A Multimodal Approach to Investigate the Role of Cognitive Workload and User Interfaces in Human-Robot Collaboration

Apostolos Kalatzis
Montana State University
Bozeman, Montana
apostoloskalatzis@smontana.edu

Vishnunarayan Girishan Prabhu
University of North Carolina
Charlotte, North Carolina
vgirisha@charlotte.edu

Saidur Rahman
Montana State University
Bozeman, Montana
saidurrahman@montana.edu

Laura Stanley
Montana State University
Bozeman, Montana
Laura.stanley@montana.edu

Mike Wittie
Montana State University
Bozeman, Montana
mike.wittie@montana.edu

ABSTRACT

One of the primary aims of Industry 5.0 is to refine the interaction between humans, machines, and robots by developing human-centered design solutions to enhance Human-Robot Collaboration, performance, trust, and safety. This research investigated how deploying a user interface utilizing a 2-D and 3-D display affects participants' cognitive effort, task performance, trust, and situational awareness while performing a collaborative task using a robot. The study used a within-subject design where fifteen participants were subjected to three conditions: no interface, display User Interface, and mixed reality User Interface where vision assistance was provided. Participants performed a pick-and-place task with a robot in each condition under two levels of cognitive workload (i.e., high and low). The cognitive workload was measured using subjective (i.e., NASA TLX) and objective measures (i.e., heart rate variability). Additionally, task performance, situation awareness, and trust when using these interfaces were measured to understand the impact of different user interfaces during a Human-Robot Collaboration task. Findings from this study indicated that cognitive workload and user interfaces impacted task performance, where a significant decrease in efficiency and accuracy was observed while using the mixed reality interface. Additionally, irrespective of the three conditions, all participants perceived the task as more cognitively demanding during the high cognitive workload session. However, no significant differences across the interfaces were observed. Finally, cognitive workload impacted situational awareness and trust, where lower levels were reported in the high cognitive workload session, and the lowest levels were observed under the mixed reality user interface condition.

KEYWORDS

Human-Robot Collaboration, user interfaces, cognitive workload, mixed reality

1 INTRODUCTION

Industry 5.0 is the latest wave of the industrial revolution that goes beyond efficiency and productivity and focuses on improving how humans can work and collaborate with advanced technology, including robots, to improve collaboration, performance, and safety, with workers as the center [48, 83]. The advancement of technologies can play a significant role in supporting and enhancing Human-Robot Collaboration (HRC). Among various areas where these advancements could significantly impact one is the user interface, which acts as the communication gateway between users and robots [59, 76]. Prior studies have reported that an ideal user interface can improve task performance and productivity and reduce the short-term memory load of the users [57, 66]. User interfaces (UI) range from simple 2-D displays to complex state-of-the-art 3-D displays developed in recent years (e.g., virtual reality, mixed reality, etc.). While the capabilities of each interface can be significantly different, a UI's core characteristics and features should aim to achieve consistency, feedback, support internal locus of control, reduce short-term memory load, etc., to develop a useful user interface [29, 66]. Studies have investigated the impact of various user interfaces on training programs, teaching, as a distraction method, etc. Most of these studies reported virtual reality (VR), mixed reality (MR), and extended reality (XR) interfaces as more realistic, immersive, assistive, and efficacious than a traditional 2-D display or no user interface [7, 31, 61]. However, it should be noted that these studies have also reported the VR/MR/XR technologies to be comparatively more cumbersome and, in some instances, induce a higher cognitive load on the users while completing a task[16, 19, 42].

The cognitive load associated with a task can be defined as the relative demand imposed by a particular task in terms of the mental resources required [70]. This is a critical factor when considering HRC, as prior studies have reported this to impact the collaboration's effectiveness and efficiency significantly[10, 27, 47]. Researchers have noted that if a robot is too complex or its interface is unclear, the cognitive load on the human collaborator may be too high, resulting in decreased performance and increased error rates[41]. Conversely, if the robot is too simple or its interface is too limited, the cognitive load on the human may be too low, leading to disengagement and reduced engagement with the task[41]. Therefore, it is important to carefully design and optimize the robot interface and interaction to manage the cognitive load of HRC. As briefly mentioned earlier, interfaces act as the communication gateway between users and robots; hence interfaces should aim to bridge the gap between the gulf of evaluation and the gulf and execution, i.e., understanding the state of the system and executing an action to accomplish a specific goal. In HRC, three critical factors affect the gulf of evaluation and execution: (i) short-term memory load and cognitive load that the interface asserts on humans, (ii) situational awareness, and (iii) human trust in the system and interface[38, 65, 71]. Studies investigating the use of XR technologies for complex HRC have observed mixed findings, where few observed that the users experience a higher cognitive workload and vice versa[3, 13, 33, 50, 52, 53, 59]. The cognitive workload cannot be directly attributed to the XR technology as various factors, including task complexity, task duration, interface design, device, etc., play a significant role [6, 18, 84]. Primary reasons attributed to an increased cognitive workload while using XR technology for HRC include cumbersome technology, time of use, the field of view, lack of feedback, delay in processing feedback, task complexity, etc. [12, 62]. Hence, researchers should consider these factors and focus on improving the interface design and other parameters (task complexity, time of use) while using XR headsets for HRC. When discussing the role of cognitive load on HRC and how it affects task performance and productivity, it is critical to focus on situational awareness (SA). Situational awareness is the ability to perceive and comprehend what is happening in the environment around you at any given time or space and predict what can happen in the future[25].

Cognitive load significantly impacts an individual's ability to maintain situational awareness, and prior studies have reported that when a person experiences a high cognitive load, their attentional resources are limited, and they may not be able to process all the information available to them fully [67, 72]. This can lead to decreased situational awareness and an increased risk of errors or accidents. Researchers have noted that a lack of situational awareness often leads to higher chances of errors, failures, and fatalities in complex systems such as aviation, manufacturing, and mining [22, 24, 55, 68]. Therefore, it is important to manage cognitive load appropriately to maintain situational awareness in high-stress situations. While opportunities for lapses in SA are common in complex environments, their likelihood in environments involving higher automation, specifically in HRC along with assistive technology, is not well-researched [23, 54]. The consensus is that increasing automation and providing assistive technology during HRC can offload the burden on the user and reduce their cognitive load leading to a better SA.

For a user to utilize any assistive technology during a complex task, including HRC, it is imperative that the user trusts the assistive tool [45, 64]. Trust is a critical factor influencing the adoption of assistive technologies, as users rely on tools they believe will perform as expected[2, 5]. Specifically, the literature on trust in HRC related to assistive technology, such as user interfaces, highlights the importance of trust in improving performance and productivity [58]. Additionally, factors including transparency, predictability, and reliability play a significant role in building trust in user interfaces [43, 56]. Further, the design of the interface, including its appearance and functionality, can impact the user's level of trust[20, 40]. Trust also affects the user's performance and productivity, as higher levels of trust lead to increased confidence in the technology, which results in more efficient and effective use[63, 87]. Furthermore, trust can be affected by the robot's behavior and communication and how it responds to the user's actions and requests[78, 81, 85]. Therefore, building trust in assistive technology is critical for achieving optimal performance and productivity in HRC. Specifically, prior studies using mixed reality interfaces reported it to enhance trust and task performance by providing users with additional information, sensory feedback, and natural representation [58, 74]. While these findings are promising, the impact of trust on productivity is still an open question and further research is needed to determine the optimal levels of trust necessary for maximizing productivity in HRC and mixed reality environments.

To investigate these research gaps and understand the impact on users' cognitive load, task performance, trust, and situational awareness while performing a collaborative pick-and-place task with various levels of cognitive demand, we conducted a systematic and comprehensive human-subject study. This study design allowed us to investigate highly relevant and interrelated human factors considerations in HRC. As such, this study employs systematic empirical manipulation of cognitive workload and the use of an assistive interface to collect multimodal responses that include multiple-task performance metrics, subjective perceptions, and physiological responses to understand their interrelations and impact on HRC. We hypothesize that MR assistive interface for HRC will improve task performance and situational awareness and reduce the cognitive workload during a human-robot collaborative pick-and-place task. Furthermore, we hypothesize MR interface will be more efficacious among the assistive displays compared to a 2-D display.

2 RELATED WORK

2.1 Human Physiology, Task Performance, and Cognitive Workload

Mixed reality has immense potential as assistive technology in HRC by providing users with helpful information when needed, and previous research has shown that guidance and holographic information can lead to better performance results and augment the HRC [3, 13, 33, 50, 52, 53, 59]. However, the effect of MR on cognitive workload and task performance is unclear as researchers have mixed findings of using AR to present information to a user. Some suggest that AR positively affects cognitive workload and task performance [14, 44, 69, 75]. Whereas some show evidence of cognitive overload using MR. Although the findings from these studies offer important insights, one primary limitation is that all these studies relied solely on subjective feedback (i.e., NASA TLX, NASA RTLX, surgical TLX) to measure cognitive workload. While subjective perceptions of cognitive workload are valuable, it also leads to many challenges since they cannot be objectively compared. Moreover, in most studies, operators often report their cognitive workload at the end of the task, which fails to capture the cognitive load in real time while performing a task.

Physiological parameters can serve as reliable, objective indicators of cognitive workload [8, 11, 32]. These can be represented as a function of the autonomic nervous system (ANS). The ANS consists of two major systems: the sympathetic and parasympathetic systems [49, 60]. The most commonly assessed indices of ANS are based on cardiovascular activity, specifically, the frequency domain metric of HRV. The low frequency (LF) band is considered a reflection of both sympathetic and parasympathetic activities with vagal modulation, while the high frequency (HF) band is regarded as an indicator of parasympathetic activity. The LF/HF ratio is commonly used to represent the sympathovagal balance [17, 26, 35, 80]. Higher LF/HF values indicate sympathetic activity dominance, while lower LF/HF values suggest a shift towards parasympathetic activity dominance [17, 26, 35, 80]. While investigating HRC, the cognitive workload can be recognized using LF/HF ratio by observing differences in cognitive state. However, it should be noted that there is scientifically no way to distinguish which factor among arousal/stress/cognitive load contributes to the LF/HF ratio because of the overlap (similarities) of these factors in how it affects human physiology.

Additionally, the monitoring of cognitive workload allows for the measurement of the cognitive cost associated with tasks and the prediction of operator performance. Prior research has shown that both underloading and overloading the brain may have a detrimental effect on human performance [4, 30, 36, 51, 86]. According to Wickens' multiple resource theory, humans have a limited amount of cognitive resources to dedicate to a task, and when the task at hand exceeds the available resources, it leads to inefficiency and deteriorates performance [79]. On the other hand, when a task requires fewer resources, it can distract users from the primary task leading to a lack of vigilance and resulting in a subpar performance [86]. Objective responses, rooted in established theories like Wickens' multiple resource theory, offer quantifiable indicators of cognitive resource allocation. In dynamically changing conditions such as HRC, these objective performance measures serve as continuous and reliable markers to ensure optimal task distribution, preventing both underloading and overloading scenarios that can compromise efficiency and safety. By systematically monitoring cognitive workload and correlating it with objective performance measures, such as task completion time and accuracy, the study provides a comprehensive and data-driven insight into the intricate interplay between human cognition and robot collaboration.

2.2 Trust

Developing intuitive user interfaces that facilitate natural and efficient interactions between humans and robots is vital for effective task completion, performance, and cognitive load management [28, 76]. Trust in a user interface is important in ensuring that humans can effectively collaborate with robots [9, 73]. Trust is based on the perceived usability, dependability, and safety of the interface, and it is crucial to the success of HRC. Therefore, there is a need to create user interfaces that not only improve task performance and cognitive workload management but also inspire trust and confidence in HRC. Previous research has shown a strong correlation between the amount of trust in human-robot teammates and their performance, as well as an effect on the quality of their interactions [15, 46]. High levels of trust between human-robot can lead to increased collaboration and better performance and decision-making [82]. On the other hand, a low degree of trust among human-robot teammates can lead to decreased efficiency and can negatively affect the user's perception of the robot, resulting in negative attitudes towards the robot and decreased willingness to collaborate [82]. Although these findings provide insights regarding trust and performance in HRC, there are no current studies that explore trust in using a user interface as an assistive tool to complete a human-robot task. However, there is evidence that a user interface that is intuitive and user-friendly can provide clear and concise instructions and therefore ensure effective HRC. In this study, we will investigate how the user interface influences the operator's trust under cognitive workload levels in order to provide insight into more robust HRC designs.

2.3 Situation Awareness

Another critical factor in HRC is situational awareness, as it enables humans and robots to make informed decisions and adapt to changing circumstances [25]. Developing an effective user interface is pivotal in supporting situation awareness in HRC [67, 72]. However, the effectiveness of user interfaces in supporting situation awareness during HRC depends on several factors, including the complexity of the task, the cognitive workload, and the level of trust between humans and robots. Cognitive load and trust significantly impact an individual's ability to maintain situational awareness [67, 72]. Specifically, a high cognitive load and low trust would lead to low situational awareness as their attentional resources are limited [21]. Additionally, in cases where humans are required to perform multiple tasks simultaneously or when tasks are highly complex, their cognitive workload increases, leading to a reduced capacity to monitor and comprehend the situation [22, 24, 55, 68]. However, other studies suggest that effective UIs can reduce cognitive workload and increase trust, situational awareness, and task performance [14, 44, 69, 75]. For instance, using mixed reality interfaces can provide humans with a more intuitive and immersive way to interact with robots and their surroundings, improving their situational awareness. Therefore it is essential to understand the impact of the user interface on situation awareness and the interconnections with the cognitive workload in order to create more efficient and safer HRC designs.

3 METHODS

3.1 Participants

The participants in this study consisted of fifteen college students (7 male and 8 female), and their age distribution of 26.20 ± 6.41 years. Every individual who took part in the study was required to sign a consent form, and their participation was contingent on adhering to the guidelines established by the University's Institutional Review Board and referenced by the IRBNO: AK090821. The participants reported their experience with industrial robots and joysticks. None of the participants reported prior experience with industrial robots. Additionally, the average participant reported a slight familiarity with joystick devices which were utilized by the participant to control the robot. There were no dropouts in this study, and participants were compensated with a $25 Amazon gift card for their time.

3.2 Collaborative Task

In this study, a human operator used a joystick to control a UR3e (Universal Robots, DK) collaborative robot (cobot) to pick up objects from a surface and place them in a bin (Figure 1). The robot's speed was set using a scaling factor that regulates the linear speed of the end-effector. During this task, the participants navigated the robot gripper to pick the objects (see Figure 1) and drop these into a bin. Objects of different colors and shapes were used and the participants received instructions regarding the object that they will have to pick up from an external screen before each trial.

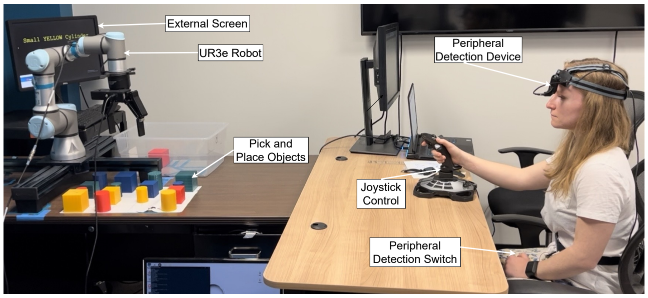

Figure 1: Experimental setup of HRC with the UR3 robot for a pick-and-place task.

3.3 Procedure

This study included two major independent variables – cognitive workload (i.e., low, high), and user interface (i.e., no interface, display, mixed reality – as well as three classes of dependent variables – task performance, subjective responses, and physiological responses. The participants attended two sessions, where each session focused on one level of the cognitive workload variable (i.e., low, high, Figure 2). At the beginning of each session we collected the baseline data, which included physiological data, to ensure that across the different days of the experiment, the participants started with the same baseline. Within each session the low cognitive workload (LCW) and high cognitive workload (HCW), the participants underwent three sub-sessions where they performed the task with and without visual assistance provided through a user interface. In the first sub-session, the participants performed the task without using a user interface (no UI). In the next sub-session participants were assisted during the task using a UI that provided location information to an external screen that shows a top-down camera view of the objects (display UI Figure 3). The UI was specifically designed to offer visual assistance for pick-and-place tasks, aiding participants in locating objects and achieving optimal gripping. A single display, positioned to the right of the user was utilized for this purpose. To ensure ergonomic comfort, the screen providing vision assistance is positioned at the height of the operator rather than on top of the robot. This placement allows for optimal viewing and interaction, enhancing the user's experience. Finally, in the last sub-session, the participants performed the task wearing the MR headset (Microsoft Hololens 2) that shows a top-down camera view of the object (mixed reality UI, Figure 3). Cognitive workload and assistance were counterbalanced and all participants performed all experimental conditions. Counterbalancing was used to address the effects of task experience and the order of cognitive workload states. We adopted a Latin square design approach to ensure that each condition occurs equally often in each possible position throughout the experiment. The participants completed 16 trials of the task in each sub-session.

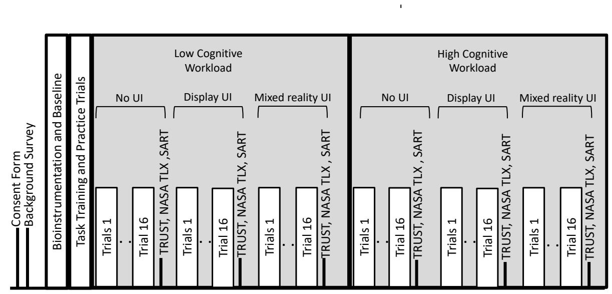

Figure 2: Study Protocol.

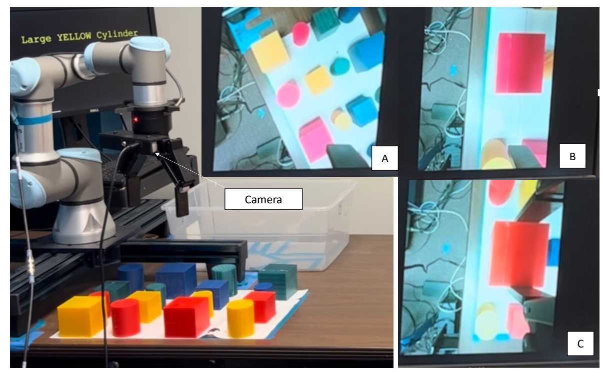

Figure 3: User Interface in 2-D or 3-D Display. View of the camera from the home location of the robot (A). Closer view before the pick up of the object (B), and after pick up (C).

3.4 Cognitive Workload Manipulation

The cognitive workload was manipulated by adding visual secondary tasks. The participants were asked to perform two tasks simultaneously. The primary task was the task of interest (bin picking), and a secondary task is performed alongside the primary task. For the secondary task, We utilized the Peripheral Detection Task (PDT) (Van Winsum et al., 1999) where a LED light, mounted on a headband in front of the participant, was shown every three to five seconds. The participants had a small switch attached to their index finger, which they had to press every time they saw the LED light. If a reaction was not detected within two seconds from the onset of the stimulus, this was coded as a missed signal (after two seconds the light turned off).

3.5 Dependent Measurements

3.5.1 Primary Task Performance. Trajectories of each individual pick-and-place object were recorded at a frequency of 100 Hz. The data provide the precise location of the six angles of the robot. At the end of each trial, this information is saved in a .csv format, resulting in one .csv file per trial. Task performance was quantified using two different measurements: travel time (time in seconds to pick-and-place the objects in each trial) and travel distance (distance in meters covered from the home location of the robot until the dropping point). All the participants completed the pick-and-place task successfully. Hence there were no binary errors.

3.5.2 Secondary Task Performance. Event detection and reaction times were recorded during the secondary task. In each trial, we calculated the number of detected signals and the reaction time of each detected signal (time to respond to stimuli in seconds).

3.5.3 Subjective Responses. After each session, the participants completed the situation awareness rating technique (SART) to measure SA and the NASA task load index questionnaire (NASA TLX) to quantify workload. SART consists of three subscales: attentional supply, attentional demand, and understanding of the task. The SART composite score was calculated as understanding (demand-supply). The NASA TLX is a multidimensional assessment tool consisting of six subjective subscales from very high to very low in terms of their: mental demand, physical demand, temporal demand, effort, performance, and frustration. The questionnaire was further analyzed by its subscales and composite scores. The overall workload is calculated by averaging the score of the six subscales. At the start of the experiment, participants were asked to complete the propensity to trust questionnaire (Jessup et al., 2019) to capture their tendency to trust in general. The propensity to trust questionnaire consists of six questions rated on a 1 (low) to 7 (high) scale. The composite score is calculated as an average of all questions (after inverting negatively framed questions). At the end of each assistance sub-session (i.e., Display, Hololens), the participants completed the trust in automation questionnaire (TRUST) by Jian et al. (2000) adapted to Display and Hololens. The TRUST survey quantifies active trust perceptions based on the condition. The TRUST questionnaire consists of 12 questions rated on a scale from 1 (low) to 7 (high), with the composite score calculated as an average of all questions (after inverting negatively framed questions).

3.5.4 Heart Rate Variability Response. HRV responses were collected using the clinically validated Biopac® MP160 system. The ECG and respiration signals were collected at a sampling rate of 1000 Hz. Post-data collection, the first step was to remove the artifacts from the physiological data. To remove the artifacts from the physiological data, we first visualized the HRV data, any ectopic beats or motion artifacts were interpolated using Biopac AcqKnowledge. In this study, we applied Wavelet Transform to scale the decomposed ECG signal into different frequency band signals. After eliminating the noises, the ECG signal was reconstructed with the original signal's useful parts [1].

3.6 Statistical Analysis

Statistical significance was determined through repeated measures analysis of variance (RM ANOVA) tests on dependent responses, with significance reported at $\alpha = 0.05$ and marginal significance at $0.05 < \alpha < 0.1$. The normality of the data was determined using a Q-Q plot. Equality of variances was verified using Levene's test and using boxplots. Independence was assured by our counterbalanced design. The RM ANOVAs were separately run on each task performance metric to test the effects of cognitive workload (i.e., low, high) and user interface (i.e., no UI, display UI, mixed reality UI). Separate RM ANOVAs were performed on all subjective responses to test the effects of cognitive workload (i.e., low, high) and user interface (i.e., no UI, display UI, mixed reality UI). Furthermore, separate RM ANOVAs were run on all secondary task performance metrics to test the effects of cognitive workload (i.e., low, high) and user interface (i.e., no UI, display UI, mixed reality UI). Finally, a non-parametric test (Friedman's test) was performed on LF/HF ratio to test the effects of cognitive workload (i.e., low, high) and user interface (i.e., no UI, display UI, mixed reality UI). Post hoc comparisons were performed where needed using Tukey HSD.

4 RESULTS

4.1 Primary Task Performance Metrics

4.1.1 Travel Time. The overall travel time of the pick-and-place task was significantly impacted by cognitive workload ($p = 0.01$, $\eta^2 = 0.05$, Figure 4), where participants had higher travel times in HCW at (32.03 ± 16.33) compared to LCW at (30.03 ± 13.21). The user interface significantly impacted the travel time ($p < 0.01$, $\eta^2 = 0.09$), where overall time was faster without an interface (29.65 ± 15.38) compared to display UI (30.75 ± 13.46) and MR UI (32.97.64 ± 15.30). There was an interaction effect between cognitive workload and interface ($p = 0.03$, $\eta^2 = 0.03$). Travel times were faster in low cognitive workload and no interface (27.15 ± 10.83) session compared to high cognitive workload and the use of interfaces in both high and low cognitive workload.

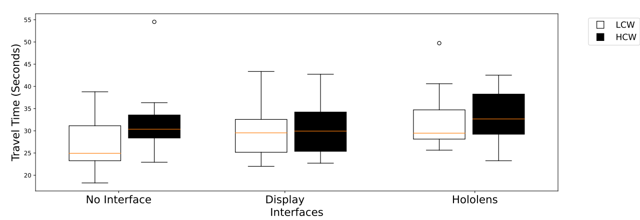

Figure 4: Effects of cognitive workload and User Interface on travel time.

4.1.2 Travel Distance. Two-way marginal interaction was observed between cognitive workload and interface ($p = 0.03$, $\eta^2 = 0.04$, Figure 5). Travel distance was higher in high cognitive workload when using the mixed reality UI compared to display UI and no UI in both low and high cognitive workload; however, no pairwise t-tests were significant after Tukey HSD Method.

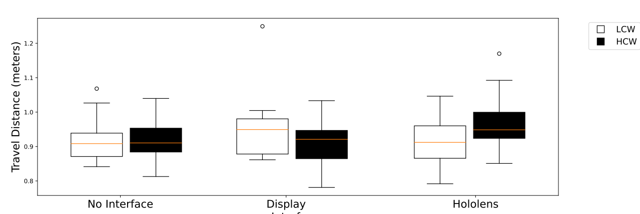

Figure 5: Effects of cognitive workload and User Interface on travel distance.

4.2 Secondary Task Performance

4.2.1 Missed Signals. There was no statistically significant difference (all $p > 0.05$) between the no UI, display UI, and mixed reality UI. However, we observed a slight decrease in missed signals when comparing the no UI (71.90 ± 44.60) with display UI (64.26 ± 42.61) and mixed reality UI (65.82 ± 38.81).

4.2.2 Response Time. All main effects and interactions were statistically identical (all $p > 0.05$) across the interfaces. There was a small increase in the response time when comparing the no UI (0.69 ± 0.12) with display UI (0.70 ± 0.13) and mixed reality UI (0.72 ± 0.10).

4.3 Physiological Data - LF/HF Ratio

All main effects and interactions were statistically identical (all $p > 0.05$). However, we observed a slight increase in LF/HF ratio during the HCW session (5.93 ± 5.71) compared to LCW (6.10 ± 6.79), which is an indicator of increased cognitive workload.

4.4 Subjective Responses- NASA-TLX

4.4.1 Composite Score. Cognitive workload affected overall workload score ($p < 0.01$, $\eta^2 = 0.07$) where LCW (24.35 ± 13.23) was rated with lower overall scores than HCW (45.69 ± 20.09, Figure 6). No other effects were observed (all $p > 0.466$).

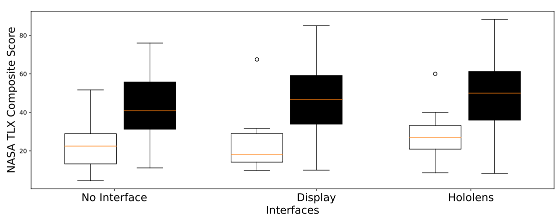

Figure 6: Effects of cognitive workload and User Interface on overall cognitive workload perception.

4.4.2 Mental Demand. Cognitive workload significantly affected the mental demand subscale ($p < 0.01$,$\eta^2 = 0.11$) with lower mental demand in LCW (30.71 ± 17.70) than in HCW (62.24 ± 24.11).

4.4.3 Physical Demand. Cognitive workload significantly affected the physical demand subscale ($p = 0.01$, $\eta^2 = 0.09$) with higher physical demand in HCW (32.53 ± 26.67) than LCW (20.22 ± 19.25).

4.4.4 Temporal Demand. There was a significant effect of cognitive workload on temporal demand ($p = <0.01$, $\eta^2 = 0.12$) where high cognitive workload resulted in higher temporal demand in HCW (49.87 ± 26.30) than LCW (22.42 ± 12.08).

4.4.5 Performance. Cognitive workload ($p = 0.01$, $\eta^2 = 0.07$) had a significant effect on perceived performance, where participants felt they performed better in LCW (41.37 ± 25.25) than in HCW (33.04 ± 31.35).

4.4.6 Effort. The cognitive workload had a significant effect ($p <0.01$, $\eta^2 = 0.06$), with more effort required for HCW at (52.80 ± 26.37) than low at (26.16 ± 16.88).

4.4.7 Frustration. There was a significant effect of cognitive workload on frustration ($p = <0.01$, $\eta^2 = 0.08$), where high cognitive workload resulted in more frustration in HCW (35.29 ± 26.65) than LCW (13.42 ± 12.24).

4.5 Subjective Responses - Situational Awareness

4.5.1 Composite Score. Cognitive workload ($p < 0.01$, $\eta^2 = 0.04$) influenced the composite score for SART where lower situation awareness was associated with HCW (13.83 ± 4.36) than LCW (18.02 ± 3.96, Figure 7). No other effects were observed (all $p > 0.586$).

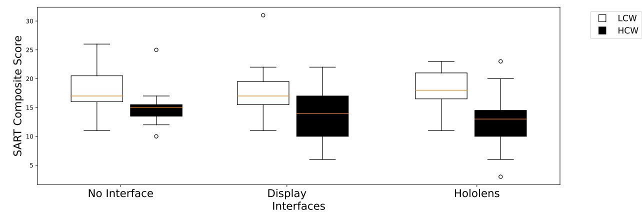

Figure 7: Effects of cognitive Workload and User Interface on Situational Awareness.

4.5.2 Attention Demand Subscale. There was a significant effect of cognitive workload on attention demand ($p = <0.01$, $\eta^2 = 0.03$), where high cognitive workload resulted in more attention demand in HCW (10.91 ± 4.19) than LCW (6.43 ± 2.11).

4.5.3 Attention Supply Subscale. There was a marginal effect of cognitive workload on attention demand ($p = 0.07$, $\eta^2 = 0.06$), where high cognitive workload resulted in more supply demand in HCW (16.78 ± 3.13) than LCW (15.57 ± 3.49).

4.5.4 Understanding Subscale. Cognitive workload ($p =0.09$, $\eta^2 = 0.03$) influenced the understanding, where lower understanding was observed with HCW (7.96 ± 2.47) than LCW (8.89 ± 2.21).

4.6 Subjective - Trust

All main effects and interactions were statistically identical (all $p > 0.218$, $\eta^2 = 0.01$). However, we observed a decrease in trust during the HCW (4.28 ± 1.18) session using both interfaces compared to LCW (4.65 ± 1.07), with the lowest decrease to be noticed using the mixed reality UI (3.96) during the HCW session.

5 DISCUSSION

Three key takeaways of this study are:

(1) High cognitive workload and User Interfaces decreased task efficiency (i.e., travel time) and accuracy (i.e., travel distance).
(2) Cognitive workload impacted situational awareness where participants perceived more attention- and supply-demand during the HCW session. The effects of cognitive workload were captured at all the subscales of NASA TLX, where participants perceived the task as more difficult during the HCW session.
(3) Cognitive workload impacted physiological data where an increase in LF/HF ratio was observed during the HCW.

5.1 High Cognitive Workload and User Interfaces decrease Task Performance

In this study, we observed that the high cognitive workload task required higher travel times to complete the task compared to the low cognitive workload task. This is in line with literature, which has noted that task demand dictates the overall impact on the performance of a human operator in a collaborative task [4, 30, 36, 51, 86]. Additionally, we observed that the user interfaces, especially the mixed reality UI, had a negative impact on task performance during high cognitive workload tasks where both travel distance and travel time were highest compared to all other scenarios. This finding suggests that UI is also a significant factor (positive/negative impact) influencing performance along with cognitive workload. Specifically, the lowest traveled time was observed in the low cognitive workload state without UI assistance, and the highest travel time was observed in the high cognitive workload state with mixed reality UI. The operator had no control over the robot's velocity as the robot moved in response to stepwise binary (on/off) joystick inputs. Hence, the change in speed directly indicates the continuity of joystick inputs as influenced by human behavior. As a result, high levels of cognitive and the use of a user interface increased the accumulation of stuttering over relatively longer trials. The main effect was also found in travel distance, where longer trials were observed in the high cognitive workload session while using mixed reality UI. Further, it was observed that the mixed reality UI's assistance in the high cognitive workload state led to the longest travel distances suggesting a not optimal trajectory path because of task difficulty and the inherent complexity of mixed reality technology.

This negative impact on both task performance aligns with the findings of previous research where the cognitive workload is a significant factor influencing human performance [4, 30, 36, 51, 86].Additionally, prior studies have shown a negative impact on task performance while using user interfaces in HRC [4, 30, 36, 51, 86]. Research has shown that user interfaces can increase the cognitive workload required to complete tasks, leading to decreased task efficiency and accuracy [4, 30, 36, 51, 86].

Therefore, it is important for robot operators who complete tasks that require a high level of mental effort to receive support to mitigate the negative effects of cognitive workload on their overall performance. Additionally, it is vital for designers to keep these factors in mind when creating interfaces to ensure that users can complete tasks quickly and accurately without becoming mentally fatigued. Finally, it should be noted that the 2-D display interfaces allowed for better or comparable task performance during high and low cognitive tasks compared to no interface scenario suggesting that future MR systems that are less cumbersome and cognitively demanding could improve HRC.

5.2 Effects of Cognitive Workload and User Interfaces on Subjective Responses

5.2.1 Impact of Cognitive Workload and User Interfaces on cognitive workload perceptions in HRC. The results collected using NASA TLX indicate that cognitive workload had a significant effect on perceived workload across all user interfaces, with a higher workload reported during the high cognitive load task. However, it is important to notice that irrespective of the cognitive demand of the task (High vs. Low), participants consistently reported the highest NASA TLX scores while using the mixed reality UI. This clearly indicates that the participants perceived the mixed reality UI to add additional workload, despite anecdotally it being the most preferred interface. Further, we observed that responses to all NASA TLX subscales showed significant differences between the groups (No interface, displays UI, and mixed reality UI) except for performance, where participants always rated their perceived performance higher. In high cognitive workload sessions, participants reported that the mixed reality UI would provide the best performance, despite the fact that the actual task performance decreased during the mixed reality UI sessions. This suggests that perceived workload and actual performance may not always align in high cognitive workload scenarios. In low cognitive workload scenarios, participants believed that the display UI would lead to the best performance, again without alignment with actual performance metrics. However, there were no significant differences between the no UI and display UI groups in low cognitive workload sessions, indicating that adding a user interface did not significantly impact task performance in low cognitive workload scenarios. The study highlights the importance of considering both subjective and objective measures to evaluate task performance during HRC, irrespective of the cognitive load associated with the task. Our findings in the high cognitive demand task clearly indicate that while participants may perceive certain interfaces as less demanding and enhance their performance, this may not always align with actual task performance. As such, designers and researchers must consider perceived workload and objective performance metrics when evaluating user interface designs in high cognitive workload scenarios.

5.2.2 The impact of cognitive workload and User Interface in Situational Awareness in HRC. In HRC, situational awareness is a critical component for effective communication and decision-making. In this research, we observed that participants reported a lower SA during the high cognitive workload session, irrespective of whether an interface was used. This observation aligns and adds to the current body of work where studies have reported high cognitive workload can lead to increased attention and reduced SA [77]. Additionally, among the high cognitive workload session, we observed that the lowest SA was reported when using the mixed reality UI in a high cognitive load setting. This suggests that high cognitive workload scenarios with immersive interfaces may negatively impact SA in HRC scenarios. However, this observation could also be because of the participant's unfamiliarity with MR devices, thus demanding more attention from users than displays UIs and no interface scenarios. Future studies should consider involving participants with prior MR experience or pursue a longitudinal study to identify the root cause of this observation. Responses from the SART subscales showed that the supply and understanding subscales did not show significant differences between interfaces and no interfaces group, indicating that participants perceived the interfaces to provide adequate information and could understand the task requirements. However, participants reported requiring higher attention demand in the mixed reality UI. Again, this observation can be because of the participants' lack of experience with MR systems calling for further investigation in this direction. Irrespective of the root cause driving the observation, the results indicate that it is important to consider the context of the task, the interface design, and the impact of cognitive workload on SA while using an assistive tool in HRC.

5.2.3 Trust in User Interfaces under low and high cognitive workload. Trust is a critical factor that affects task performance and decision-making in complex environments, including HRC, which is strongly influenced by cognitive load [9, 73]. In this study, we observed that participants reported lower levels of trust under a high cognitive load, with the lowest trust scores when using the mixed reality UI in the high cognitive workload state. This is consistent with prior research demonstrating that a high cognitive workload can reduce trust and reliance on automation in complicated systems [82]. Due to their high cognitive demands, immersive interfaces, such as mixed reality, may negatively influence trust under these circumstances. It is critical to emphasize the influence of cognitive workload on trust in complex systems, particularly in HRC settings. The findings of this study show that high cognitive workload situations, especially those employing immersive interfaces, may negatively influence trust during HRC. Hence, it is essential for designers and researchers to consider the impact of cognitive workload on trust to mitigate the detrimental impact it may have on trust during HRC scenarios when a user interface is utilized.

5.3 LF/HF ratio increased during the high cognitive workload session

We collected the physiological data of participants during an HRC task and observed an increase in the LF/HF ratio, suggesting a higher cognitive workload among the participants [17, 26, 35, 80]. However, we did not observe significant differences between the two sessions or across different interfaces. The lack of a significant difference in the data obtained could be attributed to the participants' effort level, where participants might not have put the same effort in different sessions. However, this effort is not captured in their subjective cognitive workload perception and task performance, where the results indicated higher levels of cognitive workload. Another possible explanation is the 30-second break between the pick-and-place trials. This break was provided to bring the robotic arm to a home location so all participants could start the pick-and-place task from the same location. However, this data (break) was included for analysis instead of focusing on data per trial. This approach was adopted to capture a more comprehensive assessment of the participant's physiological responses, given that the time window per trial (50 seconds) is insufficient for HRV analysis based on literature and our prior works [34]. Finally, the lack of a significant difference in the data obtained could be attributed to individual differences among the participants, such as gender and age, that might have influenced physiological responses to the HRC task [37, 39].

5.4 Study Limitations

While this study provided valuable insights into the interplay of cognitive workload and user interfaces on task performance, trust, and situational awareness during HRC, this research has a few limitations. One of the primary limitations is the number of participants in the study, and the participants recruited in this study were college students pursuing advanced degrees. Future efforts should focus on recruiting industry assembly line employees, as most are high school graduates. Additionally, efforts should be made to ensure that the participant pool includes a diverse sample of varying ages, including adults, older adults, and younger individuals. This study is also limited by the fact that none of the participants had prior experience with MR. Therefore, this novel technology may have demanded additional mental effort from participants to comprehend and complete the task. Future research should investigate the time-on-task effect, examining the learning effect based on earlier and later MR UI usage trials. Another limitation lies in the analysis of the physiological activity and, more specifically, the LF/HF ratio. The total session window was used for the analysis, which included the small break between the trials. This approach was the most suitable for analyzing these data since the HRV analysis requires at least five minutes of data for analysis, and each trial lasts approximately 50 seconds. Future studies should consider using additional physiological metrics such as pupillometry along with HRV. Finally, another limitation is that while affective states impact cognitive load, we focused on how UI could impact cognitive load and task performance in this study. This was intentional to not overburden the participants with surveys and experiments that collected affective states. With these limitations, the findings from this study can still provide researchers with a foundational background and information that should be considered.

6 CONCLUSION AND FUTURE WORK

This study focused on developing and evaluating the impact of User interfaces as assistive technology in HRC. Specifically, 2-D and mixed reality user interfaces were developed and utilized during human-robot collaborative tasks to understand their effect on cognitive workload, task performance, situation awareness, and participants' trust. Our findings indicate that, irrespective of the use of interfaces, high cognitive workload scenarios reduced task performance, situational awareness, and trust and increased the users' cognitive load. In investigating the impact of user interfaces, mixed reality UI during high cognitive workload scenarios reduced task performance, situational awareness, and trust. These findings suggest that MR interfaces can potentially negatively impact cognitively demanding tasks and complex scenarios rather than being an effective assistive technology. However, these observations could be because of the current cumbersome design of MR systems available on the market or the participants' lack of experience with the MR systems. Future research should try to involve participants with more MR experience or potentially conduct a longitudinal study with multiple interventions to understand the novelty effect of MR systems. Finally, our findings clearly show the importance of including a combination of subjective and objective performance metrics for the evaluation of the system. We highly recommend that future studies consider this to understand and enhance the HRC experience.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Science Foundation under award number 1900704.

REFERENCES

[1] Paul S Addison. 2005. Wavelet transforms and the ECG: a review. *Physiological measurement* 26, 5 (2005), R155.

[2] Areej AlHogail. 2018. Improving IoT technology adoption through improving consumer trust. *Technologies* 6, 3 (2018), 64.

[3] Angelos Argyrou, Christos Giannoulis, Andreas Sardelis, Panagiotis Karagiannis, George Michalos, and Sotiris Makris. 2018. A data fusion system for controlling the execution status in human-robot collaborative cells. *Procedia CIRP* 76 (2018), 193–198.

[4] Hasan Ayaz, Ben Willems, B Bunce, Patricia A Shewokis, Kurtulus Izzetoglu, Sehchang Hah, Atul Deshmukh, and Banu Onaral. 2010. Cognitive workload assessment of air traffic controllers using optical brain imaging sensors. *Advances in understanding human performance: Neuroergonomics, human factors design, and special populations* (May 2010), 21–31.

[5] Tammy Bahmanziari, J Michael Pearson, and Leon Crosby. 2003. Is trust important in technology adoption? A policy capturing approach. *Journal of Computer Information Systems* 43, 4 (2003), 46–54.

[6] James Baumeister, Seung Youb Ssin, Neven AM ElSayed, Jillian Dorrian, David P Webb, James A Walsh, Timothy M Simon, Andrew Irlitti, Ross T Smith, Mark Kohler, et al. 2017. Cognitive cost of using augmented reality displays. *IEEE transactions on visualization and computer graphics* 23, 11 (2017), 2378–2388.

[7] Esther MJ Bekkers, Anat Mirelman, Lisa Alcock, Lynn Rochester, Freek Nieuwhof, Bastiaan R Bloem, Elisa Pelosin, Laura Avanzino, Andrea Cereatti, Ugo Della Croce, et al. 2020. Do patients with Parkinson's disease with freezing of gait respond differently than those without to treadmill training augmented by virtual reality? *Neurorehabilitation and neural repair* 34, 5 (2020), 440–449.

[8] John G Casali and Walter W Wierwille. 1984. On the measurement of pilot perceptual workload: a comparison of assessment techniques addressing sensitivity and intrusion issues. *Ergonomics* 27, 10 (May 1984), 1033–1050.

[9] Manuel Cebrian Ramos, Nicholas A Christakis, Iain D Couzin, Matthew O Jackson, Nicholas R Jennings, Ece Kamar, Isabel M Kloumann, Hugo Larochelle, David Lazer, Richard McElreath, et al. 2019. Machine Behaviour. *Nature* (2019).

[10] George Charalambous, Sarah Fletcher, and Philip Webb. 2016. The development of a scale to evaluate trust in industrial human-robot collaboration. *International Journal of Social Robotics* 8 (2016), 193–209.

[11] Rebecca L Charles and Jim Nixon. 2019. Measuring mental workload using physiological measures: A systematic review. *Applied ergonomics* 74 (Jan. 2019), 221–232.

[12] Arzu Çöltekin, Ian Lochhead, Marguerite Madden, Sidonie Christophe, Alexandre Devaux, Christopher Pettit, Oliver Lock, Shashwat Shukla, Lukáš Herman, Zdeněk Stachoň, et al. 2020. Extended reality in spatial sciences: A review of research challenges and future directions. *ISPRS International Journal of Geo-Information* 9, 7 (2020), 439.

[13] Oscar Danielsson, Anna Syberfeldt, Rodney Brewster, and Lihui Wang. 2017. Assessing instructions in augmented reality for human-robot collaborative assembly by using demonstrators. *Procedia CIRP* 63 (2017), 89–94.

[14] Luís Fernando de Souza Cardoso, Flávia Cristina Martins Queiroz Mariano, and Ezequiel Roberto Zorzal. 2020. A survey of industrial augmented reality. *Computers & Industrial Engineering* 139 (2020), 106159.

[15] Munjal Desai. 2012. *Modeling trust to improve human-robot interaction.* Ph. D. Dissertation. University of Massachusetts Lowell.

[16] Abhiraj Deshpande and Inki Kim. 2018. The effects of augmented reality on improving spatial problem solving for object assembly. *Advanced Engineering Informatics* 38 (2018), 760–775.

[17] AK Dey and DD Mann. 2010. A complete task analysis to measure the workload associated with operating an agricultural sprayer equipped with a navigation device. *Applied ergonomics* 41, 1 (Jan. 2010), 146–149.

[18] Tien-Thong Nguyen Do, Avinash Kumar Singh, Carlos A Tirado Cortes, and Chin-Teng Lin. 2020. Estimating the cognitive load in physical spatial navigation. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 568–575.

[19] Mathilde Drouot, Nathalie Le Bigot, Emmanuel Bricard, Jean-Louis De Bougrenet, and Vincent Nourrit. 2022. Augmented reality on industrial assembly line: Impact on effectiveness and mental workload. *Applied Ergonomics* 103 (2022), 103793.

[20] Florian N Egger et al. 2001. Affective design of e-commerce user interfaces: How to maximise perceived trustworthiness. In *Proc. Intl. Conf. Affective Human Factors Design*. Citeseer, 317–324.

[21] Mica R Endsley. 1988. Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society annual meeting*, Vol. 32. Sage Publications Sage CA: Los Angeles, CA, 97–101.

[22] Mica R Endsley. 1999. Situation awareness in aviation systems. *Handbook of aviation human factors* 257 (1999), 276.

[23] Mica R Endsley. 2018. Automation and situation awareness. In *Automation and human performance: Theory and applications*. CRC Press, 163–181.

[24] Mica R Endsley and Daniel J Garland. 2000. Pilot situation awareness training in general aviation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 44. SAGE Publications Sage CA: Los Angeles, CA, 357–360.

[25] Mica R Endsley and Daniel J Garland. 2000. *Situation awareness analysis and measurement.* CRC Press.

[26] L Finsen, K Søgaard, C Jensen, V Borg, and H Christensen. 2001. Muscle activity and cardiovascular response during computer-mouse work with and without memory demands. *Ergonomics* 44, 14 (Nov. 2001), 1312–1329.

[27] Marina Fujita, Ryu Kato, and Arai Tamio. 2010. Assessment of operators' mental strain induced by hand-over motion of industrial robot manipulator. In *19th International Symposium in Robot and Human Interactive Communication*. IEEE, 361–366.

[28] Rinat R Galin and Roman V Meshcheryakov. 2020. Human-robot interaction efficiency and human-robot collaboration. In *Robotics: Industry 4.0 Issues & New Intelligent Control Paradigms*. Springer, 55–63.

[29] Wilbert O Galitz. 2007. *The essential guide to user interface design: an introduction to GUI design principles and techniques.* John Wiley & Sons.

[30] Edith Galy. 2018. Consideration of several mental workload categories: perspectives for elaboration of new ergonomic recommendations concerning shiftwork. *Theoretical issues in ergonomics science* 19, 4 (Sept. 2018), 483–497.

[31] Nirit Gavish, Teresa Gutiérrez, Sabine Webel, Jorge Rodríguez, Matteo Peveri, Uli Bockholt, and Franco Tecchia. 2015. Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments* 23, 6 (2015), 778–798.

[32] Vishnunarayan Girishan Prabhu, Kevin Taaffe, Ronald Pirrallo, and Dotan Shvorin. 2020. Stress and burnout among attending and resident physicians in the ED: a comparative study. *IISE Transactions on Healthcare Systems Engineering* 11, 1 (Aug. 2020), 1–10.

[33] Christos Gkournelos, Panagiotis Karagiannis, Niki Kousi, George Michalos, Spyridon Koukas, and Sotiris Makris. 2018. Application of wearable devices for supporting operators in human-robot cooperative assembly tasks. *Procedia CIRP* 76 (2018), 177–182.

[34] Junichiro Hayano and Emi Yuda. 2019. Pitfalls of assessment of autonomic function by heart rate variability. *Journal of physiological anthropology* 38, 1 (2019), 1–8.

[35] Nis Hjortskov, Dag Rissén, Anne Katrine Blangsted, Nils Fallentin, Ulf Lundberg, and Karen Søgaard. 2004. The effect of mental stress on heart rate variability and blood pressure during computer work. *European journal of applied physiology* 92, 1 (June 2004), 84–89.

[36] G Robert J Hockey. 1997. Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological psychology* 45, 1-3 (March 1997), 73–93.

[37] Sarah K Hopko, Riya Khurana, Ranjana K Mehta, and Prabhakar R Pagilla. 2021. Effect of Cognitive Fatigue, Operator Sex, and Robot Assistance on Task Performance Metrics, Workload, and Situation Awareness in Human-Robot Collaboration. *IEEE Robotics and Automation Letters* 6, 2 (2021), 3049–3056.

[38] Edwin L Hutchins, James D Hollan, and Donald A Norman. 1985. Direct manipulation interfaces. *Human–computer interaction* 1, 4 (1985), 311–338.

[39] Apostolos Kalatzis, Sarah Hopko, Ranjana K Mehta, Laura Stanley, and Mike P Wittie. 2022. Sex Parity in Cognitive Fatigue Model Development for Effective Human-Robot Collaboration. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 10951–10958.

[40] Kristiina Karvonen. 2000. The beauty of simplicity. In *Proceedings on the 2000 conference on Universal Usability*. 85–90.

[41] Brenden Keyes, Mark Micire, Jill L Drury, Holly A Yanco, and Daisuke Chugo. 2010. Improving human-robot interaction through interface evolution. In *Human-robot interaction*. InTech, 183–202.

[42] Kyong-Ho KIM and Kwang-Yun WOHN. 2011. Effects on Productivity and Safety of Map and Augmented Reality Navigation Paradigms. *IEICE Transactions on Information and Systems* E94.D, 5 (2011), 1051–1061. https://doi.org/10.1587/transinf.E94.D.1051

[43] Lina Kluy and Eileen Roesler. 2021. Working with industrial cobots: The influence of reliability and transparency on perception and trust. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 65. SAGE Publications Sage CA: Los Angeles, CA, 77–81.

[44] Fabrizio Lamberti, Federico Manuri, Andrea Sanna, Gianluca Paravati, Pietro Pezzolla, and Paolo Montuschi. 2014. Challenges, opportunities, and future trends of emerging techniques for augmented reality-based maintenance. *IEEE Transactions on Emerging Topics in Computing* 2, 4 (2014), 411–421.

[45] Allison Langer, Ronit Feingold-Polak, Oliver Mueller, Philipp Kellmeyer, and Shelly Levy-Tzedek. 2019. Trust in socially assistive robots: Considerations for use in rehabilitation. *Neuroscience & Biobehavioral Reviews* 104 (2019), 231–239.

[46] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80.

[47] John D Lee and Bobbie D Seppelt. 2009. Human factors in automation design. *Springer handbook of automation* (2009), 417–436.

[48] Jiewu Leng, Weinan Sha, Baicun Wang, Pai Zheng, Cunbo Zhuang, Qiang Liu, Thorsten Wuest, Dimitris Mourtzis, and Lihui Wang. 2022. Industry 5.0: Prospect and retrospect. *Journal of Manufacturing Systems* 65 (2022), 279–295.

[49] Arthur D Loewy and K Michael Spyer. 1990. *Central regulation of autonomic functions.* Oxford University Press.

[50] Sotiris Makris, Panagiotis Karagiannis, Spyridon Koukas, and Aleksandros-Stereos Matthaiakis. 2016. Augmented reality system for operator support in human–robot collaborative assembly. *CIRP Annals* 65, 1 (2016), 61–64.

[51] Lukasz M Mazur, Prithima R Mosaly, Carlton Moore, and Lawrence Marks. 2019. Association of the usability of electronic health records with cognitive workload and performance levels among physicians. *JAMA network open* 2, 4 (April 2019), e191709–e191709.

[52] George Michalos, Panagiotis Karagiannis, Sotiris Makris, Önder Tokçalar, and George Chryssolouris. 2016. Augmented reality (AR) applications for supporting human-robot interactive cooperation. *Procedia CIRP* 41 (2016), 370–375.

[53] George Michalos, Niki Kousi, Panagiotis Karagiannis, Christos Gkournelos, Konstantinos Dimoulas, Spyridon Koukas, Konstantinos Mparis, Apostolis Papavasileiou, and Sotiris Makris. 2018. Seamless human robot collaborative assembly–An automotive case study. *Mechatronics* 55 (2018), 194–211.

[54] Dave Miller, Annabel Sun, and Wendy Ju. 2014. Situation awareness with different levels of automation. In *2014 IEEE International Conference on systems, man, and cybernetics (SMC)*. IEEE, 688–693.

[55] Mohsen Naderpour, Salman Nazir, and Jie Lu. 2015. The role of situation awareness in accidents of large-scale technological systems. *Process Safety and Environmental Protection* 97 (2015), 13–24.

[56] Scott Ososky, Tracy Sanders, Florian Jentsch, Peter Hancock, and Jessie YC Chen. 2014. Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems. In *Unmanned systems technology XVI*, Vol. 9084. SPIE, 112–123.

[57] Sharon Oviatt. 2006. Human-centered design meets cognitive load theory: designing interfaces that help people think. In *Proceedings of the 14th ACM international conference on Multimedia*. 871–880.

[58] Riccardo Palmarini, Iñigo Fernandez del Amo, Guglielmo Bertolino, Gino Dini, John Ahmet Erkoyuncu, Rajkumar Roy, and Michael Farnsworth. 2018. Designing an AR interface to improve trust in Human-Robots collaboration. *Procedia CIRP* 70 (2018), 350–355.

[59] Stergios Papanastasiou, Niki Kousi, Panagiotis Karagiannis, Christos Gkournelos, Apostolis Papavasileiou, Konstantinos Dimoulas, Konstantinos Baris, Spyridon Koukas, George Michalos, and Sotiris Makris. 2019. Towards seamless human robot collaboration: integrating multimodal interaction. *The International Journal of Advanced Manufacturing Technology* 105 (2019), 3881–3897.

[60] Pedro Pinheiro-Chagas, Manuela Piazza, and Stanislas Dehaene. 2019. Decoding the processing stages of mental arithmetic with magnetoencephalography. *cortex* 114 (2019), 124–139.

[61] Vishnunarayan Girishan Prabhu, Laura Stanley, Robert Morgan, and Brayton Shirley. 2022. Comparing the Efficacy of a Video and Virtual Reality Intervention to Mitigate Surgical Pain and Anxiety. In *Human Interaction, Emerging Technologies and Future Systems V: Proceedings of the 5th International Virtual Conference on Human Interaction and Emerging Technologies, IHIET 2021, August 27-29, 2021 and the 6th IHIET: Future Systems (IHIET-FS 2021), October 28-30, 2021, France*. Springer, 1041–1048.

[62] Kevin Rehring, Malte Greulich, Laurenz Bredenfeld, and Frederik Ahlemann. 2019. Let's get in touch-decision making about enterprise architecture using 3D visualization in augmented reality. (2019).

[63] Minjin Rheu, Ji Youn Shin, Wei Peng, and Jina Huh-Yoo. 2021. Systematic review: Trust-building factors and implications for conversational agent design. *International Journal of Human–Computer Interaction* 37, 1 (2021), 81–96.

[64] Philipp Schmidt, Felix Biessmann, and Timm Teubner. 2020. Transparency and trust in artificial intelligence systems. *Journal of Decision Systems* 29, 4 (2020), 260–278.

[65] Jean Scholtz. 2003. Theory and evaluation of human robot interactions. In *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the.* IEEE, 10–pp.

[66] Ben Shneiderman and Catherine Plaisant. 2010. *Designing the user interface: Strategies for effective human-computer interaction.* Pearson Education India.

[67] Kip Smith and Peter A Hancock. 1995. Situation awareness is adaptive, externally directed consciousness. *Human factors* 37, 1 (1995), 137–148.

[68] A Sneddon, K Mearns, R Flin, and R Bryden. 2004. Safety and situation awareness in offshore crews. In *SPE International Conference on Health, Safety, and Environment in Oil and Gas Exploration and Production*. OnePetro.

[69] Pooya Soltani and Antoine HP Morice. 2020. Augmented reality tools for sports education and training. *Computers & Education* 155 (2020), 103923.

[70] John Sweller. 2011. Cognitive load theory. In *Psychology of learning and motivation*. Vol. 55. Elsevier, 37–76.

[71] Daniel Szafir. 2019. Mediating human-robot interactions with virtual, augmented, and mixed reality. In *Virtual, Augmented and Mixed Reality. Applications and Case Studies: 11th International Conference, VAMR 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings, Part II 21*. Springer, 124–149.

[72] Richard M Taylor. 2017. Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In *Situational awareness*. Routledge, 111–128.

[73] Suzanne Tolmeijer, Ujwal Gadiraju, Ramya Ghantasala, Akshit Gupta, and Abraham Bernstein. 2021. Second chance for a first impression? Trust development in intelligent system interaction. In *Proceedings of the 29th ACM Conference on user modeling, adaptation and personalization*. 77–87.

[74] Georgios Tsamis, Georgios Chantziaras, Dimitrios Giakoumis, Ioannis Kostavelis, Andreas Kargakos, Athanasios Tsakiris, and Dimitrios Tzovaras. 2021. Intuitive and safe interaction in multi-user human robot collaboration environments through augmented reality displays. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE, 520–526.

[75] DWF Van Krevelen and Ronald Poelman. 2010. A survey of augmented reality technologies, applications and limitations. *International journal of virtual reality* 9, 2 (2010), 1–20.

[76] Valeria Villani, Fabio Pini, Francesco Leali, and Cristian Secchi. 2018. Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications. *Mechatronics* 55 (2018), 248–266.

[77] Markus Wallmyr, Taufik Akbar Sitompul, Tobias Holstein, and Rikard Lindell. 2019. Evaluating mixed reality notifications to support excavator operator awareness. In *IFIP Conference on Human-Computer Interaction*. Springer, 743–762.

[78] Ning Wang, David V Pynadath, and Susan G Hill. 2016. Trust calibration within a human-robot team: Comparing automatically generated explanations. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 109–116.

[79] Wickens, Christopher D and McCarley, Jason S. 2019. *Applied attention theory*. CRC press.

[80] Glenn F Wilson and Christopher A Russell. 2003. Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Human factors* 45, 4 (Dec. 2003), 635–644.

[81] Robert H Wortham and Andreas Theodorou. 2017. Robot transparency, trust and utility. *Connection Science* 29, 3 (2017), 242–248.

[82] Anqi Xu and Gregory Dudek. 2016. Maintaining efficient collaboration with trust-seeking robots. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 3312–3319.

[83] Xun Xu, Yuqian Lu, Birgit Vogel-Heuser, and Lihui Wang. 2021. Industry 4.0 and Industry 5.0—Inception, conception and perception. *Journal of Manufacturing Systems* 61 (2021), 530–535.

[84] Zhen Yang, Jinlei Shi, Wenjun Jiang, Yuexin Sui, Yimin Wu, Shu Ma, Chunyan Kang, and Hongting Li. 2019. Influences of augmented reality assistance on performance and cognitive loads in different stages of assembly task. *Frontiers in psychology* 10 (2019), 1703.

[85] Sean Ye, Glen Neville, Mariah Schrum, Matthew Gombolay, Sonia Chernova, and Ayanna Howard. 2019. Human trust after robot mistakes: Study of the effects of different forms of robot communication. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1–7.

[86] Mark S Young and Neville A Stanton. 2002. Attention and automation: new perspectives on mental underload and performance. *Theoretical issues in ergonomics science* 3, 2 (2002), 178–194.

[87] Kun Yu, Shlomo Berkovsky, Ronnie Taib, Dan Conway, Jianlong Zhou, and Fang Chen. 2017. User trust dynamics: An investigation driven by differences in system performance. In *Proceedings of the 22nd international conference on intelligent user interfaces*. 307–317.