Designing Interface Aids to Assist Collaborative Robot Operators in Attention Management

Curt Henrichs,[1] Fangyun Zhao,[2] Bilge Mutlu[1]

***Abstract*— As collaborative robots become increasingly widespread in manufacturing settings, there is a greater need for tools and interfaces to support operators who integrate, supervise, and troubleshoot these systems. In this paper, we present an application of the Robot Attention Demand (*RAD*) metric for use in the design of user interfaces to support operators in collaborative manufacturing scenarios. Building on prior work that introduced *RAD*, we designed and implemented prototype timeline and countdown-timer interfaces to be used within a collaborative *assembly-inspection* task where an operator is also responsible for a separate *sorting* task. We performed a user evaluation to investigate the effects of displaying predictive *RAD* information on operator performance and perceptions of the task. Our results show lower perceived task load and increased usability scores compared to baseline condition without an interface. These findings suggest that predictive *RAD* should be used by designers and engineers developing operator interfaces for collaborative robot applications in manufacturing.**

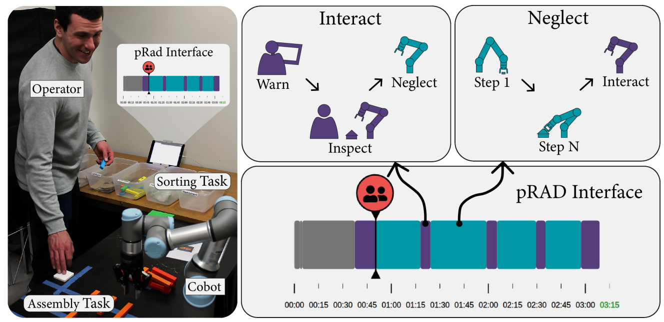

Fig. 1. Our work aims to support operators on the manufacturing shop-floor with supervising their collaborative robots. We developed two interfaces based on an extension of Robot Attention Demand, called *pRAD*, including the shown "timeline" interface. As an operator performs both an *assembly-inspection* task and a *sorting* task, the interface suggests when to switch between neglecting and interacting with the robot.

I. INTRODUCTION

As industry continues to realize the promise of increased productivity and decreased human risk by adopting collaborative robots (cobots) into the workplace [3], [30], designers and engineers will need to continue to develop solutions to improve collaborative work and to understand the effects of these solutions on people. For small- and medium-sized enterprises, cobot applications generally fall under paradigms of physically-isolated [31], shared-space-start-and-stop [23], and hand-guided [16] interactions. Engineers must consider key design decisions for human-robot coordination such as how work should be shared based on the skillsets of human and robot workers [30], [37]; how operator intent may be communicated either with verbal communication [6] or gesture [26]; and how the operator's "collaborative capacity" may be measured and communicated [20], [27].

This paper explores how "collaborative capacity" can be presented to a cobot operator to improve task performance and cognitive load. We introduce a predictive signal, *pRAD*, adapted from Robot Attention Demand (*RAD*) [14], [27], as a means to communicate the operator's upcoming collaborative obligations for manufacturing tasks, see Figure 1. We address several questions regarding *pRAD* and *RAD* as applied to the manufacturing domain: Is *RAD* a useful metric in the manufacturing domain to evaluate collaborative tasks? Does presenting the *pRAD* signal affect the operator's performance? And, does presenting the *pRAD* signal affect the operator's perceptions of the task? To address these questions, we designed and prototyped two user interfaces presenting *pRAD* based on common metaphors of timeline and countdown-timer, see Figure 2. Our first interface uses a timeline metaphor to display the relationship of *interaction time* and *neglect time* for each step of a manufacturing task. The second interface considers a countdown-timer conveying the *neglect time* of the robot and color codes to indicate task progress. As a baseline, we presented a blank screen, which displays no *pRAD* information. We performed a user study in which a human operator completed two tasks—a collaborative *assembly-inspection* task and an individual *sorting* task—in a collaborative workspace, using the interfaces to vary the amount of *pRAD* information. We tested task performance, perceived task load, and usability. Through semi-structured interviews, we examined participants' interaction strategies, preferences, and design suggestions.

Our work's contributions can be summarized as follows:

- Exploration of *RAD* as both a descriptive and predictive metric for cobots;
- The design and prototyping of alternative interfaces with varying information-detail informed by *pRAD*;
- An empirical evaluation of alternative interfaces on task and collaboration outcomes;
- An open source implementation of our prototypes using *pRAD* and task implementation.[1]

*This work was supported by National Science Foundation awards 1426824 and 1822872.

[1] Curt Henrichs and Bilge Mutlu are with Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA {cdhenrichs,bilge}@cs.wisc.edu

[2]Fangyun Zhao is with Department of Psychology, Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA olivia.zhao@wisc.edu

[1]https://github.com/Wisc-HCI/ITER

II. BACKGROUND

A. Robot Attention Demand

A solution to expressing "collaborative capacity" is Robot Attention Demand (*RAD*), proposed by Goodrich and Olsen [14], [27]. *RAD* refers to the extent to which a human worker contributes to the completion of a specific task with a particular robot, see Equation 1. Thus, *RAD* serves as a measure of the robot's task autonomy. The metric is composed of *interaction effort* (**IE**; *i.e.,* effort the operator must spend attending to the robot) and *neglect tolerance* (**NT**; *i.e.,* how well the robot can perform the task without attention or intervention from the operator). *RAD* has been used in many ways, and we are particularly interested in applying it to human-robot teaming tasks in the manufacturing setting as a time-domain implementation, similar to typical workcell measures of makespan and cycle-time. Using Olsen's [27] time-domain operationalization, *neglect tolerance* (*neglect time*) is defined as the time that a robot can complete a task without human intervention; and *interaction effort* (*interaction time*) is defined as the time period that the robot must rely on a human operator actively interacting with the robot. *Interaction effort* also includes the task switch-over delay for the human. As an example, fully autonomous systems that do not require operator interaction will have a *RAD* value of zero. Likewise, a tele-operated robot that requires operator's full attention and engagement for the duration of the task will have a *RAD* value of one. In the human-robot collaboration scenarios rapidly emerging in manufacturing, *RAD* will invariably fall between these two extremes.

$$RAD = \frac{IE}{(NT + IE)} \tag{1}$$

Prior works have explored extensions to *RAD* across several domains for both singular robots and multi-robot groups. Elara *et al.* [9] presented False Alarm Demand (*FAD*), a ratio of false alarm time over the sum of false alarm time and interaction time, and extended *RAD* by incorporating false alarm time. They found that false alarms due to either undetected or errant interaction have a negative effect on the autonomy of the controlled robot. Mohan *et al.* [24] applied extended neglect time, *RAD*, and *FAD* metrics to scenarios where a soccer-playing robot was managed by a human in tele-operated and semi-autonomous modes. They found task complexity to increase with the number of false alarms and robot performance to increase under tele-operated control (higher *RAD*). Elara and Zhou [10] applied *FAD* to an assistive walking task for a robot with tele-operation and semi-autonomous modes showing that accounting for false alarms improves performance. Many works have utilized *fan-out* in robotic systems [7], [38], [39]. *Fan-out* is defined as the number of robots an operator is able to attend to without diminishing performance [14], [27], [28]. For homogeneous robot groups, *fan-out* is the reciprocal of *RAD*. Finally, Glas *et al.* [12] evaluated a multi-robot system composed of social robots that interacted with shoppers. They used *RAD* to enable an operator to manage multiple conversations effectively.

Our work applies *RAD* to a collaborative task within the manufacturing context as both a metric and as a *predictive* signal to be displayed to the human operator acting in a supervisory role. Specifically, we leverage the time-based formulation of *RAD* to produce *pRAD* for visualization on an operator interface.

B. Operator Interfaces

An engineer designing an operator interface has to consider the information objectives of the task (*e.g.,* displaying process, diagnostic support, tool monitoring, program authoring), the modalities and technologies available (*e.g.,* augmented reality, cell-mounted monitor, teach-pendant), and available data and models (*e.g.,* user, process). For example, *Smart Workbench* [19] provides an AR-based interface for real-time feedback, but it also has several support monitors that provide secondary process and diagnostic information. Similarly, *Multi-modal Assembly Support System* (*MASS*) [25] provides operators with assembly process instructions through a cell-mounted monitor. Our work leverages cell-mounted operator interfaces to support cobot supervision for pre-existing processes.

When displaying process information, engineers have to decide on how much and what aspects to present and on a mechanism with which to synchronize the robot with the operator. A readily available solution is to use the teach pendant software provided alongside the cobot. For example, the Universal Robots Polyscope software,[2] which has a "Run" view that provides high-level program diagnostics, a "Program" view displaying the program as a hierarchy of primitives, and several views for detailed setup/diagnostics. Synchronization is handled through confirmation popups. In the literature, Tsarouchi *et al.* [36] presents a ROS-integrated interface that presents the program hierarchy. They, likewise, address coordination with a popup prompt. Ahmad *et al.* [1] recently proposed a cobot dashboard that aggregates cobots within a company for easier management and diagnosis, however the tool is not designed for an operator on the shop-floor. What is lacking in these systems is an operator-accessible representation of coordination capacity derived from task allocation and cycle-times. Our work explores the deliberate visual expression of collaborative capacity with *pRAD* to synchronize the human's and robot's work.

Interfaces that augment operator decision making and control with *RAD* have been explored in non-manufacturing contexts by Cummings and Mitchell [8]. They evaluated mechanisms to assist in decision support and task management for unmanned aerial vehicle (UAV) control demonstrating that an interface using *fan-out* helps to support operators by providing a more accurate estimate of capacity. More generally, Chen *et al.* [5] reviewed key human performance issues in monitoring multi-robot systems and concluded with research themes of operator multitasking, trust, situational awareness, operator workload, and interface design.

[2] https://www.universal-robots.com/

Fig. 2. We evaluated two *pRAD* interfaces: timeline and timer. As the task progresses through neglect into interaction phases, both interfaces presents a visual warning based on the traffic-light metaphor. Timeline automatically adjusts during interaction phases if user takes longer than prescribed.

III. DESIGN

Our approach considers *RAD* both as a task-level metric and a step-wise subtask metric. *RAD* as a task metric considers *interaction time* and *neglect time* for the entire duration of the task. Alternatively as a step-wise subtask, we break the task into a series of neglect-interaction pairs and then compute *RAD* on each pair individually. To produce a predictive signal from *RAD* (*pRAD*), we empirically measure the timing for neglect-interaction pairs and assume that future interaction would produce similar values. Refer to Figure 3 as an example of this breakdown for our evaluation task.

We expect this approach to be an acceptable representation of real manufacturing processes where allocation of work is fixed and is well-specified by engineers. Specifically, in our evaluation, we assumed that *neglect time* was measured on a static order of deterministic actions that a robot has taken. Likewise, *interaction time* is an averaged estimate of the time an operator takes at a particular step. More sophisticated methods of generating timing using a planning environment (*e.g.,* Authr [34]) could be substituted with minimal effort to produce neglect-interaction pairs. We address such limitations and extensions in the Discussion section.

To effectively display *pRAD* information to the operator, we consider two common user-interface metaphors—countdown-timer and timeline—as illustrated in Figure 2. A key assumption of our design is that presenting the step-wise *pRAD* value directly (*e.g.,* printed numerically) would be ineffective as a signal to the operator due to collapsing the temporal relationship between the neglect and interaction components. Instead, we think of *pRAD* as the conceptual relationship between components being displayed in the interface.

In the countdown-timer interface, we display the remaining predicted *neglect time*, in minutes and seconds, before the next interaction step. We warn the user of impending interaction with a traffic-light metaphor, where the background of the timer switches from green to yellow at five seconds as a warning and from yellow to red at zero seconds for interaction. The color stays red and reports zero seconds until the user finishes the interaction step. Notably, predicted *interaction time* is not displayed, instead deferring to the user to decide on how long an interaction should last.

The timeline interface provides a view of the entire task, where blue tiles represent the predicted *neglect time* and purple tiles represent predicted *interaction time*, as shown in Figure 2. Instead of presenting *pRAD* numerically, this interface aims to illustrate the signal as the relationship of neglect and interaction steps, giving operators an approximate sense of impending demand. Additionally, presenting multiple neglect and interaction steps within the timeline aids the operator in planning their work further into the future. Applying this simplification of actions in the timeline to neglect and interaction steps differs from an implementation-level timeline of task primitives (*e.g.,* grasp or move) for humans and robots, as proposed by Pearce *et al.*[30]. Designing around a metaphor of a video editing timeline, the location of the playhead on the timeline shows the relative time both for the step and the task as a whole. Exact timing is found via the track underneath the task breakdown, although it requires closer inspection to access the precise timing. If during an interaction step the participant finishes early or takes longer than initially presented, the timeline will adjust the tile size accordingly. The playhead icon performs the same signaling role as the timer interface (i.e., traffic-light warning) with the notable change that, instead of displaying remaining *neglect time*, it indicates state through three icons: check mark (neglect), exclamation point (warning), and collaborate (interact).

Based on our designs, we propose the following hypotheses:

H1—Using timeline and timer interfaces, users will perform more efficiently, such that the interaction time will be lower, relative to the baseline blank condition.

H2—Timeline and timer interfaces will decrease task load relative to the baseline blank condition.

H3—Users will perceive the timeline user interface as being more usable than timer and baseline blank conditions.

H4—The timeline user interface will have a lower interaction time and lower perceived task load relative to timer.

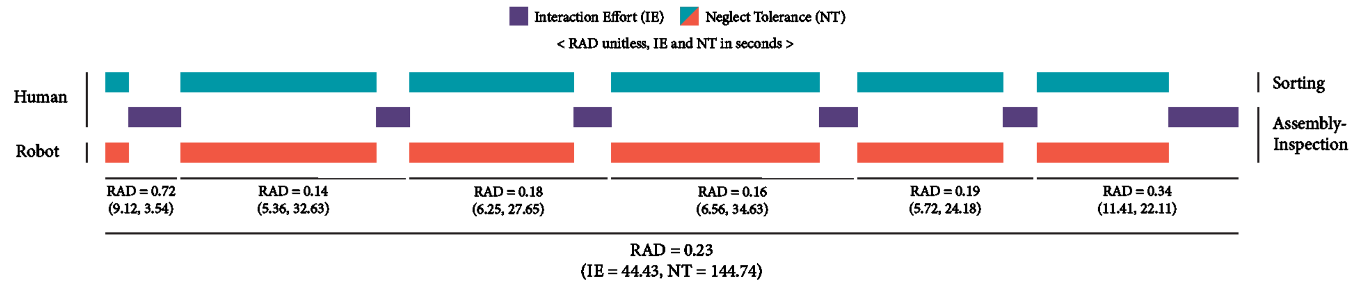

Fig. 3. Task and step-wise *RAD* breakdown for the *assembly-inspection* task and *sorting* task. *RAD* can be used both as a descriptive metric of task interaction between operator and cobot or presented to the operator as a predictive signal to guide future interaction.

IV. EVALUATION

A. Participants

A total of 34 participants (26 male, 7 female, 1 other) with an average age of 20.9 years old (SD = 2.54) were recruited from the University of Wisconsin–Madison campus. All participants had prior manufacturing or mechanical/industrial engineering experience.

B. Study Design & Procedure

Our evaluation followed a three-by-one (interface: *timeline* vs. *timer* vs. *blank*) within-subjects design, and all conditions were presented to participants in a counterbalanced order. The study was conducted as the first session of a larger protocol (see Zhao *et al.* [40] for the second session), taking 45–60 minutes per participant. Participants were shown a short instructional video describing the two experimental tasks, *assembly-inspection* and *sorting*, along with a brief description of the two user interface conditions and the baseline condition. Participants were instructed that a trial ended when both tasks were completed and asked to work as if they were at a manufacturing facility. Participants then had one short training session to showcase the robot's abilities followed by three trials each with a different interface condition. At the end of each trial, participants received a post-trial questionnaire and a semi-structured interview.

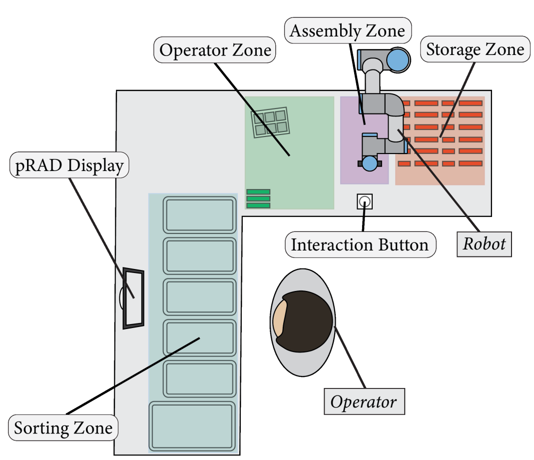

Fig. 4. Four zones of the workspace: sorting, operator, assembly, and storage. Participants perform the *sorting* task while supervising the robot in the *sorting-zone* and enter the *operator-zone* to interact with the robot. The robot makes use of the *assembly-zone* and the *storage-zone* to construct a small toy house object.

Participants acted as a supervisor and split their attention between the two tasks. This additional role reflects realistic cobot scenarios where productivity expectations for the operator is higher, as they can make use of otherwise wasted idle time waiting on the robot.

Assembly-Inspection Task–Participants were asked to inspect the robot's assembly of a small toy house composed of magnetic wooden blocks. Participants inspected the structure at each phase of assembly. After each phase, participants pressed the "interaction button" to tell the robot to continue its assembly. The robot was programmed to use a fixed set of blocks within its workspace for construction. Slight variations in initial placement of the blocks during trial setup introduced small alignment errors in the construction (*e.g.,*minor distortions in the toy house) requiring the participant to inspect the final product and make necessary corrections.

Participants were provided a reference diagram that displayed the correct configuration of blocks for each phase. They were instructed not to touch the toy house until the robot had stopped moving for their own safety, though they were allowed to visually inspect it whenever they desired. After the robot completed a neglect phase, it moved near the first block of the next phase and waited for the participant to press the "interaction button." Upon completion of the toy house, the robot returned to its initial home position.

Sorting Task–Participants were asked to sort several materials (*i.e.,* cardboard, wood blocks, wood cylinders, plastic pipes, and plastic bags) in separate labeled bins from a common unsorted bin. They were instructed not to lift the bins. Sorted bin order was randomized per trial.

Workspace–The workspace was separated into four zones: an *assembly-zone* where a Universal Robots UR3e cobot constructs the toy house; a *storage-zone* where the robot's blocks were stored; an *operator-zone* where the participant had access to several blocks and instructions; and a *sorting-zone* for the *sorting* task. Figure 4 illustrates the setup. Just outside the *assembly-zone*, we placed the "interaction button." In the *sorting-zone*, there was one bin of unsorted materials placed furthest away from the robot and five individually labeled bins to store the sorted materials. A tablet computer mounted behind the *sorting-zone* displayed the interfaces.

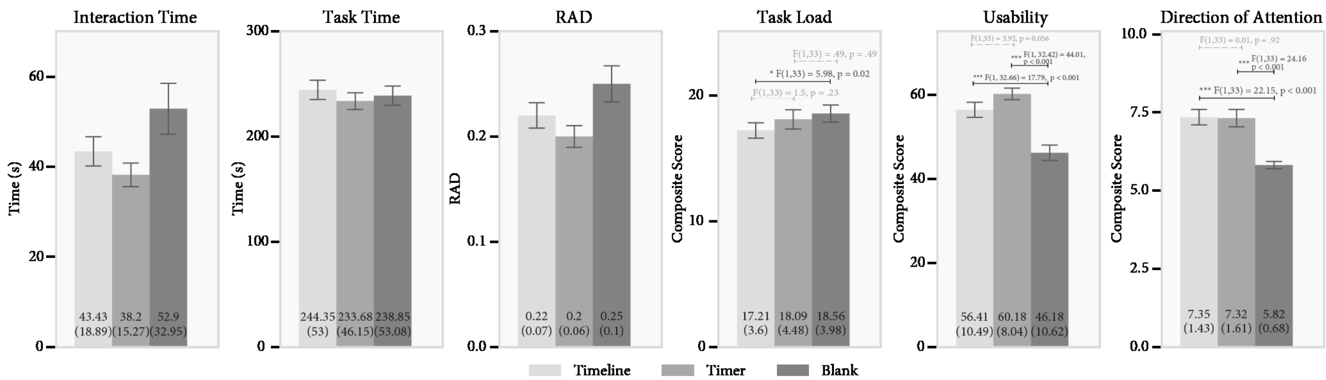

Fig. 5. Interaction time, task time, RAD, task load, usability, and direction of attention by three user interface conditions: Timeline, Timer, and Blank. Error bars denote standard error. Descriptive statistics shown as Mean (SD) in each bar. Representative *a priori* contrast analysis results are shown within figure. Significance level: $^{\cdot}$ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$.

C. Measures and Analysis

The pre-study questionnaire included the Multimedia Multitasking Index (MMI) [29] to measure participants' perceptions of their own multitasking behavior. Participants were given as much time as needed to complete both the *assembly-inspection* task and the *sorting* task, and total time to complete the task was recorded. The post-trial questionnaire included the NASA Task Load Index (TLX) [18], System Usability Scale (SUS) [2], [4], NASA Situational Awareness Rating Technique (SART) [35], and the two-item scale of direction of attention adapted from Rae *et al.* [32]. Direction of attention scale asks participants to rate their ability to keep track of the environment and whether direction of attention impacted the task. The semi-structured interview included questions about strategies for managing the robot, awareness of the robot's state, perceptions of and opinions on the interface and task, challenges with the task, and any suggestions for design or alternate interface modalities.

We tested gender, age, and ethnicity as potential covariates and found none to have a significant effect at an α level of 0.05 ($p > 0.05$). A linear mixed-effects regression model was constructed with interface condition as the input variable, MMI as a covariate, and usability, taskload, and direction of attention as response variables. We found that MMI was not a significant covariate in these models. Additionally, we constructed a linear mixed-effects regression model with interface condition and centered MMI as input variables and total task time, total *RAD*, and total interaction time as response variables, respectively. We conducted *a priori* contrast analyses to make comparisons indicated by our hypothesis. Situation awareness was not analyzable due to insufficient responses from the participants. In calculating MMI, we omitted the other-computer-applications primary matrix (a usage subscale) due to technical issues with the questionnaire. We also included condition as a by-item random-effects factor in all models. All analyses were conducted in R.

Semi-structured interviews were transcribed by four trained experimenters. All analyses of qualitative data were conducted following a Ground Theory Approach [17].

V. RESULTS

A. Quantitative Results

We provide a textual description of the results below and main-effects-test details in Table I. Figure 5 shows descriptive statistics and results from *a priori* contrast analyses.

H1—Our data did not provide support for *H1*. We found no main effect of user interface condition on total interaction time, total task time, or task *RAD*. However, our analysis found a marginal main effect of MMI on total interaction and a significant Pearson's product-moment correlation of 0.24 between centered MMI and total interaction time, $t(88) = 2.37$, $p = .02 < 0.05$. There was no main effect of centered MMI on task *RAD*. We found a significant Pearson's product-moment correlation of 0.24 between task *RAD* and centered MMI, $t(88) = 2.31$, $p = .02 < 0.05$.

H2—Our analysis offered partial support for *H2*. We found a significant main effect of user interface condition on perceived task load. Participants rated task load significantly lower in the timeline condition than in the blank condition.

H3—There was also partial support for *H3*. We found a significant main effect of user interface condition on usability. Participants rated usability to be significantly higher in the timeline and timer conditions compared to the blank condition. There was a significant main effect of user interface condition on direction of attention. Participants rated direction of attention to be significantly higher in the timeline and the timer conditions than the blank condition.

TABLE I

MAIN EFFECTS FROM EACH LINEAR MIXED REGRESSION MODEL. SIGNIFICANCE LEVEL: $^{\cdot}$ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$.

Measure	*b*	*se*	*df*	*F*	*p*	*d*
Interface Condition						
Interaction Time	10.04	6.65	(1, 29)	2.28	.14	0.17
Task Time	–537	10.81	(1, 29)	0.25	.62	–0.04
Task *RAD*	0.028	0.02	(1, 29)	1.99	.17	0.14
Task Load	1.35	0.65	(1, 33)	4.30	.046*	0.14
Usability	–10.10	2.47	(1, 32.69)	16.67	< .001***	–0.36
Direction of Attention	–1.53	0.32	(1, 33)	22.15	< .001***	–0.37
MMI						
Interaction Time	4.41	2.38	(1, 28)	3.20	.085$^{\cdot}$	0.19
Task Time	4.46	6.81	(1, 28)	0.40	.53	0.09
Task RAD	0.015	0.0091	(1, 28)	2.70	.11	0.20

H4—We found no support for *H4*. No main effect of user interface condition found for total interaction time. There was no significant difference between participants in timeline and timer conditions in task load.

B. Qualitative Results

Four themes—*strategies*, *influences*, *performance*, and *preferences*—emerged from the semi-structured interviews. Response totals are summarized in Table II.

Strategies—Participants used sound, visual cues, and the interface in their strategies to monitor the robot while also performing their own task. Participants (23 Blank, seven Timeline, seven Timer) reported that they were *"listening for noise"* (P03) from the robot. Alternatively, seven participants stated that they did not listen to the robot or stopped listening, *e.g., "I think I kind of used sound less"* (P22), throughout the session. Participants (15 Blank, 10 Timeline, three Timer) also mentioned that they *"glance[d] over"* (P12) toward the robot to check its status. Notably, P05 and P08 expressed that the robot *"always stop[ped] in kind of a position"* (P08) when it was done. Six participants mentioned that they *"didn't even bother looking at that <gestures at robot>"* (P09) robot or looked *"a little bit but not as much as the first time"* (P29).

Influences–Participant's strategies were influenced by their previous experiences, demand on their attention, and trust toward the robot. Participants appeared to learn the workflow and develop a sense of process duration from their previous trials. P08 stated, *"I was a little more familiar with how I was supposed to work"* and *"how much time I actually had to sort."* Three participants also reported having borrowed from their experience in manufacturing. Seven felt that the interfaces added to the demand for their attention. For instance, three participants had to prioritize the robot's task to *"make sure that [they] didn't keep the robot waiting"* (P26).

Participants also stated that their strategies were influenced by their level of trust toward the robot, the interface, or the interaction button. Five participants explicitly mentioned trusting the robot, *e.g., "[they] knew the robot would always work"* (P16) (one Blank, two Timeline, three Timer), while others were *"waiting for it to make a mistake"* (P31). Likewise, some participants (five Timeline, three Timer) trusted the interface. For instance, P11 *"never really doubted what the screen was telling"* them even though *"it could have been doing something completely wrong."* As for the interaction button, four participants discussed the lack of immediate feedback, *"it took a little bit to react"* (P03). Two participants stated that they *"glanced over a couple times"* (P12), as they did not know *"if [they] actually pressed it"* (P12). P15 and P30 recommended that *"when [they] pressed the button, something on the screen should appear"* (P30). Only P35 indicated trusting the button, they *"knew eventually it would work,"* as *"the timer would start again."*

Performance—Participants indicated that their performance was affected by their strategy during the task. Some participants (three Blank, one Timeline, one Timer) mentioned errors during the task, *e.g., "put wrong pieces in the wrong bucket"* (P09), while others (four Blank, three Timer) expressed uncertainty. For instance, P28 did not *"quite know how many more of [steps] there [we]re."* Several participants (five Blank, two Timeline, six Timer) stated that they felt that a specific trial was *"efficient"* (P25) or *"faster"* (P23). Six participants mentioned trials in the blank or the timer condition were *"stressful"* (P26). Seven participants suggested improvements by making changes to the workspace layout, *e.g., "rotate the robot"* (P27) and *"do the job remotely"* (P30). In all trials, the robot successfully constructed the house regardless of whether the participant adjusted it during assembly.

Preferences–We asked participants for their interface preference during the interview. One participant preferred the blank condition; 10 preferred the timeline; and 18 preferred the timer. Four participants expressed that their preferences varied based on context. For instance, P07 preferred the timeline *"in an industrial setting"* and timer *"if you really wanted to crank out parts,"* whereas P22 was split between timer for a *"long period of time"* as it is *"less like strenuous, like more straightforward"* and blank condition as it *"felt like I could focus more on sorting than looking at the screen."*

For the timeline condition, a few participants favored it because it provided *"a big picture overview"* (P15); and it was *"easy to interpret"* (P28) and quick to learn (P26). However, nine participants disliked the interface because it was *"distracting"* (P08) or *"complicated"* (P27). Six participants expressed that it was *"unnecessary to have the whole process [presented]."* Four participants suggested that if the task *"was more complicated or if it was longer, then, this one [the timeline] might have been okay"* (P33).

Sixteen participants described the timer interface as being *"extremely helpful"* (P02), *"the easiest"* (P08), or *"simpler"* (P13), as it made them aware of the time (P05). However, five participants did not *"really think [that the interface was] needed because [they could] always hear the sound"* (P27) or were *"not sure if that information [wa]s necessary"* (P21).

Participants found the traffic-light metaphor useful (eight Timeline, 11 Timer), especially *"in a factory environment"* (P35). With the color-coding warnings, eleven participants reported using peripheral vision—*"corner of [their] eye"* (P19)—to look at the display in the timer condition, whereas they had to be *"looking directly at the screen"* (P22) to see the warning for the timeline. P22 stated *"coloring on the screen was probably the most helpful."*

Finally, participants made several suggestions for improving user experience. P02, P10, and P29 suggested *"a combination"* (P10) of the timer and timeline interfaces. P05 and P31 wanted

TABLE II

UNIQUE PARTICIPANT RESPONSE COUNTS FOR MAJOR THEMES.

	Strategies	***Influences***	***Performance***	***Preferences***
Timeline	33	24	5	10 + 3*
Timer	32	19	17	18 + 3†
Blank	26	23	12	1 + 2‡

	Screen	*Sonification*	*Performance*	*Attention*	*Button*
Suggestions	9	9	7	6	2

* Or *Timer* (two); Or *Blank* (one).
† Or *Timeline* (two); Or *Blank* (one).
‡ Or *Timeline* (one); Or *Timer* (one).

more robot-specific information, *e.g., "a 2D representation of this robot"* (P31). P09 and P34 wanted more dynamic human-centered information, *e.g., "to track personal progress"* (P09). Another frequent requested change (nine participants) was to consider sonification instead of, or in addition to, the visualization, such as *"beep"* (P34) for an audio notification.

VI. DISCUSSION

This paper presented *pRAD* for cobot operators. Our evaluation found that participants incorporated auditory, visual, and interface feedback into their monitoring strategies according to their modality preferences, trust, and experience. These findings align with those from prior ethnographic work that reported similar listening and visual-inspection behaviors [33]. Our study also found that participants rated both interfaces as being significantly more usable than the baseline (Blank) condition. However, *pRAD* interfaces did not affect interaction time or task time. While performance was consistent across conditions, the reported task load was significantly lower for the timeline condition relative to baseline. This finding suggests that the additional overview information provided by the timeline helps participants form a better mental model of the task, resulting in reduced cognitive load. However, our qualitative data also suggests that some participants felt that both interfaces were confusing or distracting. Further work to explore user experience when incorporating such interfaces into a larger system is still needed.

Future work should explore *pRAD* interfaces to support the operation of multiple cobots, specifically to help operators manage increased task load and account for errors when they misinterpret robots' needs. Saturation of the operator's workload can result in a queue of robots will be waiting for interaction, as was found by Glas *et al.* [12]. In our study, MMI was found to be significantly correlated with interaction time suggesting that media-multitasking skills are applicable to monitoring the robot's state. Selecting the number of robots that an operator is assigned to monitor could be informed by their reported MMI.

A. Limitations

The work presented here has a number of limitations. First, we speculate that the lack of support for our hypotheses regarding change in task time is due to our evaluation protocol allowing participants to self-pace their work, which may have resulted in too much variance to observe the effects of providing *pRAD* information. Alternative task instructions, *e.g.,* asking participants to sort as fast as possible or to always prioritize the robot, may help identify such performance effects. However, such strategies might also conflict with participants' prior experience in industry. Allowing open-ended strategy exploration, captured through semi-structured interviews, generated rich insight into their interaction with the interfaces and the cobot.

Generalizability is also a limitation of this work. We developed our scenario following empirical findings by Michaelis *et al.* [23] where experts utilize cobots to perform simple repetitive tasks, e.g., automated assembly, in lieu of traditional automation. We conducted the study in a controlled laboratory setting. Real-world manufacturing environments involve ambient auditory noise, visual clutter, and interruptions by other workers. Additionally, tasks and workcell configurations will differ from our specific scenario. The color palette used in the design of our prototypes did not account for color-blindness. Our interfaces should adopt a more accessible color palette. Alternatively, icons (Timeline) or text (Timer) may be used to convey warnings in situations where color is ineffective.

We generated the *pRAD* signal from analysis of the task performed in a pilot study. As an alternative approach, an engineer could use a planning tool, *e.g.,* Authr [34] or Tercio [13], to generate the *neglect time* and *interaction time* for a fully specified plan so long as timing is consistent between cycles. For tasks where timing is dynamic, more sophisticated models of human and robot work is necessary, *e.g.,* using POMDPs [15], [41] to estimate predicated *neglect time* and *interaction time* during task execution.

Finally, we highlight the safety implications of asking the operator to multitask. Some secondary tasks could be dangerous if robot prioritization was important, *e.g.,* operator is performing a separate soldering task. Other complex tasks could be ineffective as the divided attention diminishes the performance on the secondary task due to increased cognitive load. Our interface designs were for short-cycle inspection tasks with an interruptable, low-cognitive-demand secondary sorting task. Using a *pRAD* signal may not generalize to more complex or safety-critical secondary tasks. Future work should address real-time monitoring and task modeling to infer when interruption is appropriate for *pRAD-based* designs.

B. Design Implications

At a high level, our results suggest that *pRAD* interfaces reduce cognitive load. Our study, particularly our analysis of the participant feedback for future designs, provides additional design implications, which we discuss below.

Several participants suggested combining the timeline and timer interfaces to provide the user both an overview of the task and the precise timer representation. Users should be able to change the prominence of the timeline and timer components to customize the interface based on the current priority, *e.g.,* getting an overview (Timeline-focused), prioritizing the robot (Timer-focused), or focusing on their work (Blank). The design should continue to provide the color-coded traffic-light metaphor with emphasis on being seen in the user's peripheral vision. Another opportunity is to visually augment the interface with notification badges [22]. These badges can indicate a change in predicted timing or communicate the priority of the robot relative to the current task when the interface transitions from neglect to interaction.

Participants also suggested sonification, such as a simple *"beep"* played at key milestones of the task. In noisy workplaces, operators could be augmented with bone conduction headphones. Implicit *pRAD* cues from the robot's joint movement noise could be mimicked for sonification [11] through headphones, e.g., serving as functional noise [21].

VII. CONCLUSION

In this paper, we investigated the applicability of *RAD* in the context of collaborative robotics as a predictive signal, called *pRAD*, used as a design component in operator interfaces. We described the process to generate a *pRAD* signal and prototyped two user interfaces to display aspects of this signal. We then conducted a user study to evaluate the effects of displaying *pRAD* on task and collaboration outcomes. Our findings highlight the usability benefits of *pRAD* in cobot operator interfaces. Importantly, these interfaces should explicitly communicate "collaborative capacity" and implicitly convey it through the robot's start-and-stop behavior.

ACKNOWLEDGMENT

We would like to thank Bengisu Cagiltay, Pragathi Praveena, Jamie Wage, Andrew Langbehn, Amelia Boruch, Zhuanghan Dong, Allie Helein, Ava Bell, and Justin Lubben for their assistance. This work was supported by National Science Foundation awards 1426824 and 1822872.

REFERENCES

[1] H. Ahmad, M. F. Khalid, R. Kandan, M. N. M. Mydin, B. I. Ismail, and O. H. Hoe. A unified dashboard for collaborative robot management system. In *P. of SCOReD*, pages 5–9, 2020.

[2] A. Bangor, P. T. Kortum, and J. T. Miller. An empirical evaluation of the system usability scale. *Int. J. Hum-Comput. Int.*, 24(6):574–594, 2008.

[3] S. Bragança, E. Costa, I. Castellucci, and P. M. Arezes. *A Brief Overview of the Use of Collaborative Robots in Industry 4.0: Human Role and Safety*, pages 641–650. Springer, 2019.

[4] J. Brooke. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

[5] J. Y. C. Chen, M. J. Barnes, and M. Harper-Sciarini. Supervisory control of multiple robots: Human-performance issues and user-interface design. *IEEE Trans. Syst. Man. Cybern.*, 41(4):435–454, July 2011.

[6] A. S. Clair and M. Mataric. How robot verbal feedback can improve team performance in human-robot task collaborations. In *P. of HRI*, pages 213–220, New York, NY, USA, 2015. ACM.

[7] J. W. Crandall and M. L. Cummings. Developing performance metrics for the supervisory control of multiple robots. In *P. of HRI*, pages 33–40, New York, NY, USA, 2007. ACM.

[8] M. L. Cummings and P. J. Mitchell. Predicting controller capacity in supervisory control of multiple uavs. *IEEE Trans. Syst. Man. Cybern.*, 38(2):451–460, March 2008.

[9] M. R. Elara, C. A. Acosta Calderon, C. Zhou, and W. S. Wijesoma. False alarm demand: A new metric for measuring robot performance in human robot teams. In *P. of ICRA*, pages 436–441, 2009.

[10] M. R. Elara, C. A. A. Calderon, and C. Zhou. Extended neglect tolerance model amp; human robot teams. In *P. of ICICS*, pages 1–6, Dec 2011.

[11] E. Frid, R. Bresin, and S. Alexanderson. Perception of mechanical sounds inherent to expressive gestures of a nao robot-implications for movement sonification of humanoids. In *P. of SMC*, 2018.

[12] D. F. Glas, T. Kanda, H. Ishiguro, and N. Hagita. Teleoperation of multiple social robots. *IEEE Trans. Syst. Man. Cybern.*, 42(3):530–544, May 2012.

[13] M. C. Gombolay, R. J. Wilcox, and J. A. Shah. Fast scheduling of robot teams performing tasks with temporospatial constraints. *IEEE Trans. Robot.*, 34(1):220–239, 2018.

[14] M. A. Goodrich and D. R. Olsen. Seven principles of efficient human robot interaction. *IEEE Trans. Syst. Man Cybern.*, 4:3942–3948 vol.4, Oct 2003.

[15] N. Gopalan and S. Tellex. Modeling and solving human-robot collaborative tasks using pomdps. In *P. of RSS*, volume 32, pages 590–628, 2015.

[16] V. Gopinath, F. Ore, S. Grahn, and K. Johansen. Safety-focussed design of collaborative assembly station with large industrial robots. *Procedia Manufacturing*, 25:503 – 510, 2018.

[17] J. F. Gubrium, J. A. Holstein, A. B. Marvasti, and K. D. McKinney. *The SAGE handbook of interview research: The complexity of the craft*. Sage Publications, 2012.

[18] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[19] J. Hoecherl, T. Schlegl, T. Berlehner, H. Kuhn, and B. Wrede. Smartworkbench: Toward adaptive and transparent user assistance in industrial human-robot applications. In *P. of ISR*, pages 1–8, June 2016.

[20] G. Hoffman. Evaluating fluency in human–robot collaboration. *IEEE Trans. Hum. Mach. Syst.*, 49(3):209–218, 2019.

[21] M. Joosse, M. Lohse, and V. Evers. Sound over matter: the effects of functional noise, robot size and approach velocity in human-robot encounters. In *P. of HRI*, pages 184–185, 2014.

[22] A. Matallaoui. Towards more effective gamification: Does deploying semiotics help design better perceivable badges? In *P. of ICCTA*, pages 131–135, 2018.

[23] J. Michaelis, A. Siebert-Evenstone, D. Shaffer, and B. Mutlu. Collaborative or simply uncaged? understanding human-cobot interactions in automation. In *P. of CHI*, 2020.

[24] R. Mohan, W. S. Wijesoma, C. A. A. Calderon, and C. Zhou. Validating extended neglect tolerance model for human robot interactions in humanoid soccer robots. *Robotica*, 29(3):421–432, 2011.

[25] M. Morioka and S. Sakakibara. A new cell production assembly system with human–robot cooperation. *CIRP Annals*, 59(1):9 – 12, 2010.

[26] P. Neto, M. Simão, N. Mendes, and M. Safeea. Gesture-based human-robot interaction for human assistance in manufacturing. *Int. J. Adv. Manuf. Technol.*, 101(1):119 – 135, Mar 2019.

[27] D. R. Olsen and M. A. Goodrich. Metrics for evaluating human-robot interactions. In *P. of PERMIS*, 2003.

[28] D. R. Olsen and S. B. Wood. Fan-out: Measuring human control of multiple robots. In *P. of CHI*, pages 231–238, New York, NY, USA, 2004. ACM.

[29] E. Ophir, C. Nass, and A. D. Wagner. Cognitive control in media multitaskers. *Proc. Natl. Acad. Sci. U.S.A.*, 106(37):15583–15587, 2009.

[30] M. Pearce, B. Mutlu, J. Shah, and R. Radwin. Optimizing makespan and ergonomics in integrating collaborative robots into manufacturing processes. *IEEE Trans. Autom. Sci. Eng.*, 15(4):1772–1784, 2018.

[31] A. Perzylo, M. Rickert, B. Kahl, N. Somani, C. Lehmann, A. Kuss, S. Profanter, A. B. Beck, M. Haage, M. R. Hansen, M. T. Nibe, M. A. Roa, O. Sornmo, S. G. Robertz, U. Thomas, G. Veiga, E. A. Topp, I. Kessler, and M. Danzer. Smerobotics: Smart robots for flexible manufacturing. *IEEE Robot. Autom. Mag.*, 26(1):78–90, March 2019.

[32] I. Rae, B. Mutlu, and L. Takayama. Bodies in motion: mobility, presence, and task awareness in telepresence. In *P. of CHI*, pages 2153–2162, 2014.

[33] A. Sauppé and B. Mutlu. The social impact of a robot co-worker in industrial settings. In *P. of CHI*, page 3613–3622, New York, NY, USA, 2015. ACM.

[34] A. Schoen, C. Henrichs, M. Strohkirch, and B. Mutlu. Authr: A task authoring environment for human-robot teams. In *P. of UIST*, pages 1194–1208, 2020.

[35] R. M. Taylor. Situational awareness rating technique (sart): The development of a tool for aircrew systems design. In *P. of the AGARD AMP Symp. SAAO*. Seuilly-sur Seine: NATO AGARD, 1989.

[36] P. Tsarouchi, S. Makris, George Michalos, A. Matthaiakis, X. Chatzigeorgiou, A. Athanasatos, M. Stefos, P. Aivaliotis, and G. Chryssolouris. Ros based coordination of human robot cooperative assembly tasks-an industrial case study. *Procedia CIRP*, 37:254 – 259, 2015.

[37] P. Tsarouchi, A. Matthaiakis, S. Makris, and G. Chryssolouris. On a human-robot collaboration in an assembly cell. *Int. J. Compit. Integr. Manuf.*, 30(6):580–589, 2017.

[38] P. Velagapudi, P. Scerri, K. Sycara, H. Wang, M. Lewis, and J. Wang. Scaling effects in multi-robot control. In *P. of IROS*, pages 2121–2126, Sep. 2008.

[39] J. M. Whetten, M. A. Goodrich, and Y. Guo. Beyond robot fan-out: Towards multi-operator supervisory control. *IEEE Trans. Syst. Man Cybern.*, pages 2008–2015, Oct 2010.

[40] F. Zhao, C. Henrichs, and B. Mutlu. Task interdependence in human-robot teaming. In *P. of RO-MAN*, pages 1143–1149. IEEE, 2020.

[41] W. Zheng, B. Wu, and H. Lin. Pomdp model learning for human robot collaboration. pages 1156–1161, 2018.