Effect of Cognitive Fatigue, Operator Sex, and Robot Assistance on Task Performance Metrics, Workload, and Situation Awareness in Human-Robot Collaboration

Sarah K. Hopko[1], Riya Khurana[2], Ranjana K. Mehta[1], Prabhakar R. Pagilla[2], *Member, IEEE*

***Abstract*—Advancements in robot technology are allowing for increasing integration of humans and robots in shared space manufacturing processes. While individual task performance of the robotic assistance and human operator can be separately optimized, the interaction between humans and robots can lead to emergent effects on collaborative performance. Thus, the performance benefits of increased automation in robotic assistance and its impact by human factors need to be considered. As such, this paper examines the interplay of operator sex, their cognitive fatigue states, and varying levels of automation on collaborative task performance, operator situation awareness and perceived workload, and physiological responses (heart rate variability; HRV). Sixteen participants, balanced by sex, performed metal polishing tasks directly with a UR10 collaborative robot under different fatigued states and with varying levels of robotic assistance. Perceived fatigue, situation awareness, and workload were measured periodically, in addition to continuous physiological monitoring and three task performance metrics: task efficiency, accuracy, and precision, were obtained. Higher robotic assistance demonstrated direct task performance benefits. However, unlike females, males did not perceive the performance benefits as better with higher automation. A relationship between situation awareness and automation was observed in both the HRV signals and subjective measures, where increased robot assistance reduced the attentional supply and task engagement of participants. The consideration of the interplay between human factors, such as operator sex and their cognitive states, and robot factors on collaborative performance can lead to improved human-robot collaborative system designs.**

***Index Terms*—Autonomous Agents, ECG, Human Factors and Human-in-the-Loop, Human-Robot Collaboration, Industrial Robots, Manufacturing, Sex Differences, Surface Finishing**

I. Introduction

Advancements in automation and robotics have made it increasingly possible to incorporate industrial robots into shared space manufacturing facilities alongside humans. In manufacturing processes, robots have the comparative advantage of performing precise movements with repetition and uniformity, while humans retain the advantage of improved cognition, recognition, and creative decision making. Active teaming between humans and robots allows for utilization of the respective benefits of humans and robots to allow for more robust manufacturing processes. Surface finishing operations, such as deburring, chamfering, sanding, grinding, and polishing are important processes for producing finished products. Currently, many of these operations are performed manually by the operator due to sensing and uncertainty handling limitations for automation, complex part geometries, and difficulties in workpiece registration in the workspace [1], [2]. Human-robot collaboration (HRC) is therefore envisioned in such cases to improve task performance [3]. However, the introduction of humans and robot into a shared work environment has implications on the safety of the collaborative system in addition to how the varying implementations of robotic assistance can affect respective performances of humans and robots, and effectively their shared collaborative performance.

One of the essential human factor considerations in HRC is the situation awareness of a human operator due to the risk of operation near industrial machinery and the comparative benefit of higher situation awareness in humans than robots. Situation awareness (SA) is the ability to perceive and comprehend what is happening around you and the ability to predict what will happen in the future [4]. SA is impacted by the supply of attentional resources, demands on attentional resources, and the operator's understanding of the task [5]. SA affects the operator's internal mental model of their surroundings and is known to be a strong driver in the decision making process [6], [7]. Lapses in SA have been identified as a root cause in aviation fatalities [8], train derailments [9], and other large-scale technology systems mishaps [6], and thus should also be considered in HRC systems. Along with safety factors, SA directly impacts the quality of task performance, efficiency of task completion, and ability of the operator to safely take back control from the robot [4], [10]. SA often has a tradeoff with higher automation and autonomy in robots as lower automation levels tend to be better at keeping the human 'in-the-loop', although reliable and proper implementation of higher levels of automation can provide direct system performance benefits [11].

Variation in the level of assistance provided by the robot directly impacts overall task performance, operator SA, and operator workload [10], [11]. Increasing levels of automation in robots is often intended to offload the physical burden on the operator and allow for highly repetitive tasks with heavier payloads [12]; however, the introduction of industrial robots in collaborative manufacturing can result in an increase in cognitive workload due to the use of more complex tasks or higher burden of cognition and decision making on the operator [13]. Sustaining higher levels of cognitive load may

Manuscript received: October 15, 2020; Revised: January 18, 2021; Accepted: February 16, 2021.

This paper was recommended for publication by Editor Gentiane Venture upon evaluation of the Associate Editor and Reviewers' comments.

This work was supported by the NSF under grant number 1900704.

[1]S. K. Hopko, R. K. Mehta, are with the College of Engineering, Industrial and Systems Engineering Department, Texas A&M University, College Station, USA. (email: s.hopko@tamu.edu)

[2]R. Khurana, P. R. Pagilla are with the College of Engineering, Mechanical Engineering Department, Texas A&M University, College Station, USA.

Digital Object Identifier (DOI): see top of this page.

in turn make the operators susceptible to other sources of fatigue, such as cognitive fatigue and overload with negative implications on performance [14], [15]. Within the manufacturing sector, the top three reported root causes of fatigue include work related stress, lack of sleep, and schedule shifts, all related to cognitive fatigue with primary recovery strategies including caffeinated drinks (reported by more than half of fatigued workers) and talking with coworkers (reported by ~30%) [13]. While the source of a physical stressor may be different, physical stressors also have ergonomic and cognitive implications. Cognitive fatigue can also impact perceptions of task difficulty resulting in behavioral changes that can affect system performance [15], [16], such as reducing the efficiency with which the human completes the task and inducing more human error [17]. Despite this evidence, there is a lack of understanding of how cognitive fatigue can impact human-robot interactions in shared space HRC tasks.

Cognitive functions and operator acceptance are largely impacted by the dispositional characteristics of the operators including age and sex [18]–[20]. Operator perceptions of robots and robot behavior have been shown to be impacted by the sex of the operator, and in some scenarios, sex has a larger effect on subjective perceptions than the age of the operator [18], [19]. Males and females have also been shown to have proxemic behavioral differences when interacting with robots [21]. Despite these encouraging results, sex differences in operator workload, SA, and task efficiency is largely understudied in HRC. According to the United States Census, women have consistently encompassed around one-third of the workforce in the manufacturing industry [22], and thus sex is an inevitable human factor that needs further investigation.

Investigations into objective methods to quantify varying human factors has gained popularity in human-robot interaction due to the need for non-interrupting and continuous measurements to facilitate adaptive HRC. To quantify cognitive states, heart rate variability (HRV) has been employed to classify fatigue [23], [24] and workload [25], [26] in a direct and continuous manner with minimal task interference. In addition to objective measures, subjective experiences must be considered as they can provide key behavioral information that objective measures may not able to explicitly capture. Few studies have employed a systematic and comprehensive approach of examining multiple highly relevant and interrelated factors, which allow for understanding of human factor considerations in HRC. As such, this study employs systematic empirical manipulation of cognitive fatigue, operator sex, and assistance level, with a multimodal response employing multiple task performance metrics, subjective perceptions, and physiological responses to understand their interrelations and impact on effective HRC.

II. METHODS

A. Participants

This study recruited sixteen participants, balanced by sex, with an age distribution of 25.12 ± 3.31 years from the engineering population at Texas A&M University. Eleven of the participants were seeking advanced degrees, and five were seeking undergraduate degrees, with majors in industrial, biomedical, safety, chemical, aerospace or mechanical engineering. All participants were right-hand dominant. IRB approval (IRB2020-0097DCR) and COVID-19 human subjects testing safety plan approval were both received prior to starting the experiment. Upon consent, participants reported their prior experience with industrial robotics. Three participants reported prior experience ranging from a little experience to a lot of experience. Additionally, the average participant reported slight familiarity with joystick devices which were utilized by the participant to control the robot.

B. Collaborative Task

The task employed in this study was a metal surface polishing task where participants interacted with a UR10 robot (Universal Robots, DK) and controlled the robot through right-handed joystick input controls (Fig. 1). Participants had access to six degrees of freedom, but were asked to control the X, Y and Z directions. Furthermore, the movement speed of the robot was kept uniform and was not controllable by the participants beyond stepwise binary inputs (on – 1cm/s, off – brakes to 0cm/s). In each trial participants navigated a squared S-shaped trajectory following traced markings (Fig. 1, top left inset) consisting of five main events: two U-shaped turns and three horizontal lateral movements. During the low assistance conditions, participants controlled all X and Y navigation around these events, and the robot was programmed to prevent a downward force larger than 15 N. During the high assistance conditions, the participants were responsible for lateral maneuvering of the tool (Events 1, 3, 5) and automatic control was responsible for predicting and maneuvering around corners (Events 2, 4) in addition to preventing downward force past 15 N. During the high assistance, a blue dialog box appeared in clear view of the participant when assistance took over around the turns and disappeared when control was handed back to the operator. These assistance levels were designed to keep the human operator in-the-loop while allowing the automatic control to take over the more difficult aspects of the task: maintaining uniform contact force, judging the distance before turning, and the control of turning itself.

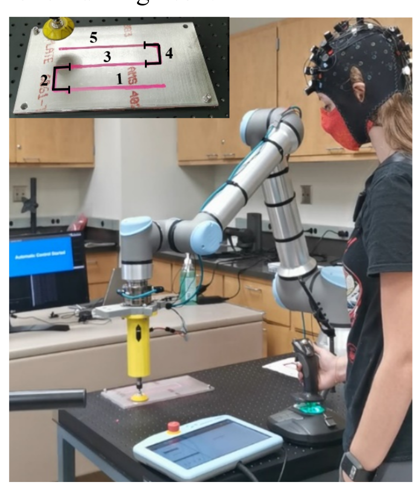

Fig 1: Experimental setup of HRC with the UR10 robot for a metal surface polishing task with a S-shaped trajectory composed of three lateral movements (events 1,3,5) and two U-turns (events 2,4; top left inset)

Participants attended two sessions, where each session focused on one level of the fatigue variable (fatigue, no fatigue), and within each session, participants underwent two robot assistance conditions (low, high; Fig. 2). Fatigue and assistance conditions were counterbalanced within both the male and female participant pool and all participants performed all experimental conditions. Participants completed ten trials in each condition (fatigue/no fatigue, low/high assistance), and each trial took approximately 60-70 s to complete. To mitigate learning or order effect, participants were allowed to practice the tasks until they felt comfortable with the controls and their performance across each assistance level. Requested training practice runs ranged from one to three trials for each level of assistance.

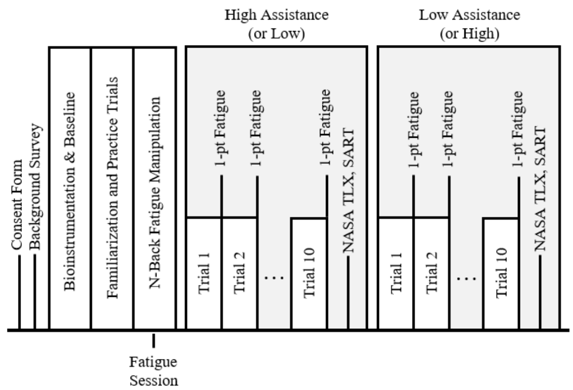

Fig. 2: Protocol timeline for each session

C. Cognitive Fatigue Manipulation

At the fatigue session, participants completed a computer based 1-hour sustained spatial 2-back test prior to interacting with the collaborative robot. The n-back test manipulates working memory, a critical cognitive component of information processing [27], [28], and sustained n-back tests have been shown to manipulate cognitive fatigue that can result in task disengagement and performance declines [29]. The spatial version of the n-back was selected to manipulate spatial working memory [27], which can influence performance across other tasks requiring spatial processes, such as navigating part metal polishing trajectories. The spatial 2-back test was given on a black background with white circular stimuli that were randomly presented within a 3x3 grid and changed location every two seconds, visible for one second (Fig. 3). The task required participants to remember the location of the stimuli and press the space bar when the current stimulus matched the location of one that appeared two events back. Perceptions of fatigue are used to validate fatigue manipulation.

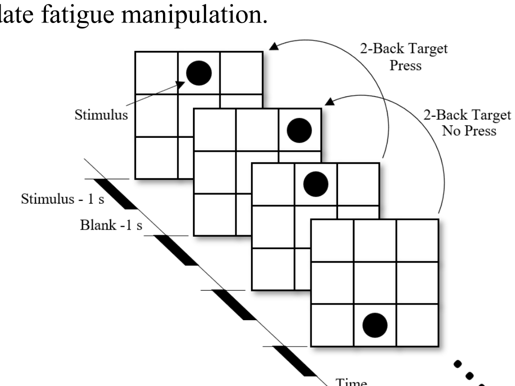

Fig. 3: Spatial 2-Back Test Diagram

D. Measurements

Task Performance

Trajectories executed by the participants were recorded from the robot for all trials at a frequency of 1 Hz. Task performance was quantified with three measures: 1) an efficiency metric, measured by the overall speed of each trial. Overall speed is defined as the traveled trajectory length divided by trial completion time.; 2) an accuracy metric, measured by deviation from the defined trajectory for each type of event (i.e. lateral or U-turn); and 3) a precision metric, measured by the variance in deviation from the defined trajectory.

Subjective Reponses

Following each trial participants were asked one question about fatigue: "What is your level of fatigue?" rated on an integer locked scale from 1 (low) to 7 (high). The ratings from the fatigue question were averaged for all trials in a condition prior to conducting statistical analysis. Following each condition (i.e., after every ten trials) participants completed the situation awareness rating technique (SART) [5] to measure SA, and the NASA task load index questionnaire (NASA TLX) [30] to quantify cognitive workload. Both questionnaires were further analyzed by their subscales and composite scores. SART consists of three subscales: attentional supply, attentional demand, and understanding of the task. The SART composite score is calculated as understanding - (demand - supply). NASA TLX has six subscales: mental demand, physical demand, temporal demand, performance, effort, and frustration. The overall workload score is calculated as the sum of subscales.

HRV Responses

A 2-lead chest affixed device, Actiheart (Actiheart 5, Camntech, UK), was used to record electrocardiogram (ECG) signals. Three participates (one male, two female) were removed from HRV analysis due to missing data. The ECG signals were corrected for ectopics and missing beats, motion related artifacts, and interpolated using the recommended settings in the software packages by Marked [31], Strasser et al. [32], and Cuiwei Li et al. [33].

Following signal correction, each condition was segmented into an early block, starting at the beginning of trial 1, and a late block, starting at the beginning of trial 6, each lasting exactly 5-minutes from their respective start point. HRV features were extracted from each block and included frequency domain HRV metrics, namely, low frequency (LF; 0.04-0.015 Hz), high frequency (HF; 0.15-0.40 Hz) and LF/HF ratio. HF is an index of activation in the parasympathetic nervous system, i.e., the 'rest and digest' system [34]. The LF measurement has been shown to index both the sympathetic nervous system, the 'fight or flight' system, in addition to the parasympathetic system. As such the LF/HF ratio is often used rather than LF alone in indexing the sympathetic response with respect to parasympathetic. The frequency domain HRV data were selected as they have been shown to vary with mental fatigue through a predominate decrease in parasympathetic activity or increase in LF/HF ratio [24], [25].

E. Statistical Analysis

Statistical significance was determined through separate repeated measures analysis of variance (RM ANOVA) on all dependent responses with significance reported at $\alpha = 0.05$ and marginal significance at $0.05 < \alpha < 0.1$. Separate RM ANOVAs were run on each task performance metric to test the effects of the three independent variables, fatigue (no fatigue/fatigue), assistance (low/high), and sex (male/female), in addition to fatigue order (fatigue first/fatigue second) and event type (lateral/U-turn). Separate RM ANOVAs were performed on all subjective responses to test the effects of the independent variables: fatigue, assistance, and sex. Finally, separate RM ANOVAs were run on all HRV measures to test the effects of fatigue, assistance, and sex in addition to phase (early/late blocks). Post hoc comparisons were performed where needed using Bonferroni corrections.

III. Results

A. Task Performance Metrics

Efficiency Metric - Overall Speed

The overall speed of the task was significantly impacted by fatigue ($p = 0.031$, $\eta^2 = 0.386$), where participants had higher speed in no fatigue (NF) at 1.435 ± 0.005 cm/s compared to fatigue (F) at 1.421 ± 0.006 cm/s. Assistance level significantly impacted the overall speed ($p < 0.001$, $\eta^2 = 0.860$), where overall speed was faster with high assistance (1.468 ± 0.007 cm/s) compared to low assistance (1.388 ± 0.007 cm/s). Fatigue order also had a significant effect on speed ($p = 0.04$, $\eta^2 = 0.358$) where participants who started in the NF condition had higher task speed in both sessions than those that started in the F condition. No effect of sex or interactions were observed (all $p > 0.132$).

Accuracy Metric - Deviation from Defined Trajectory

Assistance had a significant impact ($p < 0.001$, $\eta^2 = 0.870$) with more deviation in low assistance, 0.8 ± 0.04 cm than high assistance 0.4 ± 0.1 cm. Event type was also significant ($p < 0.001$, $\eta^2 = 0.800$) where lateral events had lower deviation (0.4 ± 0.04 cm) than turns (0.7 ± 0.1 cm). An assistance, event type interaction was observed ($p = 0.033$, $\eta^2 = 0.379$; Fig. 4) where high assistance significantly reduced deviation in both events as compared to low assistance, and U-turns were a greater source of deviation.

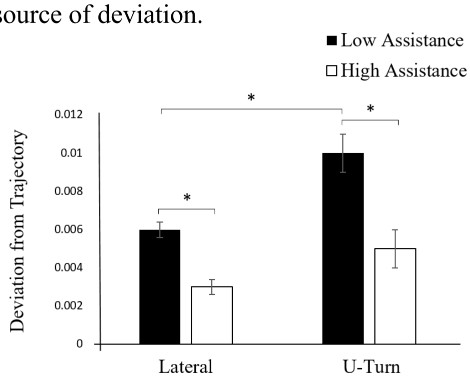

Fig. 4: Effects of Assistance and Event Type on Deviation. The error bars represent standard error. * represents significant difference

Precision Metric - Variance in Deviation from Trajectory

Event type was significant ($p = 0.007$, $\eta^2 = 0.533$) where there was higher variability in turns (0.028 cm^2) than in lateral movements (0.02 cm^2), and a significant interaction between assistance and event type was also observed ($p = 0.017$, $\eta^2 = 0.452$). A marginal three-way interaction between fatigue, assistance and sex was observed ($p = 0.092$, $\eta^2 = 0.258$), with no pairwise corrections significant after Bonferroni, and a four-way interaction was observed between assistance, event type, sex, and fatigue order ($p = 0.052$, $\eta^2 = 0.326$).

B. Subjective Responses

1-pt Fatigue

Fatigue perceptions were significantly different between NF and F ($p < 0.001$, $\eta^2 = 0.622$) with higher fatigue in the F conditions (4.325 ± 0.531) than the NF conditions (2.066 ± 0.277). Assistance also impacted fatigue perceptions ($p = 0.032$, $\eta^2 = 0.289$) with higher fatigue during low assistance at 3.409 ± 0.343 than high assistance at 2.981 ± 0.382. There was a marginal interaction between fatigue and sex ($p = 0.072$, $\eta^2 = 0.213$) where females reported greater fatigue than males in the F condition (females = 5.216 ± 2.2, males = 3.35 ± 1.8).

NASA Task Load Index (TLX)

Composite Score: Fatigue marginally affected overall workload score ($p = 0.063$, $\eta^2 = 0.225$) where NF was rated with lower overall scores than F. No other effects were observed (all $p > 0.105$).

Mental Demand Subscale: Fatigue marginally affected the mental demand subscale ($p = 0.067$, $\eta^2 = 0.225$) with lower mental demand in NF than F, in addition to a marginal fatigue and sex interaction ($p = 0.064$, $\eta^2 = 0.224$) where females perceived higher mental demand when fatigued than males.

Temporal Demand Subscale: There was a marginal effect of assistance on temporal demand ($p = 0.082$, $\eta^2 = 0.200$) where high assistance resulted in higher temporal demand. A significant three-way interaction between fatigue, assistance, and sex ($p = 0.036$, $\eta^2 = 0.278$; Fig. 5) was also found, however, post hoc analysis did not reveal any significant comparisons after Bonferroni corrections, although the difference is likely driven by the fatigue condition where males experienced higher temporal demand during high assistance than females.

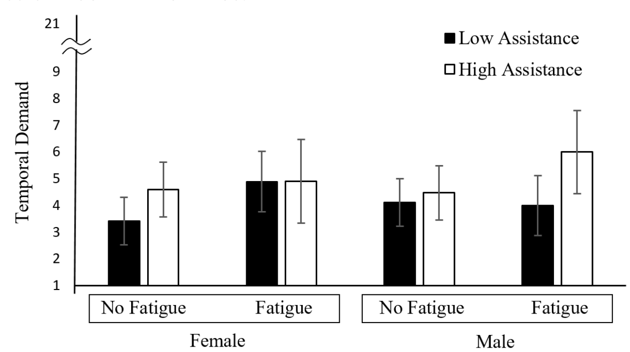

Fig. 5: Effects of Fatigue, Assistance, and Sex on Subjective Temporal Demand. The error bars represent standard error.

Performance Perception Subscale: Fatigue (p = 0.022, $\eta^2 = 0.321$) and assistance (p = 0.010, $\eta^2 = 0.389$) had a significant effect on perceived performance where participants felt they performed better in NF (15.372 ± 1.214) than F condition (14.0052 ± 1.365) and performed better with high assistance (15.372 ± 1.214) than low assistance (14.277 ± 1.335). There was also a significant two-way effect between assistance and sex (p = 0.019, $\eta^2 = 0.333$) where females felt they performed better with increased assistance (16.094 ± 1.713 in high, 14.453 ± 1.888 in low), but the same effect was not observed in men (14.201 ± 1.713 in high, 14.102 ± 1.888 in low), however all pairwise ttests were insignificant (all p > 0.296).

Effort Subscale: Assistance had a significant effect (p = 0.026, $\eta^2 = 0.308$) with more effort required for low assistance at 7.178 ± 0.966 than high at 5.794 ± 0.900. No other effects were significant for effort (all p > 0.113).

Frustration Subscale: A three-way marginal interaction was observed between fatigue, assistance, and sex (p = 0.088, $\eta^2 = 0.194$). Females rated higher frustration with the high assistance and when fatigued as compared to males, and females rated higher frustration in the low assistance when not fatigued than males; however, no pairwise ttests were significant after Bonferroni corrections.

Situation Awareness Rating Technique (SART)

Composite Score: Both fatigue (p = 0.090, $\eta^2 = 0.205$) and assistance (p = 0.089, $\eta^2 = 0.206$) influenced the composite score for SART, where higher situation awareness was associated with no fatigue (NF = 20.500 ± 1.460, F = 17.424 ± 1.529), and lower assistance (Low = 19.634 ± 1.082, High = 18.290 ± 1.469).

Attentional Supply Subscale: Assistance had a significant effect on available attentional supply, including questions such as arousal level and task engagement, (p = 0.011, $\eta^2 = 0.400$). Participants felt they had higher supply in the low assistance conditions at 11.254 ± 0.831 than in the high assistance condition at 10.134 ± 0.722.

Attentional Demand Subscale: The effects of fatigue were significant for perceived attentional demand of the task (p = 0.035, $\eta^2 = 0.299$) where the task had lower demand during NF at 6.219 ± 0.604 than F at 8.278 ± 1.001. There was also a marginal interaction between fatigue and assistance (p = 0.064, $\eta^2 = 0.211$; Fig. 6).

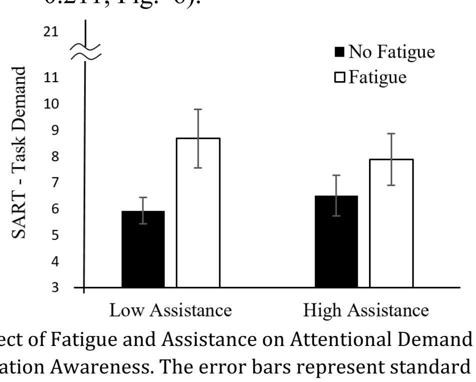

Fig. 6: Effect of Fatigue and Assistance on Attentional Demand Subscale of Situation Awareness. The error bars represent standard error.

Understanding Subscale: Fatigue had a marginal effect (p = 0.099, $\eta^2 = 0.195$) where NF had higher understanding (16.366 ± 0.525) than F (15.558 ± 0.713).

C. HRV Responses

Parasympathetic Activity - HF

A marginal three-way interaction was observed between fatigue, assistance, and phase (p = 0.067, $\eta^2 = 0.297$; Fig. 7), likely driven by higher HF in late F trials than NF trials in high assistance only. There was also a marginal four-way interaction between fatigue, assistance, phase, and sex (p = 0.068, $\eta^2 = 0.296$). All other main effects and interactions were statistically identical (all p > 0.180).

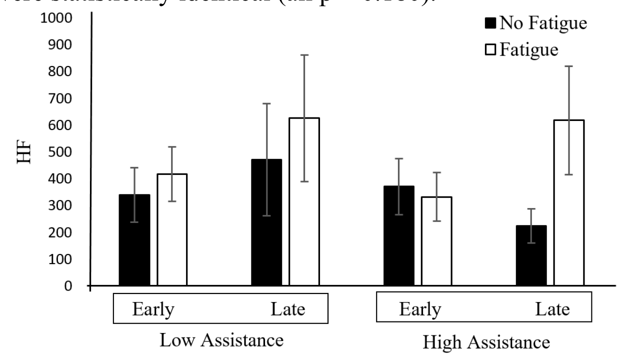

Fig. 7: Effect of Fatigue, Assistance, and Phase on Parasympathetic Activity – HF in ms². The error bars denote standard error.

Sympathetic Activity - LF

All main effects and interactions were statistically identical (all p > 0.118) excluding sex, which had a significant main effect (p = 0.032, $\eta^2 = 0.381$) where females had higher LF than males at 1972 ± 194 ms^2 verses 1338 ± 164 ms^2.

LF/HF Ratio

There was an assistance, sex marginal interaction in LF/HF ratio (p = 0.068, $\eta^2 = 0.295$) likely due to females having lower ratio than males during low assistance (post hoc p = 0.0145, Bonferroni α = 0.0125). There was also an interaction between phase and sex (p = 0.097, $\eta^2 = 0.251$) with a main phase effect (p = 0.012, 0.482) where early trials had lower ratio than late trials. A significant three-way interaction between fatigue, phase and sex was also observed (p = 0.036, $\eta^2 = 0.371$; Fig. 8), likely driven by females having higher LF/HF in late blocks as compared to males when not fatigued (p = 0.045, Bonferroni α = 0.006). All other main effects and interactions were not significant (all p > 0.114).

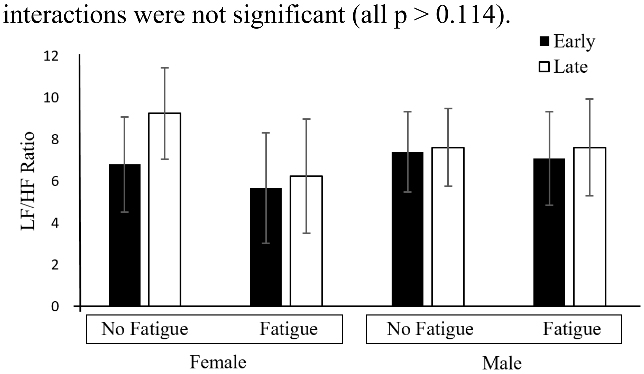

Fig. 8: Effect of Fatigue, Phase and Sex on LF/HF Ratio. The error bars represent standard error.

IV. DISCUSSION

This study investigated the impacts of operator sex, cognitive fatigue status, and robot assistance levels on operator behavior, task performance metrics, and situation awareness in shared spaced robotics. Four key takeaways of this study include:

1. High robot assistance successfully improved task efficiency and accuracy, and this was accompanied by higher perceived performance and lower efforts, but greater temporal demands.
2. There is a differential effect of assistance and fatigue on set task performance measures (i.e., efficiency, accuracy and precision) and the measures are disparately sensitive to human behavior in HRC.
3. Robot assistance directly impacted situation awareness where participants perceived less attentional supply (i.e., task engagement and arousal) with higher assistance, thereby reducing situation awareness. The effects of assistance on SA were also effectively captured by physiological measures.
4. Females perceived greater performance benefits from utilizing high automation, while males did not perceive a benefit from the assistance, despite demonstrating comparable objective performance. Sex differences were also found in the subjective and objective measures of fatigue.

A. *Higher Assistance Improves Operator Task Performance*

Assistance level dictates the extent to which automation aids the human operator in a collaborative task. In this study, the effects of assistance level on the HRC task performance were captured by the efficiency and accuracy metrics, as well as operator perceptions of performance. Regardless of sex, participants performed significantly better when using high robot assistance; the efficiency measure, i.e., overall speed, was found to have a strong main effect (improved speed of 0.08 cm/s with higher assistance) with a large effect size of 0.86. Speed itself was not a controllable dimension by the operator as the robot moved based on stepwise binary (on/off) joystick inputs. Therefore, the change in speed is a direct measure of the continuity of joystick inputs as influenced by the human behavior. As such, the higher assistance was able to consistently reduce the stuttering behavior accumulated over a relatively short trial (~1-min); over an 8-hour workday with 90% operator productivity, 0.08 cm/s equates to an additional 35 metal plates polished. This is in line with previous research that also found productivity benefits with higher assistance [11]. The consequence of manual controls or lower assistance may be more apparent in systems with continuous scale speed controls, in more difficult tasks, or on the cumulative effect of reduced efficiency on long term productivity.

The main impact of assistance level was also found for the accuracy metric, deviation from trajectory, reducing deviation by half with a large effect size of 0.87. The effect is also significant when looking at individual event types, where deviation was observed more in turn events (difficult aspect of the task) than lateral events; however, higher assistance was able to reduce much of this variance. The use of robot assistance was able to enhance the operator's performance by reducing stuttering inputs from the operator, and by improving task accuracy, more obvious around the difficult aspect of the task. Even with the minimal difficulty of the task employed in this study, the benefits of higher assistance were observed.

B. *Cognitive Fatigue and its Interactive Effect with Increased Assistance*

Operator fatigue is a critical human factors challenge that impacts productivity and safety across numerous industrial domains [35], [36]. While a prevalent safety risk in the manufacturing sector, cognitive fatigue is rarely considered in the design and implementation of highly instrumented collaborative robotic systems. Cognitive fatigue is associated with declines in attentional resources and impaired situation awareness, which are both important human attributes required for effective HRC [10], [37]. In the present study, the accuracy and precision metrics of HRC task performance were not found to be impacted by operator fatigue states. However, the efficiency metric captured the detrimental impacts of fatigue on HRC. For example, the HRC task completion speed was slower when operators were fatigued. This was further elucidated by the impact of fatigue order on task completion speeds, i.e., the participants who started with fatigue session had lasting effects on task performance where they continued to perform worse in both sessions. These results highlight the importance of utilizing appropriate task performance metrics that are sensitive to and can effectively capture critical operator states and their influences on effective HRCs.

The fatiguing task employed in this study was a sustained spatial 2-back test, which was highly cognitively fatiguing, and every participant reported increasing levels of discomfort with taking the test. It is likely that the effects of the 60-minute cognitive task were so fatiguing in comparison to the collaborative task, that participants began recovery during the HRC portion of the experiment. Parasympathetic activity has been shown to decrease with mental fatigue [24], therefore the increase is HF might suggest recovery when the stressor is removed. The effect of recovery on HRV was only visible during the fatigue session, particularly in the high assistance condition, through an increase in parasympathetic activity in later trials. Furthermore, participants rated lower fatigue and lower mental demand in the high assistance condition than low assistance. High assistance allowed for "human-out-of-the-loop" during the HRC due to the full autonomy of the robot around turns and remaining reasonably engaged for the remainder of the task via lateral navigations [11].

C. *Elements of Situation Awareness are Impacted by Assistance Level and Operator Cognitive Fatigue*

While automation-aided "human-out-of-the-loop" alleviated perceptions of cognitive fatigue in the present study, there are critical considerations regarding dynamically changing human in/out of the loop in HRCs, such as that manipulated here. Human behavior can be affected by both human and robot factors and the effect of human-out-of-the-loop of the task can reduce situation awareness and make it more difficult for the operator to reengage with the task [4], [10], [11]. In this study, higher robot assistance, regardless of operator fatigue states, were found to reduce operator situation awareness. Decomposing SA subscales revealed that participants reported lower perceived attentional supply (i.e. low task engagement, arousal) with high assistance. Interestingly, cognitive fatigue states resulted in greater demand subscale of SA, which was further exacerbated in the low assistance condition as expected [38]. These results highlight that robot automation level and operator fatigue impact different attributes of SA, thereby providing opportunities and guidance for developing closed loop engineering solutions to support cognitive processes to maintain or augment operator SA during HRC that are adaptive to operator fatigued states.

D. *Cognitive Fatigue Differentially Impacts the Sexes*

In general, female participants reported greater levels of fatigue and mental demand throughout the fatigue session than males irrespective of the robot assistance levels. The physiological responses also captured these sex differences, whereby females exhibited higher LF/HF ratio in late blocks when not fatigued as compared to men (Fig 8). Historically, both physiological and subjective perceptions of fatigue states have been shown to vary by sex [39], [40], which was supported by this study. An additional sex difference observed here was the perceptions of performance. Despite improved task performances in the high versus low robot assistance conditions, males did not perceive substantial improvements in their performance with high assistance, whereas females did perceive better task performance in the high robot assistance. Sex did not have an impact on perceived SA during the HRC even though sex differences were observed in response to fatigue and task performance outcomes. These findings highlight that considerations of operator sex can help identify ways that male and female operators respond to HRC [41], both behaviorally and physiologically, such that more effective HRCs can be designed that address and accommodate for such group differences. These findings also provide insights on how different population groups perceive benefits and costs of automation rated high assistance as improving their performance.

E. *Study Limitations and Future Work*

Limitations of this study need to be acknowledged. The participants recruited in this study were college students predominately seeking advanced degrees in engineering. Future work should focus on industry workers as the majority of jobs in manufacturing are taken up by high-school graduates or less [42]. However, the findings presented in this study are relevant as they set the stage for future hypothesis driven work. While one HRC use case is provided, i.e., a metal polishing task, the fundamental implications of operator factors (cognitive fatigue, sex) and robot factors (assistance level) are relevant to other HRC use cases given specific outcomes (i.e., task performance, situation awareness, workload, physiological responses), although the metrics and uncertainties in other HRC tasks may require further investigation. The generalization of the results to physical shared-space tasks where the robot and the human operator are within physical reach of each other requires further investigation due to the implications on trust and resulting allocations of attentional resources. Additionally, further investigation into contextually relevant environmental factors that affect user acceptance, such as propensity to trust, and their interaction with fatigue and assistance should be considered.

V. CONCLUSION

This work systematically examined operator fatigue, operator sex, and robot assistance level, all highly relevant and interrelated factors for optimizing HRC system designs with respect to task performance and user experience. The various task performance metrics (i.e., efficiency, accuracy, and precision) were able to selectively capture various attributes of the relationship between operator fatigue and assistance level. Our findings indicate that assistance through high automation significantly improves task accuracy and efficiency but does not change precision, whereas fatigue impacts task efficiency, but not accuracy or precision. Operator perceptions varied by robot assistance level but were different for males and females. Females perceived greater performance benefits from utilizing high automation, while males did not perceive a benefit from the assistance. Furthermore, higher automation aided “human-out-of-the-loop”, which allowed for operator fatigue recovery, measured using HRV signals and subjective perceptions; however, this resulted in lower operator situation awareness with increased perceived temporal demand. These findings demonstrate that effective HRC can be achieved by examining collaborative task performances through addressing factors at the intersection of the human level (i.e., operator sex and cognitive states), and the robot level. The systematic approach was able to capture the interdependence between the examined factors through changes in task performance metrics, subjective experiences, and physiological measures.

ACKNOWLEDGMENT

This study was supported by the National Science Foundation under grant number 1900704.

REFERENCES

[1] R. Bogue, “Finishing robots: a review of technologies and applications,” *Industrial Robot: An International Journal*, vol. 36, no. 1, pp. 6–12, Jan. 2009, doi: 10.1108/01439910910924611.

[2] Y. Wen, J. Hu, and P. Pagilla, “A Novel Robotic System for Finishing of Freeform Surfaces,” *2019 International Conference on Robotics and Automation (ICRA)*, 2019, doi: 10.1109/ICRA.2019.8793734.

[3] D. A. Abbink *et al.*, “A Topology of Shared Control Systems—Finding Common Ground in Diversity,” *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 5, pp. 509–525, Oct. 2018, doi: 10.1109/THMS.2018.2791570.

[4] M. R. Endsley and D. J. Garland, *Situation Awareness Analysis and Measurement*. CRC Press, 2000.

[5] R. M. Taylor, “Situation awareness rating technique (SART): the development of a tool for aircrew systems design,” vol. Situational Awareness in Aerospace Operations (Chapter 3), no. Frace: Neuilly sur-Seine, p. NATO-AGARD-CP-478, 1990.

[6] M. Naderpour, S. Nazir, and J. Lu, “The role of situation awareness in accidents of large-scale technological systems,” *Process Safety and*

Environmental Protection, vol. 97, pp. 13–24, Sep. 2015, doi: 10.1016/j.psep.2015.06.002.

[7] K. Smith and P. A. Hancock, "Situation Awareness Is Adaptive, Externally Directed Consciousness," *Hum Factors*, vol. 37, no. 1, pp. 137–148, Mar. 1995, doi: 10.1518/001872095779049444.

[8] National Transportation Safety Board, "Aircraft accident report-Northwest Airlines- McDonnell Douglas DC-9–82, N312RC, Detroit Metropolitan Wayne Co., Airport, Romulus MI, August 16, 1987, Washington, DC," 1988. Accessed: Oct. 12, 2020. [Online]. Available: https://libraryonline.erau.edu/online-full-text/ntsb/aircraft-accident-reports/AAR88-05.pdf.

[9] "Train Engineer's Loss of Situational Awareness Led to Amtrak Derailment; NTSB Says Technology Could Have Prevented Fatal Accident." https://www.ntsb.gov/news/press-releases/Pages/PR20160517.aspx (accessed Oct. 12, 2020).

[10] M. Gombolay, A. Bair, C. Huang, and J. Shah, "Computational design of mixed-initiative human–robot teaming that considers human factors: situational awareness, workload, and workflow preferences.," *International Journal of Robotics Research*, vol. 36, no. 5–7, pp. 597–617, Jun. 2017.

[11] D. Kaber and M. Endsley, "The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task," *Theoretical Issues in Ergonomics Science*, vol. 5, pp. 113–153, Mar. 2004, doi: 10.1080/1463922021000054335.

[12] M. Baker and H. A. Yanco, "Autonomy mode suggestions for improving human-robot interaction," in *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, Oct. 2004, vol. 3, pp. 2948–2953 vol.3

[13] L. Lu, F. M. Megahed, R. F. Sesek, and L. A. Cavuoto, "A survey of the prevalence of fatigue, its precursors and individual coping mechanisms among U.S. manufacturing workers," *Applied Ergonomics*, vol. 65, pp. 139–151, Nov. 2017

[14] G. Borragán, H. Slama, M. Bartolomei, and P. Peigneux, "Cognitive fatigue: A Time-based Resource-sharing account," *Cortex*, vol. 89, pp. 71–84, Apr. 2017, doi: 10.1016/j.cortex.2017.01.023.

[15] M. Mouloua and P. A. Hancock, *Human Performance in Automated and Autonomous Systems: Current Theory and Methods*. CRC Press, 2019.

[16] A. Rajavenkatanarayanan *et al.*, "Towards a robot-based multimodal framework to assess the impact of fatigue on user behavior and performance: a pilot study," in *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, New York, NY, USA, Jun. 2019, pp. 493–498, doi: 10.1145/3316782.3322776.

[17] A. Murata, "Ergonomics and cognitive engineering for robot-human cooperation," in *Proceedings 9th IEEE International Workshop on Robot and Human Interactive Communication. IEEE RO-MAN 2000 (Cat. No.00TH8499)*, Sep. 2000, pp. 206–211

[18] I. H. Kuo *et al.*, "Age and gender factors in user acceptance of healthcare robots," in *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, Sep. 2009, pp. 214–219, doi: 10.1109/ROMAN.2009.5326292.

[19] M. Strait, P. Briggs, and M. Scheutz, "Gender, more so than Age, Modulates Positive Perceptions of Language-Based Human-Robot Interactions," p. 8.

[20] P. J. Caplan, M. Crawford, J. S. Hyde, and J. T. E. Richardson, *Gender Differences in Human Cognition. Counterpoints: Cognition, Memory, and Language Series*. Oxford University Press, 2001 Evans Road, Cary, NC 27513 (ISBN-0-19-511291-1, $19, 1997.

[21] D. S. Syrdal, K. Lee Koay, M. L. Walters, and K. Dautenhahn, "A personalized robot companion? - The role of individual differences on spatial preferences in HRI scenarios," in *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*, Aug. 2007, pp. 1143–1148, doi: 10.1109/ROMAN.2007.4415252.

[22] U. C. Bureau, "Women in Manufacturing," *The United States Census Bureau*. https://www.census.gov/newsroom/blogs/random-samplings/2017/10/women-manufacturing.html (accessed Oct. 14, 2020).

[23] F. Ding, N. Fu, S. Alsamarai, and Y. Xu, "Correlating Heart Rate Variability with Mental Fatigue," *Major Qualifying Projects (All Years)*, Apr. 2012, [Online]. Available: https://digitalcommons.wpi.edu/mqp-all/2501.

[24] H. M. Melo, L. M. Nascimento, and E. Takase, "Mental fatigue and heart rate variability (HRV): The time-on-task effect," *Psychology & Neuroscience*, vol. 10, no. 4, pp. 428–436, 2017, doi: 10.1037/pne0000110.

[25] S. Delliaux, A. Delaforge, J.-C. Deharo, and G. Chaumet, "Mental Workload Alters Heart Rate Variability, Lowering Non-linear Dynamics," *Front Physiol*, vol. 10, May 2019

[26] A. Murata, "Experimental discussion on measurement of mental workload - evaluation of mental workload by HRV measures -," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E77-A, no. 2, pp. 409–416, Feb. 1994.

[27] K. A. Ellis, R. B. Silberstein, and P. J. Nathan, "Exploring the temporal dynamics of the spatial working memory n-back task using steady state visual evoked potentials (SSVEP)," *NeuroImage*, vol. 31, no. 4, pp. 1741–1751, Jul. 2006, doi: 10.1016/j.neuroimage.2006.02.014.

[28] X. Caseras *et al.*, "Probing the Working Memory System in Chronic Fatigue Syndrome: A Functional Magnetic Resonance Imaging Study Using the n-Back Task," *Psychosomatic Medicine*, vol. 68, no. 6, pp. 947–955, Dec. 2006, doi: 10.1097/01.psy.0000242770.50979.5f.

[29] J. F. Hopstaken, D. van der Linden, A. B. Bakker, and M. A. J. Kompier, "A multifaceted investigation of the link between mental fatigue and task disengagement," *Psychophysiology*, vol. 52, no. 3, pp. 305–315, 2015, doi: 10.1111/psyp.12339.

[30] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in *Advances in Psychology*, vol. 52, P. A. Hancock and N. Meshkati, Eds. North-Holland, 1988, pp. 139–183.

[31] V. Marked, "Correction of the heart rate variability signal for ectopics and missing beats," *Heart rate variability*, 1995.

[32] F. Strasser, M. Muma, and A. M. Zoubir, "Motion artifact removal in ECG signals using multi-resolution thresholding," in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, Aug. 2012, pp. 899–903.

[33] Cuiwei Li, Chongxun Zheng, and Changfeng Tai, "Detection of ECG characteristic points using wavelet transforms," *IEEE Transactions on Biomedical Engineering*, vol. 42, no. 1, pp. 21–28, Jan. 1995

[34] C. Zhao, M. Zhao, J. Liu, and C. Zheng, "Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator," *Accident Analysis & Prevention*, vol. 45, pp. 83–90, Mar. 2012, doi: 10.1016/j.aap.2011.11.019.

[35] L. M. Barker and M. A. Nussbaum, "Fatigue, performance and the work environment: a survey of registered nurses," *Journal of Advanced Nursing*, vol. 67, no. 6, pp. 1370–1382, 2011

[36] K. Baker, J. Olson, and D. Morisseau, "Work Practices, Fatigue, and Nuclear Power Plant Safety Performance," *Hum Factors*, vol. 36, no. 2, pp. 244–257, Jun. 1994, doi: 10.1177/001872089403600206.

[37] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. C. Chen, E. J. de Visser, and R. Parasuraman, "A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction," *Hum Factors*, vol. 53, no. 5, pp. 517–527, Oct. 2011, doi: 10.1177/0018720811417254.

[38] D. B. Van de Merwe, L. Van Maanen, F. B. Ter Haar, R. J. E. Van Dijk, N. Hoeba, and N. Van der Stap, "Human-Robot Interaction During Virtual Reality Mediated Teleoperation: How Environment Information Affects Spatial Task Performance and Operator Situation Awareness," in *Virtual, Augmented and Mixed Reality. Applications and Case Studies*, Cham, 2019, pp. 163–177, doi: 10.1007/978-3-030-21565-1_11.

[39] W.-L. Huang, L.-R. Chang, T. B. J. Kuo, Y.-H. Lin, Y.-Z. Chen, and C. C. H. Yang, "Gender differences in personality and heart-rate variability," *Psychiatry Research*, vol. 209, no. 3, pp. 652–657, Oct. 2013, doi: 10.1016/j.psychres.2013.01.031.

[40] J.-M. Woo and T.-S. Kim, "Gender Plays Significant Role in Short-Term Heart Rate Variability," *Appl Psychophysiol Biofeedback*, vol. 40, no. 4, pp. 297–303, Dec. 2015, doi: 10.1007/s10484-015-9295-8.

[41] Y. Wang, "Gendering Human-Robot Interaction: exploring how a person's gender impacts attitudes toward and interaction with robots," 2014, Accessed: Oct. 15, 2020. [Online]. Available: https://mspace.lib.umanitoba.ca/xmlui/handle/1993/24446.

[42] E. Richards and D. Terkanian, "Occupational Employment Projections to 2022," United States Census Bureau, Bureau of labor statistics, Dec. 2013. Accessed: Oct. 15, 2020. [Online]. Available: https://www.bls.gov/opub/mlr/2013/article/pdf/occupational-employment-projections-to-2022.pdf.