking Metrics to Changes in Task Difficulty and Learning

Among the eye-tracking workload measures, stationary gaze entropy (SGE) and pupil diameter (PD) increased with task difficulty and gradually reduced over multiple trials, although PD did not reduce monotonically. A number of past studies have found SGE to increase while monitoring highly variable or complex visual environments (Shiferaw et al., 2019) or more difficult surgical procedures (Wu et al., 2019). In this study, higher SGE in the HD condition may have been due to the need to monitor and track the robot's (more unpredictable) behavior, or to monitor the environment for potential collisions. Higher fixation count (FC) in the HD condition may have been the result of more frequent collision-monitoring. FC did not change significantly over time in the HD condition, suggesting that the visual monitoring demands in the HD condition may have been persistently high throughout the experiment. Interestingly, FC also increased in the last three trials of the LD condition. While this may suggest an increased visual monitoring workload, collisions remained infrequent in these trials and the NASA-TLX ratings did not rise. A possible explanation is that after learning to minimize errors during the first half of the LD condition, participants may have shifted to more frequent visual sampling of the plate's orientation to ensure more precise placement, or tracking the ball, or improving other (secondary) aspects of performance. Fixations have been found to be not only sensitive to workload but also be task-specific and increase due to visual monitoring strategies in prior work (Rivecourt et al., 2008; Zhang et al., 2022), which may be a reason for the dissociation of FC with collisions and NASA-TLX ratings in our study. It should be noted that although all our workload metrics showed significant differences across task conditions and most measures showed an effect of learning trial, all effect sizes were small (Cohen, 1988), with GTE and FC effects likely being too small to be practically important. Pupil dilation had a medium-to-large effect of learning trial ($\eta_p^2 = .085$).

On the other hand, AOI-based measures of gaze behavior had medium-to-large effect sizes for task difficulty, specifically the plate-fixations ($\eta_p^2 = .0797$) and robot-fixations ($\eta_p^2 = .0793$). Further insight into the nature of task demands and the source of workload is provided by these AOI-based measures. It was found that the plate accrued the most fixations of all the task-relevant AOI. This was likely because participants relied on visual feedback of the plate in an effort to maintain their virtual "grip," as well as to prevent the ball from rolling off. In the HD condition, participants seemed to significantly reduce plate-fixations and increase robot-fixations (Figure 3(a) and (c)). Further analysis revealed a significantly higher likelihood of collisions in the HD condition, which might have led participants to look away from the plate to monitor the robot-arms and avoid collisions with targets. Additionally, participants may also have directed their visual attention towards the robot in order to better evaluate, and to develop an internal model of, the robot's control dynamics. This is supported by the observation that some participants tended to move the robot-arms in a random, exploratory manner in the HD condition for some time initially, presumably to better understand their dynamics. Robot-arm fixations reduced significantly over the course of the trials, suggesting that participants reduced their dependence on vision for monitoring the arms, likely as a result of forming an improved internal model. The lack of a significant difference in target-fixations between conditions may be due to a fixed target-monitoring strategy in both conditions, and not devoting additional visuomotor effort towards accurate placement in the HD condition. Regarding gender differences, females tended to fixate more on the robot compared to males in both the LD and HD conditions. Previous research has found females to be more risk-averse (Byrnes et al., 1999), which may have contributed to greater "checking" behavior towards the movement of the robot.

In summary, to detect differences due to task difficulty and learning, measures of gaze behaviors may need to be included in addition to traditional eye-tracking measures of mental workload in future pHRI studies.

Reliability of Eye-Tracking Metrics

The relative reliability of PD (ICC = .68) was highest among all metrics, followed by SGE (.44) and FC (.41), and considered to be good based on criteria stated in previous research on eye-tracking measures (Zargari Marandi et al., 2018). However, it should be noted that PD is generally more reliable in studies involving cognitive (as opposed to motor-control) tasks (Krejtz et al., 2018; Matthews et al., 2015). The SEM value for PD was .16 mm, which is considered acceptable, since mental workload effects on PD typically range from .1 mm to .5 mm from baseline (Pfleging et al., 2016). However, in our study, the highest difference in the least squares means of pupil dilation between each trial in the LD and HD conditions was ~.1 mm (Figure 2(c)). Thus, the SEM of PD may have been too large (indicating a large within-subject variance) for PD to be a reliable measure of workload. The small effect size of condition on PD further supports this conclusion. However, the effect of Trial on PD was significantly larger, indicating that PD may be a better indicator of changes in mental workload that occur over learning, as opposed to those due to task difficulty.

Performance Prediction Using Eye-Tracking Metrics

The model with eye-tracking measures was more sensitive, specific, and accurate, clearly outperforming the model with task variables, for classifying the outcome of each trial to be a success or failure. This is likely because eye-tracking measures continuously capture individual variation in workload and visual behavior, which our task-based independent variables cannot. Much existing work in adaptive robotics, largely in the

TABLE 3: Confusion Matrices for Logistic Regression Models Comparing the Predictive Ability of Manipulated Task Variables (a; Model 1) versus Eye-Tracking Measures (b; Model 2)

	Model 1: Condition, Trial and Condition × Trial (R-squared = .062)				
		Predicted success/failure			
		1	0		
Actual success/failure	1	.516	.484	Recall = .52	
	0	.328	.672	Specificity = .67	
		Precision = .52	NPV = .67	Accuracy = .60	F1 score = .54
	Model 2: Eye-Tracking Measures (R-squared = .304)				
		Predicted success/failure			
		1	0		
Actual success/failure	1	.704	.296	Recall = .70	
	0	.294	.706	Specificity = .71	
		Precision = .68	NPV = .73	Accuracy = .71	F1 score = .69

rehabilitation domain (Brown et al., 2016), tends to adapt robot behavior based solely on task performance. However, in industrial use-cases, adapting robot behavior based on performance may prove to be too late due to the high safety costs of failures. Thus, using eye-tracking workload measures may provide proactive, anticipatory information about mental and visuomotor effort preceding a success or a failure, potentially leading to more timely and accurate adaptation of robot behavior.

LIMITATIONS AND FUTURE WORK

A limitation of our study is the potential effect of physical workload on our results, specifically PD, which has been shown to increase with the perception of physical effort (Zénon et al., 2014). Although we attempted to control for physical workload by offering rest periods after each trial, multiple participants reported wrist strain during the study, which may have affected PD in addition to mental workload. We acknowledge that a real-world comanipulation task may inevitably require a certain degree of physical effort, and that the effectiveness of pupil dilation may be a more effective measure during tasks that largely involve motor-cognitive workload. Future research should consider the extent to which cobots may reduce or eliminate physical workload and define specific use-cases in which pupil dilation may be an effective measure. A second limitation is that this study only recruited college students as participants, thus, limiting the generalizability of our results to other populations. With the rising age of industrial workers, it is expected that older adults will frequently use, and benefit from, physically coupled robots in the future workplace. Thus, future research needs to examine mental workload and strategies associated with pHRI across diverse users. Finally, since our task environment was virtual, there may be some differences in terms of eliciting appropriate gaze behaviors and cognitive responses between the virtual and physical world. Specifically, the absence of haptic feedback from the movement of the ball usually utilized for physically manipulating an object may have increased reliance on visual feedback and monitoring, thus potentially overestimating the increase in plate-fixations in the VR tasks studied here. However, the lack of haptic feedback may also occur in a real-world scenario—a collaborative robot that provides a high degree of physical assistance may still abstract the physical dynamics of the manipulated object from the human user. Thus, an increase in plate-fixations may still be a useful indicator of workload in real-world scenarios, although this needs to be confirmed by future work. Another related consideration is that manipulating a physical object may require additional physical effort and lead to increases in pupil dilation beyond those due to mental workload. Future studies could be conducted in real-world environments or in higher-fidelity VR environments to determine the extent to which these findings can be generalized to real-world robot manipulation scenarios.

CONCLUSION

Overall, this work found that SGE and PD may be usable indicators of mental workload in pHRI, specifically under precision and monitoring demands. Pupil dilation may be a more sensitive indicator of workload changes due to learning, than those due to task difficulty. Visual-attention metrics, specifically those that quantify the number of fixations in different AOIs, may be highly useful and informative in a pHRI context. Interestingly, the sensitivity of these metrics to both task difficulty and learning was comparable to, or higher than our workload metrics. Although AOI-based metrics are not direct measures of workload, they provide important information regarding visual strategies and performance, which are important mediators of workload (Loft et al., 2007; Tsang & Vidulich, 2006). Thus, their validity should be explored further, especially in the pHRI domain where motor-control and internal-model formation will likely be strong contributors to workload.

KEY POINTS

- Physically coupled robots may be controlled to adapt to the operator's mental state as it changes over the course of time due to differences in task difficulty and learning, but there is little work exploring the measures that may be most suitable for continuously tracking operator state/workload in pHRI.
- We propose that eye-tracking measures can be good candidates for measuring workload in pHRI due to their relative unintrusiveness and relevance to motor performance.
- In the current study, we aimed to investigate the sensitivity and reliability of eye-tracking metrics of workload, the use of visual attentional measures for drawing inferences regarding visuomotor strategies, and the ability of these measures to predict performance in pHRI.
- SGE and PD were the most sensitive and reliable eye-tracking measures—SGE was equally sensitive to task difficulty manipulations and learning, and PD was more sensitive to changes over learning. Measures of visual attention were able to quantify changes in visuomotor strategy due to task difficulty, learning, and between genders.
- A combination of eye-tracking measures of workload and visual attention were able to predict performance (success/failure) with a sensitivity of 70% and accuracy of 71%.
- Future work should investigate the usefulness of eye-tracking workload measures in different pHRI task contexts, and in different populations such as older adults. This work should leverage involuntary measures of workload such as pupil dilation, as well as measures of gaze behavior and visual strategies, which are tightly coupled with motor performance.

APPENDIX

APPENDIX A

Eye-Tracking Data Processing

Although the sampling frequency of the eye-tracker was set at 90 Hz, the actual sampling frequency depended on the graphics rendering speed of the computer and was hence variable over time (Llanes-Jurado et al., 2020). Hence, data were uniformly resampled at 90 Hz using a custom MATLAB script. Next, all samples which corresponded to a pupil diameter value of -1 (which meant that the eyes were not tracked for those times) were removed. Any remaining gaps in the data, except those larger than 100 ms were linearly interpolated (Hessels et al., 2017). We classified the gaze samples into fixations using the I-VDT velocity-dispersion algorithm (Komogortsev & Karpov, 2013). Prior to applying the algorithm, we computed the visual angle between successive gaze points using the scalar product of the gaze point vectors. Then, we obtained angular velocity by applying the Savitzky–Golay filter to the visual angle data (Holmqvist et al., 2011). Finally, as specified in the I-VDT algorithm, gaze points with an angular velocity greater than 75 deg/s were classified as saccades, and the remaining data were classified into fixations using a dispersion threshold of 1 deg and a minimum fixation duration of 80 ms (Aronson et al., 2018). For each fixation, we recorded the most frequently-hit AOI as the "fixated-AOI," which resulted in a time series of successive fixated-AOIs for each trial.

Computation of Dependent Measures

To quantify performance, each transfer was first categorized as a *success* if the plate was transferred to the target within 10-s or a *failure* if either a drop or collision occurred, or if the 10-s timer ran out before the plate was transferred. Numbers of successes and failures were computed per 30-s interval and expressed as a proportion of the total number of transfers in each 30-s interval. *Proportions of successes and failures* were used as measures of task performance. *Fixation count* was calculated per 30-s interval during the trial. We computed the *median pupil dilation (PD)* of the left eye from each 30 s interval, by subtracting the baseline pupil diameter from the raw pupil diameter during the 30 s interval. Equations that were used to compute *stationary gaze entropy (SGE)* and *gaze transition entropy (GTE)*, adapted from (Shiferaw et al., 2019), are given below

$$H_s(x) = -\sum_{i=1}^{N} p(i) \times \log_2 p(i) \tag{1}$$

where $H_s(x)$ is the SGE value for a particular time-bin "x" (equal to 30 s), "i" represents the successive AOI-fixations in the interval "x" and "p" is the proportion of fixations on the i^{th} AOI in the interval "x."

$$H_g(x) = -\sum_{i=1}^{N} p(i) \left[\sum_{j=1}^{N} p(i|j) \times \log_2 p(i|j) \right] \tag{2}$$

where $H_g(x)$ is the value of GTE, "p(i)" is the stationary distribution of fixation locations, and "p (i | j)" is the probability of transitioning to AOI "j" given current position of "i." SGE and GTE were both normalized to the maximum theoretical entropy for each 30-s interval, equal to $\log_2 N$ (Shiferaw et al., 2019). The total number of unique VR objects that were fixated during the study comprised the state space for the entropy calculation. Lastly, we also computed the ratio of fixations on the following AOI—the plate (which included the ball), the robot-arms, and the top surface of the target pillars, per 30-second interval, in order to understand the relative attentional focus directed towards these AOI. These AOI were selected because they were directly related to participants' goals and actions, and hence were critical components of the task. Fixations on the table and the floor were coded as "miscellaneous" fixations that were not task-relevant.

APPENDIX B

Computation of Reliability Measures

The total variability of the data set was partitioned according to the mixed effects model (Searle et al., 2009)

$$E_{cstr} = \mu + \beta_c + \gamma_{tr} + \alpha_s + e_{cstr} \tag{3}$$

where E_{cstr} is the measured value of an eye-tracking measure for a specific condition *c* in trial *tr* of subject *s*; μ is the grand mean; β_c is the fixed effect due to condition (task difficulty); γ_{tr} is the fixed effect of the learning trial; α_s is the random effect of subject; and e_{cstr} is the residual. Both random effects (α_s, e_{cstr}) were assumed to be independently and identically distributed, to have zero covariance between any pair of values and to have a mean of zero. The mixed-effects model was resolved using a REML procedure in JMP® Pro (Version 14, SAS Institute Inc., Cary, NC) to estimate the variance between subjects (S^2_{BS}, i.e., the variance of α_s) and within subject (S^2_{WS}, i.e., the variance of e_{cstr}) with 95% confidence intervals. These variance components were then used to calculate reliability metrics.

SEM was calculated as the square-root of the within-subjects variance, and ICC was computed using the following equation (Shrout & Fleiss, 1979)

$$ICC = \frac{S^2_{BS}}{S^2_{BS} + S^2_{WS}} \tag{4}$$

ACKNOWLEDGMENTS

This study was supported by the National Science Foundation (grants CMMI-2024772 and CMMI-2242610). The sponsor was not involved in data analysis, interpretation, or the decision for publication. Address correspondence to Divya Srinivasan, Clemson University, 277A Freeman Hall, Clemson, SC 29631, USA; e-mail: sriniv5@clemson.edu

ORCID iD

Satyajit Upasani https://orcid.org/0000-0002-3328-5968

REFERENCES

Aronson, R. M., & Admoni, H. (2018 June 26-30). Gaze for error detection during human-robot shared manipulation. In *Proceedings of RSS Workshop: Towards a Framework for Joint Action,* Pittsburgh, PA, United States. Retrieved September 27, 2023 from http://packages.personalrobotics.ri.cmu.edu/assets/pubs/fja_rss2018_aronson.pdf

Aronson, R. M., Santini, T., Kübler, T. C., Kasneci, E., Srinivasa, S., & Admoni, H. (2018). *Eye-hand behavior in human-robot shared manipulation* (p. 10).

Becser, N., Sand, T., & Zwart, J.-A. (1998). Reliability of cephalic thermal thresholds in healthy subjects. *Cephalalgia: An International Journal of Headache*, *18*(8), 574–582. https://doi.org/10.1046/j.1468-2982.1998.1808574.x

Bommer, S. C., & Fendley, M. (2018). A theoretical framework for evaluating mental workload resources in human systems design for manufacturing operations. *International Journal of Industrial Ergonomics*, *63*, 7–17. https://doi.org/10.1016/j.ergon.2016.10.007

Brown, D. A., Lee, T. D., Reinkensmeyer, D. J., & Duarte, J. E. (2016). Designing robots that challenge to optimize motor learning. In D. J. Reinkensmeyer & V. Dietz (Eds.), *Neurorehabilitation technology*, (pp. 39–58). Springer International Publishing. https://doi.org/10.1007/978-3-319-28603-7_3

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin*, *125*(3), 367–383. https://doi.org/10.1037/0033-2909.125.3.367

Chadwell, A., Kenney, L., Thies, S., Galpin, A., & Head, J. (2016). The reality of myoelectric prostheses: Understanding what makes these devices difficult for some users to control. *Frontiers in Neurorobotics*, *10*, 7. https://doi.org/10.3389/fnbot.2016.00007

Clay, V., König, P., & König, S. U. (2019). Eye tracking in virtual reality. *Journal of Eye Movement Research*, *12*(1). Article 1.https://doi.org/10.16910/jemr.12.1.3

Cognolato, M., Atzori, M., & Müller, H. (2018). Head-mounted eye gaze tracking devices: An overview of modern devices and recent advances. *Journal of Rehabilitation and Assistive Technologies Engineering*, *5*, 2055668318773991. https://doi.org/10.1177/2055668318773991

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Cornwall, W. (2015). In pursuit of the perfect power suit. *Science*, *350*(6258), 270–273. https://doi.org/10.1126/science.350.6258.270

De Rivecourt, M., Kuperus, M. N., Post, W. J., & Mulder, L. J. M. (2008). Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. *Ergonomics*, *51*(9), 1295–1319. https://doi.org/10.1080/00140130802120267

Ettinger, U., Kumari, V., Crawford, T. J., Davis, R. E., Sharma, T., & Corr, P. J. (2003). Reliability of smooth pursuit, fixation, and saccadic eye movements. *Psychophysiology*, *40*(4), 620–628. https://doi.org/10.1111/1469-8986.00063

Haddadin, S., & Croft, E. (2016). Physical human–robot interaction. In B. Siciliano & O. Khatib (Eds.), *Springer handbook of robotics*, (pp. 1835–1874). Springer International Publishing. https://doi.org/10.1007/978-3-319-32552-1_69

Hessels, R. S., Niehorster, D. C., Kemner, C., & Hooge, I. T. C. (2017). Noise-robust fixation detection in eye movement data: Identification by two-means clustering (I2MC). *Behavior Research Methods*, *49*(5), 1802–1823. https://doi.org/10.3758/s13428-016-0822-1

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Halszka, J., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press. http://lup.lub.lu.se/record/1852359

Just, M. A., Carpenter, P. A., & Miyake, A. (2003). Neuroindices of cognitive workload: Neuroimaging, pupillometric and event-related potential studies of brain work. *Theoretical Issues in Ergonomics Science*, *4*(1–2), 56–88. https://doi.org/10.1080/14639220210159735

Kao, P.-C. (2009). *Principles of motor adaptation when walking with a powered exoskeleton* (p. 120).

Komogortsev, O. V., & Karpov, A. (2013). Automated classification and scoring of smooth pursuit eye movements in the presence of fixations and saccades. *Behavior Research Methods*, *45*(1), 203–215. https://doi.org/10.3758/s13428-012-0234-9

Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C., & Krejtz, I. (2018). Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PLoS One*, *13*(9), Article e0203629. https://doi.org/10.1371/journal.pone.0203629

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. 863. https://doi.org/10.3389/fpsyg.2013.00863

Land, M. F. (2009). Vision, eye movements, and natural behavior. *Visual Neuroscience*, *26*(1), 51–62. https://doi.org/10.1017/S0952523808080899

Law, B., Atkins, M. S., Kirkpatrick, A. E., & Lomax, A. J. (2004, March 22-24). Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment. *Proceedings of the Eye Tracking Research & Applications Symposium on Eye Tracking Research & Applications*, *ETRA'2004*, 41–48. https://doi.org/10.1145/968363.968370

Llanes-Jurado, J., Marín-Morales, J., Guixeres, J., & Alcañiz, M. (2020). Development and calibration of an eye-tracking fixation identification algorithm for immersive virtual reality. *Sensors*, *20*(17), 4956. Article 17.https://doi.org/10.3390/s20174956

Loft, S., Sanderson, P., Neal, A., & Mooij, M. (2007). Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications. *Human Factors*, *49*(3), 376–399. https://doi.org/10.1518/001872007X197017

Marchand, C., De Graaf, J. B., & Jarrassé, N. (2021). Measuring mental workload in assistive wearable devices: A review.

Journal of NeuroEngineering and Rehabilitation, *18*(1), 160. https://doi.org/10.1186/s12984-021-00953-w

Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich, J. (2015). The psychometrics of mental workload: Multiple measures are sensitive but divergent. *Human Factors*, *57*(1), 125–143. https://doi.org/10.1177/0018720814539505

Novak, D., Beyeler, B., Omlin, X., & Riener, R. (2015). Workload estimation in physical human–robot interaction using physiological measurements. *Interacting with Computers*, *27*(6), 616–629. https://doi.org/10.1093/iwc/iwu021

Pfleging, B., Fekety, D. K., Schmidt, A., & Kun, A. L. (2016). A model relating pupil diameter to mental workload and lighting conditions. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 5776–5788). https://doi.org/10.1145/2858036.2858117

Sailer, U., Flanagan, J. R., & Johansson, R. S. (2005). Eye-hand coordination during learning of a novel visuomotor task. *Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *25*(39), 8833–8842. https://doi.org/10.1523/JNEUROSCI.2658-05.2005

Searle, S. R., Casella, G., & McCulloch, C. E. (2009). *Variance components*. John Wiley & Sons.

Shiferaw, B., Downey, L., & Crewther, D. (2019). A review of gaze entropy as a measure of visual scanning efficiency. *Neuroscience & Biobehavioral Reviews*, *96*, 353–366. https://doi.org/10.1016/j.neubiorev.2018.12.007

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Sobuh, M. M. D., Kenney, L. P. J., Galpin, A. J., Thies, S. B., McLaughlin, J., Kulkarni, J., & Kyberd, P. (2014). Visuomotor behaviours when using a myoelectric prosthesis. *Journal of NeuroEngineering and Rehabilitation*, *11*(1), 72. https://doi.org/10.1186/1743-0003-11-72

Srinivasan, D., & Martin, B. J. (2010). Eye–hand coordination of symmetric bimanual reaching tasks: Temporal aspects. *Experimental Brain Research*, *203*(2), 391–405. https://doi.org/10.1007/s00221-010-2241-3

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., & Goodrich, M. (2006). Common metrics for human-robot interaction. In Proceeding of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction - HRI '06 (p. 33). https://doi.org/10.1145/1121241.1121249

Stirling, L., Kelty-Stephen, D., Fineman, R., Jones, M. L. H., Daniel Park, B.-K., Reed, M. P., Parham, J., & Choi, H. J. (2020). Static, dynamic, and cognitive fit of exosystems for the human operator. *Human Factors*, *62*(3), 424–440. https://doi.org/10.1177/0018720819896898

Thomas, J. R., & French, K. E. (1985). Gender differences across age in motor performance: A meta-analysis. *Psychological Bulletin*, *98*(2), 260–282. https://doi.org/10.1037/0033-2909.98.2.260

Tsang, P. S., & Vidulich, M. A. (2006). Mental workload and situation awareness. In *Handbook of human factors and ergonomics* (3rd ed., pp. 243–268). John Wiley & Sons, Inc. https://doi.org/10.1002/0470048204.ch9

Van Acker, B. B., Bombeke, K., Durnez, W., Parmentier, D. D., Mateus, J. C., Biondi, A., Saldien, J., & Vlerick, P. (2020). Mobile pupillometry in manual assembly: A pilot study exploring the wearability and external validity of a renowned mental workload lab measure. *International Journal of Industrial Ergonomics*, *75*, 102891. https://doi.org/10.1016/j.ergon.2019.102891

Webb, G. I. (2010). Model evaluation. In *Encyclopedia of machine learning* (pp. 683–683). Springer.

White, M. M., Zhang, W., Winslow, A. T., Zahabi, M., Zhang, F., Huang, H., & Kaber, D. B. (2017). Usability comparison of conventional direct control versus pattern recognition control of transradial prostheses. *IEEE Transactions on Human-Machine Systems*, *47*(6), 1146–1157. https://doi.org/10.1109/THMS.2017.2759762

White, O., & French, R. M. (2017). Pupil diameter may reflect motor control and learning. *Journal of Motor Behavior*, *49*(2), 141–149. https://doi.org/10.1080/00222895.2016.1161593

Wilson, M., McGrath, J., Vine, S., Brewer, J., Defriend, D., & Masters, R. (2010). Psychomotor control in a virtual laparoscopic surgery training environment: Gaze control parameters differentiate novices from experts. *Surgical Endoscopy*, *24*(10), 2458–2464. https://doi.org/10.1007/s00464-010-0986-1

Wolpert, D. M., Diedrichsen, J., & Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience*, *12*(12), 739–751. Article 12.https://doi.org/10.1038/nrn3112

Wu, C., Cha, J., Sulek, J., Sundaram, C. P., Wachs, J., Proctor, R. W., & Yu, D. (2021). Sensor-based indicators of performance changes between sessions during robotic surgery training. *Applied Ergonomics*, *90*, 103251. https://doi.org/10.1016/j.apergo.2020.103251

Wu, C., Cha, J., Sulek, J., Zhou, T., Sundaram, C. P., Wachs, J., & Yu, D. (2020). Eye-tracking metrics predict perceived workload in robotic surgical skills training. *Human Factors*, *62*(8), 1365–1386. https://doi.org/10.1177/0018720819874544

Zargari Marandi, R., Madeleine, P., Omland, Ø., Vuillerme, N., & Samani, A. (2018). Reliability of oculometrics during a mentally demanding task in young and old adults. *IEEE Access*, *6*, 17500–17517. https://doi.org/10.1109/ACCESS.2018.2819211

Zénon, A., Sidibé, M., & Olivier, E. (2014). Pupil size variations correlate with physical effort perception. *Frontiers in Behavioral Neuroscience*, *8*, 286. https://doi.org/10.3389/fnbeh.2014.00286

Zhang, Y., Hopko, S., Y, A., & Mehta, R. K. (2022). Capturing dynamic trust metrics during shared space human robot collaboration: An eye-tracking approach. *Proceedings of the Human Factors and Ergonomics Society - Annual Meeting*, *66*(1), 536–536. https://doi.org/10.1177/1071181322661296

Zheng, B., Jiang, X., Bednarik, R., & Atkins, M. S. (2021). Action-related eye measures to assess surgical expertise. *BJS Open*, *5*(5), zrab068. https://doi.org/10.1093/bjsopen/zrab068

Zhou, T., Zhu, Q., & Du, J. (2020). Intuitive robot teleoperation for civil engineering operations with virtual reality and deep learning scene reconstruction. *Advanced Engineering Informatics*, *46*, 101170. https://doi.org/10.1016/j.aei.2020.101170

Satyajit Upasani is a PhD student in the Industrial and Systems Engineering Department, Virginia Tech, Blacksburg, VA. He earned an MS in industrial engineering from Virginia Tech in 2017. His research interests include human factors and user experience, with a focus on eye-movement research, mental workload, human–robot interaction, and usability engineering.

Divya Srinivasan is currently the McQueen Quattlebaum professor in industrial engineering at Clemson University, Clemson, SC. Her research interests include human factors, ergonomics, and biomechanics, with a focus on human motor control, learning, performance assessments, and injury prevention. She received MS degrees in biomedical engineering and industrial and operations engineering in 2008 and PhD in biomedical engineering in 2010 from the University of Michigan, Ann Arbor, MI, USA.

Qi Zhu is currently a researcher in the Public Safety Communications Research (PSCR) division within NIST's Communications Technology Laboratory. She received a PhD degree in civil and coastal engineering from the University of Florida and an MS degree in computer science and information engineering from the National Taiwan University. Her research interests include the following topics: 1) Mixed Reality (XR) -based robotic simulation; 2) Human factors engineering for augmented cognitive and physical functions in human–robot collaboration tasks; 3) Human–robot interaction designs for shared sensory perception.

Jing Du is an associate professor in the Department of Civil and Coastal Engineering at the University of Florida, Gainesville, FL. He received his PhD degree in construction engineering from Michigan State University (2012), Master's degree in enterprise management (2007) and Bachelor's degree in civil engineering (2004), both from Tianjin University in China. His current research interests focus on the development of the next generation of Intelligent Information Systems for future Civil Engineering works.

Alexander Leonessa is the Associate Department Head for Strategic Initiatives, and associate professor in the Department of Mechanical Engineering at Virginia Tech, Blacksburg, VA. He obtained a PhD degree in aerospace engineering 1999, and MS degrees in applied mathematics (1997) and aerospace engineering (1999), all from Georgia Tech, Atlanta, GA. His primary research interests are in robotics, and autonomous and dynamical systems.

Date received: February 6, 2023
Date accepted: September 5, 2023