Applied Ergonomics 67 (2018) 91–103

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

Enhancing the effectiveness of human-robot teaming with a closed-loop system

Grace Teo [a,*], Lauren Reinerman-Jones [a], Gerald Matthews [a], James Szalma [b], Florian Jentsch [b], Peter Hancock [b]

[a] Institute for Simulation and Training, University of Central Florida, Orlando, USA
[b] Department of Psychology, University of Central Florida, Orlando, USA

ARTICLE INFO

Article history:
Received 15 August 2016
Received in revised form
12 June 2017
Accepted 19 July 2017
Available online 3 October 2017

Keywords:
Adaptive systems
Workload model
Inter-individual variability

ABSTRACT

With technological developments in robotics and their increasing deployment, human-robot teams are set to be a mainstay in the future. To develop robots that possess teaming capabilities, such as being able to communicate implicitly, the present study implemented a closed-loop system. This system enabled the robot to provide adaptive aid without the need for explicit commands from the human teammate, through the use of multiple physiological workload measures. Such measures of workload vary in sensitivity and there is large inter-individual variability in physiological responses to imposed taskload. Workload models enacted via closed-loop system should accommodate such individual variability. The present research investigated the effects of the adaptive robot aid vs. imposed aid on performance and workload. Results showed that adaptive robot aid driven by an individualized workload model for physiological response resulted in greater improvements in performance compared to aid that was simply imposed by the system.

1. Introduction

The population of robots in the world reached 8.6 million in 2010 (Guizzo, 2010). In 2016, the global robotics market was worth $25.9 billion USD. This is expected to reach $31.5 billion USD in 2021, expanding at a compound annual growth rate of 4.0% (Wilson, 2016). Although traditionally robots have been assigned tasks that are typically “dirty, dangerous, and dull” (Takayama et al., 2008), in recent years, increased robot functionality has resulted in their deployment in a greater variety of domains such as in bomb disposal, search and rescue missions, manufacturing (Guizzo and Ackerman, 2012), as surgical robots in healthcare (da Vinci surgery, 2013), and as robot assistants for the elderly or disabled (Kumar et al., 2006) in which their more intelligent contributions are now being mandated.

Despite these advances, robots are still largely teleoperated via remote control and require explicit commands, which confines their use to relatively structured tasks. To enable robots to operate in more novel environments, performing less structured tasks, robots must be capable of richer human-robot communications that may approach that exhibited in human teams (DRC, 2013). Robots with such capability should be more responsive to humans and exhibit behaviors that approximate teaming, including sensing the human operator's psychological status such as experienced workload and fatigue. One possible strategy for enhancing capability is through the use of a closed-loop system that adapts the robot to provide appropriate support when the human becomes overloaded (see Hancock and Chignell, 1988).

1.1. New technology for human-robot communication

A closed-loop system uses feedback or error signals to drive corrective actions that maintain a desired system state (homeostasis). Such systems have been in existence for several centuries, with many modern examples (e.g., thermostats, cruise control in cars). However their use in the human-robot teaming context is relatively recent. A closed, feedback loop, using measures of the human operator's workload as input, could allow selection of robot aiding behaviors that maintain the operator's workload state at a moderate target level (see Hancock and Chignell, 1987). Hence, if the operator is experiencing high workload to the point that

* Corresponding author. Institute for Simulation and Training, 3100 Technology Parkway, Orlando, FL 32826, USA.
E-mail address: gracet@knights.ucf.edu (G. Teo).

http://dx.doi.org/10.1016/j.apergo.2017.07.007
0003-6870/

jeopardizes his/her performance, the robot can aid by relieving the operator of certain tasks. Robot teammates' behavioral repertoire can include aiding with subtasks without being instructed to do so, such as anticipating needs by appropriately escalating information, and initiating actions such as messaging for help instead of requiring the operator to do so him/herself. When operator workload state returns to more manageable levels, the robot can adapt its behaviors to return full task control to the human teammate. Adaptive reversion of task control to the human teammate may minimize the often-cited problems associated with having the human "out-of-the-loop" such as the loss of situational or system awareness, increased complacency, over-reliance issues, skill atrophy, performance degradations, and unbalanced mental workload (Carmody and Gluckman, 1993; Endsley and Kiris, 1995; Parasuraman and Wickens, 2008; Smith and Hancock, 1995).

Critically, the robot must be capable of detecting this human overload implicitly, without overt instructions. Implicit communications is crucial where the operator is unable to issue explicit instructions. For instance, the operator may be experiencing such high workload but task demands prevent him or her from being able to instruct the robot on how to assist. It is also possible that operators may not be aware of their own workload state when intensely engaged in the task at hand.

The closed-loop system therefore requires a workload model that assesses and classifies operator workload without operator input. Subjective workload responses tend to disrupt performance as they require the operators' explicit reports. In contrast, physiological measures of workload allow continuous assessment, provide high temporal resolution, and rarely require any disruptive overt response from the operator. Thus, they are particularly suitable as indicators of operator workload in adaptive systems (Byrne and Parasuraman, 1996; Hancock and Chignell, 1987, 1988). Despite these advantages, physiological assessments also have limitations (Cain, 2007). Depending on the measure, these may include lower sensitivity to task demands relative to subjective scales, or sensitivity to certain characteristics of taskload only. Measures may also be contaminated by general stress responses. Conceptual linkages from physiology to performance may also be insufficiently specified. Nevertheless, developments in recording and processing physiological signals, together with accumulating evidence for validity, have heightened interest in the physiological approach (e.g., Chen and Barnes, 2014).

The basis of physiological workload measures (e.g., heart-rate, ocular activity, brain activity, hemodynamics) lies in the notion that, with the activation of various cognitive processes required to process task demands and execute the required responses, there are correspondng physiological responses. Commonly-used measures include heart rate, heart rate variability, respiration rate, brain activity, pupil size (diameter), and electrodermal activity, among many others. Rationales for specific measures may be found in reviews by Abich (2013), Borghini et al. (2014), Warm et al. (2012), Meshkati et al. (1995), and Young et al. (2015). For the present study, the physiological workload measures selected address responses in both central (i.e., brain activity, cerebral perfusion as indicated by level of oxygen saturation, and cerebral bloodflow velocity) and peripheral responses (i.e., cardiac and ocular responses) that index such cognitive activity.

1.2. Development of a workload model that accommodates individual variability in physiological responses

A challenge for ergonomic applications is the complexity of the neuropsychological workload construct (e.g., Young et al., 2015). Different metrics for workload may dissociate from one another, and from performance as task demands change (Hancock and Scallen, 1996; Horrey et al., 2009; Szalma and Teo, 2012). In the adaptive aiding context, it is essential to distinguish (1) objective external task demands (which we call "taskload"), (2) objective performance, and (3) workload as subjective and physiological indicators reflective of operator neurocognitive state. In some circumstances, loss of performance may be used to drive an automated aid directly, without the need for workload assessment. By contrast, use of workload rather than performance as the driver may be more effective in contexts in which (1) it is difficult to monitor performance continuously, (2) performance is influenced by multiple factors, and/or (3) it is important to anticipate future performance degradation as initial compensation for high task load becomes increasingly difficult (Cain, 2007; Hancock and Warm, 1989).

However, workload-driven adaptive aiding will only be effective if there is a negative taskload – performance association, so that mitigating taskload enhances performance. There are several circumstances in which taskload dissociates from performance. At moderate levels of demand, people often compensate for changing taskload levels to maintain constant performance, although low workload appears to be especially hard to manage (Hancock and Warm, 1989; Saxby et al., 2013) and may contribute to loss of situation awareness (Young and Stanton, 2007). Indeed, workload may reflect the operator's strategies for active management of task demands, strategies that may change dynamically during the course of performance (Hockey, 1997; Saxby et al., 2013). Especially in real-life settings, high workload may be experienced as enjoyably challenging and motivating (Matthews, 2016), potentially leading to positive workload-performance associations (Abich et al., 2017). Thus, workload is primarily useful for driving aiding or other automation in task settings that produce congruent reactions indicative of overload: subjective workload and stress, little strategic compensation, and performance impairment.

Even within the subset of task environments in which workload is diagnostic of performance, there are assessment challenges, e.g., different workload measures do not always concur. Workload may be assessed with subjective scales such as the NASA-TLX (Hart and Staveland, 1988), but such measures do not adequately capture physiological response (Matthews et al., 2015). Different forms of taskload, such as working memory demands, multi-tasking, and signal salience may all provoke feelings of overload. However, different taskload factors may elicit different patterns of physiological responses so that an algorithm based on responses to increasing working memory demands, for example, might not be effective in driving an adaptive system for handling multi-tasking.

A further major challenge associated with the use of physiological workload measures involves the large inter-individual variability of such responses (e.g. Hancock et al., 1985; Moray, 1984; Meshkati and Loewenthal, 1988; Roscoe, 1993; Johannes and Gaillard, 2014). There are multiple, weakly-correlated workload responses associated with indices of autonomic and central nervous sytem functioning. For example, one individual might show a strong electroencephalographic (EEG) response but a weak electrocardiac response, whereas another person might show the opposite pattern (Matthews et al., 2015). The workload model driving the closed-loop system would need to accommodate this inter-individual variability in physiological responses to workload across multiple measures in order to be fully effective.

The present study used a task environment represented by a simulation of an unmanned vehicle operation. This met the criteria we have defined for application of workload-driven adaptive aiding. The participant monitored one or more computer-screen windows for critical signals. The task imposed relatively high event rates to make the task attentionally demanding and limit possible strategic compensation. Taskloads are not so low as to impair situation awareness or effort-regulation (Hancock and Warm, 1989; Saxby et al., 2013). Abich et al. (2017) compared single- and dual-task performance and found that dual-tasking induced a consistent pattern of outcome change indicative of overload as demonstrated by substantial performance impairment, increased subjective workload, and increased distress.

Data from two previous studies that have used the same task and context as the present study were exploited to develop the workload model (Abich, 2013; Matthews et al., 2015; Teo et al., 2016). Data from one of these studies served as the training dataset, while data from the other served as the validation dataset. Both studies provided data on individuals' physiological workload responses, as well as subjective, and performance workload measures during both single- and dual-task conditions.

The first study (training dataset) used a Change Detection Task (CDT) that required icon change recognition on a map, and a Threat Detection Task (TDT) that required detecting threat characters in a video feed. The second study (validation dataset) consisted of an entirely different sample of participants but featured the same CDT with the same task parameters. However, instead of detecting threat characters, the second task required participants to answer auditory prompts about characters in the video feed. In both studies, tasks were performed in a simulation of a military mission undertaken by a Soldier and an Unmanned Ground Vehicle (UGV).

Analyses of these studies demonstrated that "low" and "high" workload, as assessed by various physiological and subjective workload measures, could be reliably induced by single- (low taskload) and dual- (high taskload) tasks respectively, indicating a positive relationship between taskload and workload. Results also showed an inverse relationship between performance and both psychological and physiological workload measures. As well as higher subjective workload, elevated taskload also induced a range of physiological workload responses including lower heart rate variability, higher EEG frontal theta activity, shorter duration eye fixations, and higher cognitive activity as measured by pupillometry (Abich, 2013; Matthews et al., 2015).

The workload model, developed from the training dataset, involved comparing two sets of change scores from the individual. The first set of change scores ("baseline difference score") was the individual's change in physiological workload responses induced by the first pair of single-task (low workload) and dual-task (high workload). The second set of change scores ("test difference score") reflected the change in the individual's physiological workload responses computed from the same single-task and a new task. If the "baseline difference score" and "test difference score" matched, then the new task was considered to have induced the same high workload as the dual task. This enabled the workload of a new task to be classified on an individualized basis. A classification of "high workload" could then be used to trigger adaptive robot aid. When the workload model was tested with the data from the second study (validation dataset), the conditions in that study served as the "new tasks". Results showed that the dual task condition was more likely than the single task condition to trigger the adaptive aid, validating the workload model.

In the process of workload modeling, an algorithm was developed to combine the workload data from multiple physiological measures to derive a single workload index. To enable data from the various physiological measures to be combined, the scores from the different physiological measures were first converted to standard z-scores (using the group data from previous studies which used the same tasks and task parameters). To accommodate the large inter-individual variability in physiological workload responses, the algorithm entailed identifying the individual's own specific set of physiological workload markers. These markers were the physiological workload measures sensitive to the individual's workload changes between the low (i.e., single-task) and high (i.e., dual-task) taskload conditions (see Teo et al., 2016). Workload data on the individual's markers were then used to compute the workload index for the individual. Hence, the workload index of different individuals were computed from different sets of physiological measures. The validation data set confirmed that the workload model and algorithm selected proved robust for use in this present study (see Teo et al., 2016).

1.3. Research hypotheses

Thus, the objective of this present study was to determine the effects of adaptive aiding, based on multiple physiological measures of workload, on performance under low and high taskloads. A closed-loop system, based on a workload model previously described, drove the adaptive aiding. Adaptive aid to change detection performance was triggered when the system classified the participant as experiencing high workload. For participants who did not trigger adaptive aid, an imposed aid was given in the last block. Specific hypotheses were as follows:

Hypothesis 1. **There would be a positive relationship between taskload and workload, both of which would be inversely related to performance**. Previous studies of such simulated UGV environment as used here (i.e., Abich, 2013) showed that a dual-tasking manipulation (taskload) elevated workload, on multiple physiological and subjective measures, and also impaired performance. Since the present study used the same tasks and task parameters, we expected that these relationships would be confirmed. However, this manipulation check was necessary where defining the workload model depended on the single and dual task conditions eliciting low and high workload respectively.

Hypothesis 2. **Performance would improve with aid.** The aid was an auditory "beep" that occurred at the same time as each change event on the change detection task. We hypothesized that performance would improve with the provision of robot aid of any kind, which would act to relieve workload by reducing taskload. Benefits of aid were anticipated to increase with taskload; i.e., greater performance enhancement in dual- rather than single-task conditions.

Hypothesis 3. **Performance improvements from the aid would be greater for those provided adaptive aid compared to those upon whom the aid was imposed**. Results from previous studies with the same tasks (i.e., Abich et al., 2017) showed that high workload adversely impacted performance, consistent with an insufficiency of attentional resources (Matthews et al., 2000). Provision of the aid should thus reduce taskload, and hence associated workload. So, the adaptive aid should be more beneficial for those initially experiencing high workload compared to those with manageable workload. Aid that is adapted to the individual's own level of workload (i.e,. adaptive aid) should result in greater performance improvements compared to aid that disregards the individual's workload level (i.e., imposed aid).

2. Method

2.1. Study scenarios

There were four study scenarios, two of which served as baseline scenarios, and two which were experimental scenarios: (i) Scenario S, (ii) Scenario D, (iii) Scenario SSS, and (iv) Scenario SDD (see Fig. 1).

Both single-task (i.e., low taskload) and dual-task (i.e., high taskload) scenarios were included to confirm that dual-tasking did result in higher workload and poorer performance (Hypothesis 1). Baseline scenarios S and D were 5 min each in duration; no aid was provided. Data from these conditions allowed the individual's physiological workload markers to be identified, as described in 2.4 below. The two experimental scenarios, Scenario SSS and Scenario SDD, were each 15 min in duration.

Scenario SSS required participants to perform a single-task (change detection) throughout, while Scenario SDD began as a single-task (change detection) for the first 5 min before taskload was increased to a dual-task (threat detection added) for the remaining 10 min. For each experimental scenario, adaptive aid was provided in the first 10 min if the individual's workload index reached threshold. Once triggered, the adaptive aid was not revoked for the rest of the scenario. If no adaptive aid was triggered after the 10 min, aid was imposed by the system for the last 5 min. The design allowed tests of the overall impact of aid on single- and dual-task performance in scenarios SSS and SDD (Hypothesis 2). It also afforded an analysis of whether such aid effects were moderated by type of aid (imposed or adaptive: Hypothesis 3).

2.2. Experimental tasks and context

Tasks were administered via the Mixed Initiative eXperimental testbed (MIX testbed; Reinerman-Jones et al., 2010). Participants were told that they were Soldiers in a human-robot team on an intelligence, surveillance and reconnaissance (ISR) mission. In this scenario, their UGV robot teammate patrolled an area-of-interest and transmitted reconnaissance information back to the Soldier. The feed from the UGV was presented as an aerial map with icons representing enemy activities/entities, and a video feed of the UGV's environment, corresponding to the Change Detection Task (CDT) and Threat Detection Task (TDT) environments respectively (see Fig. 2).

2.2.1. Change detection task

The participant was required to monitor the Change Detection Task (CDT) window (see Fig. 2) to detect and identify the changes to these icons as soon as such changes occurred. Icons could (i) appear, (ii) disappear, or (iii) move locations. The participant then clicked on the appropriate ”appeared”, “disappeared”, “movement” button as soon as s/he detected the change. The event rate, set size, and saliency of change events were set at levels established as imposing medium workload in previous studies (i.e., Abich, 2013).

2.2.2. Threat detection task

The Threat Detection Task (TDT) required the participant to monitor the video feed for threat characters in the environment (see Fig. 3). Threats comprised enemy soldiers and armed civilians/insurgents (see Fig. 4 for examples from each category). The event rate and threat probability were set at a medium-workload level based on previous studies (i.e., Abich, 2013). Participants clicked on the characters they identified as threats. Duplicates occurred when the same character was clicked on more than one occasion.

2.3. Workload monitoring and robot aid

During training, participants, designated as Soldiers, were assigned to a “supervisor” role while the UGV assumed the role of “subordinate” in the human-robot team. Participants were informed that the robot teammate would monitor their level of workload throughout the mission, and would switch to “peer” mode when the level of workload was high. An auditory message was employed as a notification of this mode switch, i.e., “Robot teammate now in peer mode and will be assisting with change detection task”.

The robot aid rendered in “peer” mode was in the form of an auditory “beep” that occurred concurrently with each change event, to increase the saliency of those events. The decision to use

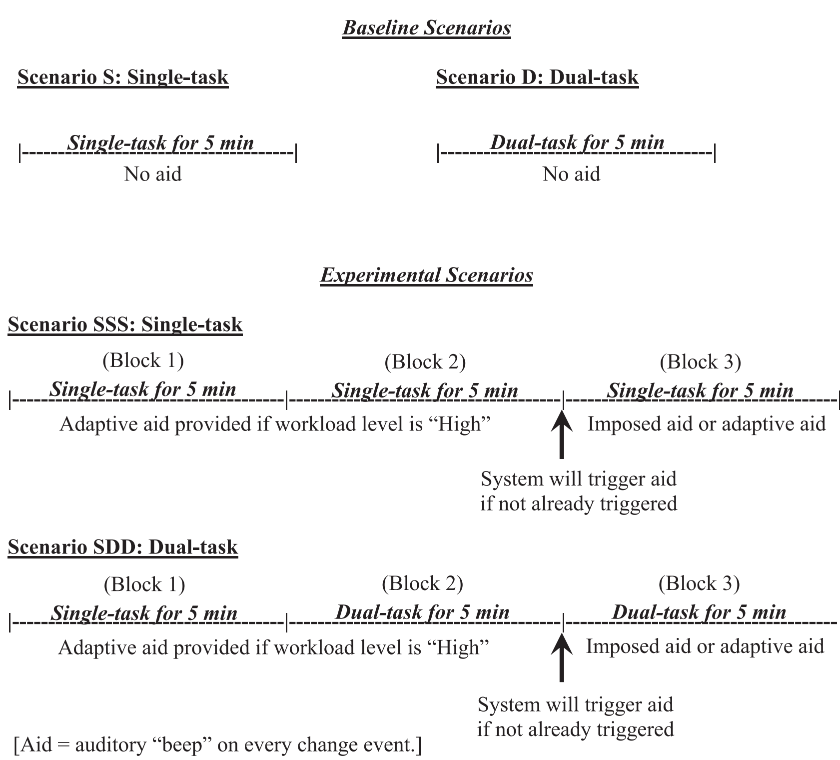

Fig. 1. Study scenarios.

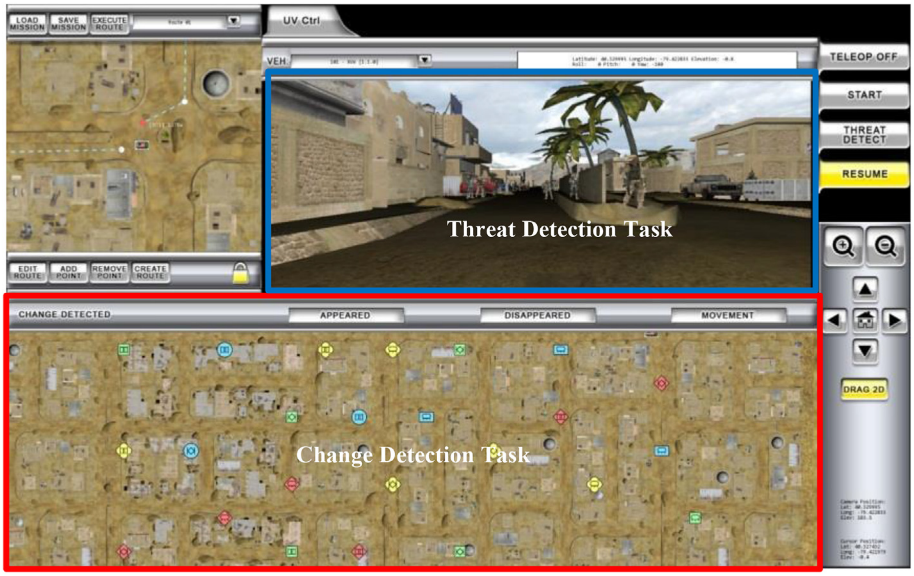

Fig. 2. Screen capture of the MIX testbed.
(outlines overlaid to differentiate each window but were not present in the actual task).

Fig. 3. Video feed showing characters in the “ground level” view.

Fig. 4. Categories of characters in the TDT.
From left to right: (i) friendly Soldiers, (ii) friendly civilians, (iii) enemy soldiers, (iv) armed civilians/insurgents.

an auditory “beep” as the aid was made to avoid adding further visual information to the tasks which were already heavily visual in nature (cf., Multiple Resource Theory; Wickens, 2008). Results of previous studies with the same change detection task showed that this auditory aid improved detection rates but also resulted in poorer change type classification accuracy (Taylor, 2012; Taylor et al., 2013). Hence, the aid could improve some aspects of performance, but could adversely affect other aspects of performance.

2.4. Closed-loop system

Baseline scenarios S (single-task) and D (dual-task) induced the low and high workload states respectively that were needed to identify the individual's own set of physiological workload markers. For each physiological measure, the difference in workload values between these baseline scenarios constituted the "baseline difference score" for that measure. The physiological markers of workload for the individual comprised measures that were sensitive to the workload changes between the low (i.e., single-task) and high (i.e., dual-task) taskload conditions. (see Teo et al., 2016). During each of the experimental scenarios (i.e., Scenarios SSS and SDD), the "test difference score" for each of the markers was computed with samples of the individual's workload state obtained at regular intervals. With each real-time sample, a workload index that quantified degree of similarity in the physiological response pattern, as captured by the set of "baseline difference scores" and "test difference scores", was computed. The index was calculated by using the best-performing algorithm from Teo et al.'s (2016) study, and was based on calculating the proportion of markers that showed a change of ±0.5 SD or greater magnitude response to increased taskload (see Teo et al., 2016; for further details).

The samples comprised data blocks that were 2 min in duration, sampled every 30 s, beginning from the second minute into the scenario (i.e., 2 min rolling window). If the index was equal or greater than the threshold, the workload state for that data sample would be "high". When three consecutive data samples (i.e., "debouncing factor" of 3) were "high", the closed-loop system would classify the participant as being in a stable "high" workload state, and the adaptive aid was evoked immediately after the participant was notified that the robot switched to the "peer" mode. Due to the "debouncing" and sampling rate, the earliest point in which the aid could be triggered in the experimental scenarios was 3 min after the start of the scenario. Once the aid was provided here, it was never revoked, and would continue for the remainder of the scenario. If the aid was not triggered in the first 10 min of the experimental scenarios, then aid would be "imposed" by the system in the last 5 min. Participants were told that aid might be provided, but they were not informed of the criterion used to trigger it.

2.5. Participants

Sixty-two psychology students (34 males, 27 females, 1 unknown) aged from 18 to 52 years ($M = 20.31$) from a large southeastern U.S. university served as participants. They were awarded course credit for their participation. Participants were self-reported to be right-handed, had self-reported normal or corrected-to-normal vision and normal color vision. They were required to refrain from consumption of caffeine 2 h prior, and any alcohol or sedative medication 24 h prior to their participation.

2.6. Procedure

After obtaining participants' informed consent, physiological sensors were attached. A 5 min physiological resting baseline was taken during which participants were told to relax and to keep their eyes open. They were then briefed on the mission, and trained on the tasks which included familiarization on how the robot teammate would switch to "peer" mode. After training, participants undertook three short practice missions with (i) the CDT alone, (ii) the CDT and TDT simultaneously, and (iii) both tasks together with the robot aid. This was followed by baseline scenarios S and D, the order of which was counterbalanced. After the baseline scenarios, the individualized, quantitative workload model based on the participant's own set of physiological workload markers was derived. This procedure took less than 2 min on average. The model was then applied to the two experimental scenarios that followed (i.e., Scenarios SSS and SDD, which were themselves counterbalanced). The NASA-TLX perceived workload questionnaire was administered after each of the four scenarios. The entire experimental session lasted approximately 3 h.

2.7. Measures

2.7.1. Performance measures

Performance on the tasks was evaluated with the measures summarized in Table 1:

2.7.2. Subjective measures

Perceived workload was assessed with the NASA Task Load Index (TLX; Hart and Staveland, 1988). Six sources of workload, i.e., mental demand, physical demand, temporal demand, performance, effort, and frustration, were rated on 0–100 scales. Overall workload was assessed as an unweighted average of these ratings.

2.7.3. Physiological measures

The physiological measures of workload were as specified in Table 2. Each has been validated as a workload metric in multiple studies. Measures such as electrodermal activity that may be primarily sensitive to sympathetic activation were not included. Further details of the sensors used and validation evidence have been provided by Abich (2013) and Matthews et al. (2015).

3. Results

Data were examined for outliers and Winsorized by setting outliers to the 5th and 95th percentile values (Tabachnick and Fidell, 2006). For all analyses, the degrees of freedom were adjusted for violations of sphericity using Box's epsilon (Maxwell and Delaney, 2004). Inflation of Type I error from multiple pairwise comparisons was minimized with the Bonferroni correction. All physiological measures were computed as percentage change from resting baseline. The flowchart in Fig. 5 summarizes the datasets that were analyzed. All participants ($N = 62$) first performed the tasks in baseline scenarios without aid. Then, they performed each of the experimental scenarios, SSS and SDD, in counterbalanced order. For each of these scenarios, participants were divided into those who received adaptive aid and those with imposed aid: Fig. 5 shows the Ns in each category. Overall, six participants received adaptive aid in both experimental scenarios, 13 received adaptive aid in only one experimental scenario, and 43 received no adaptative aid in either experimental scenarios.

Analysis Plan

Results are presented in two sections, as shown in Table 3. First, we report the analyses of taskload effects in the subset of participants whose workload state did not reach the threshold to trigger adaptive aid, but whom had aid imposed in block 3 of both experimental scenarios ($N = 43$). This group received the same schedule of task demands and imposed aid, so that the analysis of task load effect was not confounded by variability in aid onset time. (cf., those participants who received adaptive aid had it triggered at different onset times during blocks 1–2). To strengthen conclusions about the effects of taskload observed, we also checked for order

Table 1
Performance measures and their descriptions.

Measure	Description
Change detection task (CDT) performance	
1. Percent correct	Percentage of change events that were correctly detected and the change type was correctly identified.
2. Percent misclassified	Mean of the following: • Percent of appeared events misclassified as disappeared or moved • Percent of disappeared events misclassified as appeared or moved • Percent of moved events misclassified as appeared or disappeared
3. Average response time	Average response time for correct detections (i.e. change events detected and correctly classified).
4. Number of false alarms	Number of false alarms committed.
Threat detection task (TDT)performance	
5. Total correct	Number of threats correctly detected.
6. Total duplicates	Number of threats detected more than once.

Table 2
Physiological workload measures.

Physiological workload measure	Sensor	Sampling rate (Hz)
1. *Brain activity*: alpha, beta, theta band spectral powers in the frontal, parietal and occipital lobes	Electroencephalogram (EEG): Advanced Brain Monitoring (ABM) B-Alert X10 system	256
2. *Cardiac activity measures*: heart rate, heart rate variability (SD of interbeat intervals)	Electrocardiogram (ECG):ABM B-Alert X10 system	256
3. Cerebral oxygen saturation (rSO_2) levels of the left and right prefrontal cortex	Functional near-infrared spectroscopy (fNIRS): Somanetics Invos Cerebral/Somatic Oximeter, Model 5100C	0.2
4. Cerebral blood flow velocity (CBFV) from middle cerebral arteries in the left and right hemispheres	Transcranial Doppler ultrasonography (TCD): Spencer Technologies ST3 Digital Transcranial Doppler, Model PMD150	1
5. *Ocular activity measures*: number of fixations, mean and SD of fixation durations,"Index of cognitive activity" (ICA) based on pupil diameter (Marshall, 2002)	Eyetracker: Seeing Machines faceLAB 5	60

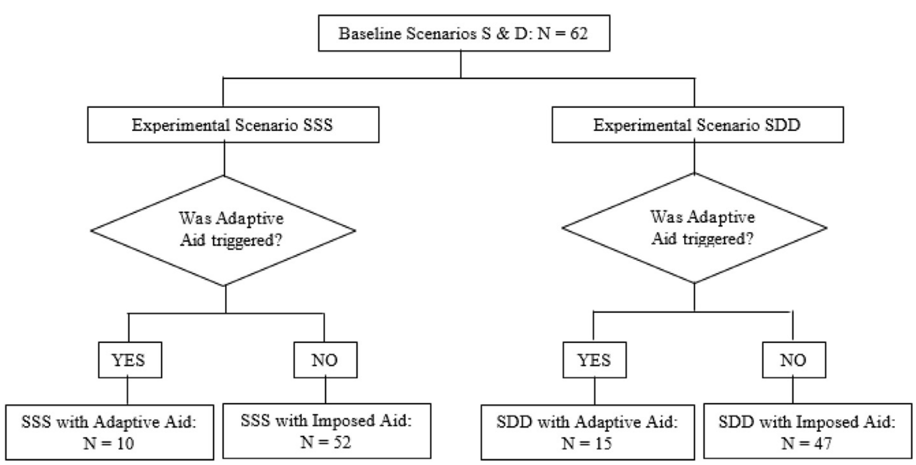

Fig. 5. No. of participants with adaptive aid or imposed aid by scenario.

effects in this group. These analyses included baseline scenario data, as well as experimental scenario data divided into three successive 5 min blocks as shown in Fig. 1.

Second, we analyzed data from each of the two 15 min experimental scenarios (SSS and SDD) to test for effects of aid. Analyses were performed by the 5 min blocks (see Fig. 1). Separate analyses were conducted for each scenario, because different sets of participants received adaptive aid in each one. For the SSS scenario, there were 10 (see Fig. 5) with adaptive aid, and 52 with imposed aid. For the SDD scenario, there were 15 (see Fig. 5) with adaptive aid, and 47 with imposed aid. These analyses differentiated the effect of receiving aid of any kind from the impact of aid type (imposed vs. adaptive).

3.1. Order and taskload effects

3.1.1. Order effects

Table 4 shows performance levels in all baseline and experimental conditions for the group receiving imposed aid throughout.

Comparisons of baseline and experimental blocks to determine taskload and aid effects could be confounded by order effects since the baseline blocks were always administered first. Results of a 2(Taskload: Single-task vs. Dual-task) x 4(Block: Baseline vs. Block 1 vs. Block 2 vs. Block 3) repeated measures ANOVA revealed that performance in the baseline blocks were not significantly different from that in the experimental blocks for all performance measures, so long as the taskload of the blocks in the comparison were of the

Table 3
Analyses plan and participants for each analysis type.

Principal independent variables	Principal dependent variable	Participants
Taskload	• Change Detection task - Percent correct - Percent misclassified - Average response time - Number of false alarms • Physiological workload index • NASA-TLX Global workload	Only participants with imposed aid since all had the same schedule of stimuli throughout all scenarios (N = 43).
Aid and Aid-type	• Change Detection task - Percent correct - Percent misclassified - Average response time - Number of false alarms • Threat Detection task - Total correct - Total duplicates	Participants grouped into Imposed aid vs. Adaptive aid groups. Separate analyses conducted for Single task (N = 10 vs. 52) and Dual task (N = 15 vs. 47) since groupings for SSS and SDD scenarios were different.

Table 4
Means (and SEs) for CDT performance measures by taskload and block (imposed aid group; N = 43).

	Percent Correct	Percent Misclass.	Avg. resp. time	No. false alarms
Single-task blocks (low taskload)				
• Scen. S Block (no aid)	64.61 (1.80)	6.17 (0.46)	1.51 (0.03)	0.88 (0.19)
• Scen. SSS Block 1 (no aid)	65.22 (2.10)	6.37 (0.49)	1.49 (0.03)	0.91 (0.14)
• Scen. SSS Block 2 (no aid)	62.64 (2.04)	5.70 (0.42)	1.49 (0.03)	0.61 (0.16)
• Scen. SSS Block 3 (imposed aid)	**67.79 (1.88)**	**9.11 (0.84)**	**1.43 (0.03)**	**0.49 (0.18)**
• Scen. SDD Block 1[a] (no aid)	63.33 (2.11)	5.47 (0.47)	1.48 (0.03)	0.72 (0.20)
Dual-task blocks (high taskload)				
• Scen. D Block (no aid)	44.06 (1.64)	7.94 (0.57)	1.58 (0.03)	0.47 (0.12)
• Scen. SDD Block 2 (no aid)	44.45 (2.13)	6.54 (0.47)	1.60 (0.03)	0.77 (0.20)
• Scen. SDD Block 3 (imposed aid)	**67.79 (1.73)**	**12.65 (1.03)**	**1.54 (0.02)**	**0.42 (0.12)**

[a] The SDD scenario began as a single-task in Block 1 and taskload was increased to dual-task in Blocks 2 and 3 (see Fig. 1).

same level (i.e., all single-task blocks or all dual-task blocks), $p > 0.05$. For instance, there were no significant differences in performance among the Single Baseline block, and the Experimental Blocks 1 or 2 of the SSS condition, or the Experimental Block 1 of the SDD scenario. Likewise there were no performance differences between the Dual Baseline block and Experimental Block 2 of the SDD scenario, $p > 0.05$ for all. Thus, order effects were not significantly influential in the present experiment.

3.1.2. *Taskload effects*

The impact of taskload was tested by comparing the physiological workload index, perceived workload ratings, and performance in single- and dual-task conditions. To minimize confounding results with the various aid onset times in the adaptive aid group who had aid triggered by their physiological workload state, the analysis was restricted to participants whose aid was triggered by the system in Block 3 (i.e., imposed aid; N = 43). They experienced the same schedule of stimuli throughout the experiment. Taskload effects were tested for performance, physiological workload index, and subjective workload measures.

3.1.2.1. Change detection performance. The series of 2(Taskload: Single-task vs. Dual-task) x 4(Block: Baseline condition vs. Block 1 vs. Block 2 vs. Block 3) repeated measures ANOVAs revealed a significant main effect of Taskload for the measures of CDT performance (see Table 4 for descriptive statistics). There were higher detection rates, $F(1,42) = 228.88$, $p < 0.001$, $\eta_p^2 = 0.85$, lower percent misclassifications, $F(1,42) = 11.31$, $p = 0.002$, $\eta_p^2 = 0.21$, and faster response times, $F(1,42) = 21.80$, $p < 0.001$, $\eta_p^2 = 0.34$, in the single-task blocks relative to the dual-task blocks. The taskload effect was not significant for the number of false alarms, $F(1,42) = 2.03$, $p = 0.161$, $\eta_p^2 = 0.046$. Main effects of block were also found for several measures, but simply reflected taskload differences across those blocks (see Table 4).

3.1.2.2. Physiological workload index. The trending of the physiological workload index over the SSS and SDD scenarios revealed a positive relationship between taskload and workload as the workload index was always higher during dual-task (high workload) blocks (see Fig. 6). The index rose slightly in both scenarios at the start of block 3 at the introduction of the

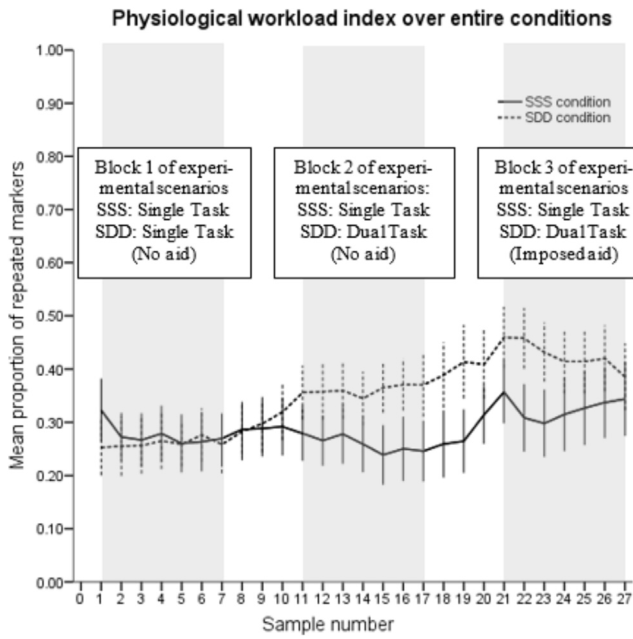

Fig. 6. Physiological workload index of imposed aid group over all samples in scenarios SSS and SDD.
(Shaded areas correspond to the data used in the block analyses. Error bars show 95% C.I.).

imposed aid, and levelled off after that.

3.1.2.3. Perceived workload. The NASA-TLX was administered after each of the four scenarios so analysis by block could not be performed (see Table 5). The one-way repeated measures ANOVA performed for the raw Global workload score revealed a significant effect of scenario, $F(18,25) = 5.38$, $p < 0.001$, $\eta_p^2 = 0.80$. Although the global workload during the shorter dual-task baseline scenario (i.e., Scenario D) was comparable to that during the longer single-task experimental scenario (i.e., Scenario SSS), global workload was generally higher in the dual-task scenarios compared to the single-task scenarios (see Table 5).

3.2. Aid and aid-type effects

For each of the two experimental scenarios, SSS and SDD, participants were divided into those with the adaptive aid and those with the imposed aid. As the groupings were different for scenarios SSS and SDD (see Fig. 5), separate analyses were conducted for single-task and dual-task (see Tables 7 and 8 for descriptive statistics for single-task and dual-task respectively).

Initial analyses with one-way between-participants ANOVA on CDT performance for single- and dual-task scenarios in turn indicated that there were no significant differences between the adaptive aid and imposed aid groups on CDT performance. Hence, the groups with adaptive and imposed aid showed similar levels of CDT proficiency prior to delivery of aid. Single-task: $F(1,60) = 0.218$, $p = 0.642$ for percent correct, $F(1,60) = 1.397$, $p = 0.242$ for percent misclassified, $F(1,60) = 0.193$, $p = 0.662$ for average response time, $F(1,60) = 0.008$, $p = 0.928$ for number of false alarms). Dual-task: $F(1,60) = 2.71$, $p = 0.105$ for percent correct, $F(1,60) = 0.175$, $p = 0.677$ for percent misclassified, $F(1,60) = 3.44$, $p = 0.069$ for average response time, $F(1,60) = 0.689$, $p = 0.410$ for number of false alarms), To evaluate the effects of aid and aid-type, a 2(Aid: Without vs. With Aid) x 2(Aid-type: Adaptive aid vs. Imposed aid) mixed-model ANOVA with the second factor as a between-subjects factor, was conducted for the single-task and dual-task blocks in turn. The baseline condition was designated as the “without aid” block (order effects were minimal), and the “with aid” block comprised the first 5 min after the onset of the aid. For those with the imposed aid, this coincided with Block 3, whereas for those with adaptive aid, this was determined individually. Doing so ensured that the length of time participants were exposed to the aid was held constant across all participants (see Table 6).

3.2.1. Change detection performance

3.2.1.1. Single-task. Results revealed a significant main effect of aid for percent misclassified, $F(1,60) = 7.50$, $p = 0.008$, $\eta_p^2 = 0.11$, and average response time, $F(1,60) = 6.13$, $p = 0.016$, $\eta_p^2 = 0.09$, but not for percent correct, $F(1,60) = 2.85$, $p = 0.097$, $\eta_p^2 = 0.045$, nor for number of false alarms, $F(1,60) = 0.343$, $p = 0.560$, $\eta_p^2 = 0.006$. For both groups of participants, percent misclassified was higher and average response times were faster with aid (see Table 7).

There was a significant aid by aid-type interaction effect for the number of false alarms, $F(1,60) = 4.767$, $p = 0.033$, $\eta_p^2 = 0.074$. Analyses of simple effects by aid showed that the number of false alarms for the groups were similar when there was no aid, but was significantly higher for the group with adaptive aid after the provision of aid. For the imposed aid group, the number of false alarms declined with the provision of aid by an average of 0.40 ($SD = 1.46$), but for the adaptive aid group, the number of false alarms increased instead by an average of 0.70 ($SD = 1.49$) when aid was provided (see Fig. 7).

3.2.1.2. Dual-task. The ANOVA yielded a significant effect of aid for percent correct, $F(1,60) = 58.43$, $p < 0.001$, $\eta_p^2 = 0.493$, percent misclassified, $F(1,60) = 26.57$, $p < 0.001$, $\eta_p^2 = 0.31$, average response time, $F(1,60) = 14.07$, $p < 0.001$, $\eta_p^2 = 0.19$, but, as with previous analyses, not for number of false alarms, $F(1,60) = 0.121$, $p = 0.729$,

Table 5
Means (and SEs) for NASA-TLX Global workload across Scenarios (N = 43).

	Scenario S (Single-task baseline)	Scenario D (Dual-task baseline)	Scenario SSS (Single-task experimental)	Scenario SDD (Dual-task experimental)
Global workload	38.92 (2.61)	53.16 (2.35)	51.22 (3.29)	63.33 (2.20)

Table 6
Description of levels for the factors aid and aid-type.

		Aid-type	
		Adaptive aid	Imposed aid
Aid	Without Aid	Baseline block of adaptive aid group	Baseline block of imposed aid group
	With Aid	First 5 min from start of aid onset (variable by participant)	First 5 min from start of aid, i.e., Block 3 of experimental scenario

Table 7
Means (SE) for CDT performance without and with aid across groups: Single task (N = 10 vs. 52).

	Group	Without aid	With aid
Percent Correct	Adaptive Aid	63.75 (3.61)	67.58 (3.55)
	Imposed Aid	65.59 (1.58)	68.69 (1.56)
	All	64.67 (1.97)	68.14 (1.94)
Percent Misclassified	Adaptive Aid	7.27 (0.98)	9.42 (1.58)
	Imposed Aid	6.01 (0.43)	9.02 (0.69)
	All	**6.64 (0.53)**	**9.22 (0.87)**
Avg. response time	Adaptive Aid	1.43 (0.02)	1.44 (0.05)
	Imposed Aid	1.51 (0.03)	1.43 (0.02)
	All	**1.49 (0.03)**	**1.43 (0.03)**
No. of false alarms	*Adaptive Aid*	*0.90 (0.43)*	*1.60 (0.38)*
	Imposed Aid	*0.94 (0.19)*	*0.54 (0.17)*
	All	0.92 (0.23)	1.07 (0.21)

*Means in bold (main effects) and italics (interaction effects) are sig. different, $p < 0.05$.

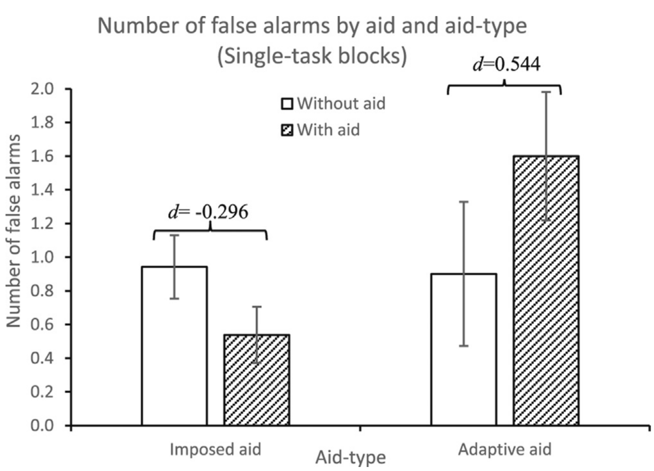

Fig. 7. Number of false alarms by aid and aid-type in single-task blocks. (error bars denote SEs).

Table 8
Means (and SEs) for CDT performance without and with aid across groups: Dual task (N = 15 vs. 47).

	Group	Without aid	With aid
Percent Correct	*Adaptive Aid*	38.81 (2.59)	55.93 (2.93)
	Imposed Aid	43.69 (1.46)	52.15 (1.66)
	All	**41.25 (1.49)**	**54.04 (1.68)**
Percent Misclassified	Adaptive Aid	7.39 (0.90)	11.90 (1.56)
	Imposed Aid	7.82 (0.51)	12.60 (0.88)
	All	**7.60 (0.51)**	**12.25 (0.90)**
Avg. response time	*Adaptive Aid*	1.68 (0.05)	1.52 (0.04)
	Imposed Aid	1.57 (0.03)	1.53 (0.03)
	All	**1.63 (0.03)**	**1.53 (0.03)**
No. of false alarms	Adaptive Aid	0.67 (0.21)	0.80 (0.23)
	Imposed Aid	0.47 (0.12)	0.45 (0.13)
	All	0.57 (0.12)	0.62 (0.13)

*Means in bold (main effects) and italics (interaction effects) are sig. different, $p < 0.05$.

$\eta_p^2 = 0.002$. Detection rates increased and response times decreased with aid, although misclassification rates also increased with aid.

A significant aid by aid-type interaction effect was observed for percent correct, $F(1,60) = 6.708$, $p = 0.01$, $\eta_p^2 = 0.10$, and for average response time, $F(1,60) = 4.48$, $p = 0.04$, $\eta_p^2 = 0.07$. Without the aid, the adaptive aid group had lower percent correct, and slower average response times compared to the imposed aid group. With aid, the detection rate for the adaptive aid group surpassed that of the imposed aid group, and their average response times decreased to be comparable to that of the imposed aid group (see Table 8).

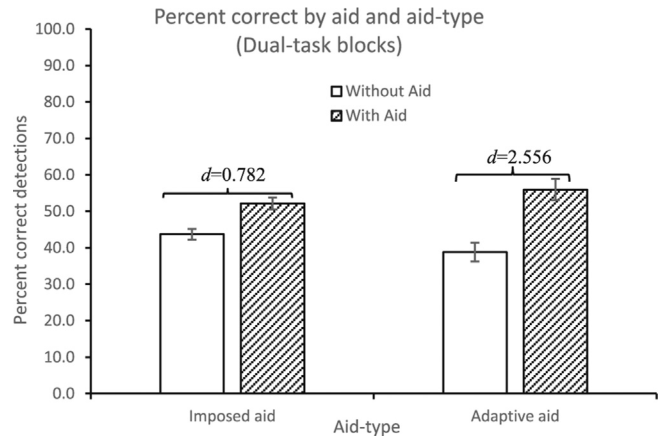

Fig. 8. Percent correct by aid and aid-type in dual-task blocks. (error bars denote SEs).

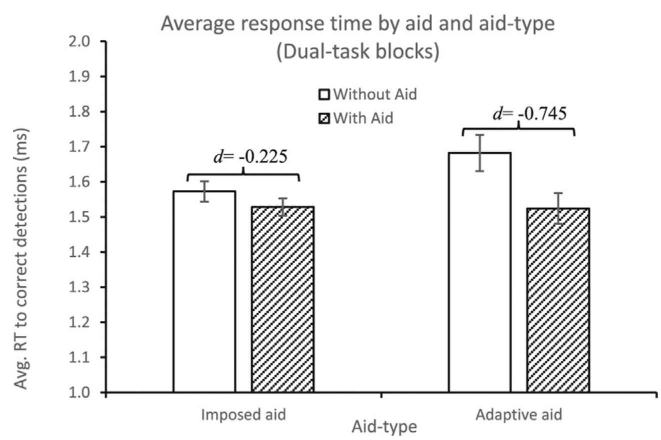

Fig. 9. Average response time by aid and aid-type in dual-task blocks. (error bars denote SEs).

For the imposed aid group, the percent correct increased by 8.46% on average ($SD = 11.72$), but for the adaptive aid group, percent correct improved by more than double, at 17.12% on average ($SD = 9.72$) with the provision of aid. The average response time for the imposed aid group decreased by 0.04 s on average ($SD = 0.18$), while the average response time for the adaptive aid group was reduced by 0.16 s on average ($SD = 0.20$) when aid was provided (see Fig. 8 and Fig. 9).

3.2.2. Threat detection performance

Threat detection performance was only analyzed for dual-task blocks because the TDT was only administered in these dual-task conditions (i.e. Scenarios D and SDD). Comparison of TDT performance between the groups receiving imposed aid versus adaptive aid was accomplished via a one-way between subjects ANOVA on the baseline scenario D. Results revealed no significant differences between the groups, indicating that the groups were equivalent on TDT performance at the outset, $F(1,60) = 0.138$, $p = 0.712$ for total correct, $F(1,60) = 0.599$, $p = 0.442$ for total duplicates.

Designating the baseline scenario D as the "without aid" block, and the first 5 min after aid onset as the "with aid" block, a 2(Aid: Without aid vs. With aid) x 2(Aid-type: Imposed aid vs. Adaptive aid) mixed-model ANOVA with the second element as a between-

Table 9
Means (and SEs) for TDT performance without and with aid across groups: Dual task (N = 15 vs. 47).

	Group	Without aid	With aid
Total Correct	Adaptive Aid	16.13 (0.61)	14.40 (0.70)
	Imposed Aid	15.87 (0.35)	15.43 (0.40)
	All	**16.00 (0.35)**	**14.91 (0.40)**
Total Duplicates	Adaptive Aid	1.73 (0.59)	2.93 (0.74)
	Imposed Aid	1.21 (0.33)	1.36 (0.42)
	All	**1.47 (0.34)**	**2.15 (0.43)**

*Means in bold (main effects) are sig. different, $p < 0.05$.

subjects factor was conducted. Results indicated a significant aid effect for total correct, $F(1,60) = 7.04$, $p = 0.01$, $\eta_p^2 = 0.11$, and total duplicates, $F(1,60) = 5.74$, $p = 0.02$, $\eta_p^2 = 0.09$. Total correct decreased and total duplicates increased when aid was provided. This finding suggests that the aid for the CDT may have diverted attentional resources away from the TDT. No other effects were significant (Table 9).

The effects of the aid could not be evaluated for perceived workload since the subjective workload measure was only administered after each scenario, while the aid was provided in the midst of the experimental scenarios.

4. Discussion

The present study has demonstrated that individualized workload modeling enhances the utility of an alerting aid, but it also identified challenges for practical deployment of such a system. Analyses of dual-tasking effects revealed that increasing taskload elevated workload and impaired CDT performance, providing support for the first hypothesis. The effects of providing aid (whether adaptive or imposed) depended on taskload. The aid used in the present study improved overall CDT accuracy (percent correct) only in dual-task blocks. This implies that the aid may only be helpful when attention is taxed by dual-tasking, consistent with the second hypothesis. However, both dual- and single-task blocks showed a pattern of faster response times but also more misclassified change events, suggesting that aid may also encourage a tendency towards more impulsive responding. The aid did not improve TDT performance, but instead resulted in fewer correct threat detections and an increase in duplicate threat identifications, suggesting that the aid may have diverted attentional resources away from the TDT to the CDT. Hence, the hypothesis that aid would improve performance is only partially supported here.

The effects of the aid used in the present study depended on whether the aid was adaptive or imposed. The adaptive aid proved more beneficial when taskload was high. During low taskload (i.e., single-task), adaptive aiding increased false alarms, while imposed aid led to reduced false alarms. However, under higher taskload (i.e., dual-task), the aid improved detection rates and speed of response by a larger extent in the adaptive aid group. When provided aid, mean detection rate for this group surpassed that of the imposed aid group. Further, their slower response times then became equated to that of the imposed aid group. Thus, the hypothesis that adaptive aid was superior to imposed aid was largely supported. Despite this, it must be kept in mind that these findings may be specific to tasks and aid formats of similar nature. Thus, further investigations are needed to determine if those results generalize to other task contexts.

Overall, the effects of the aid used in the present study were to a degree equivocal. Closed-loop adaptive automation may help to optimize potential benefits from an aid to performance, but the study also identifies barriers to be overcome in transitioning to practical application. The introduction of the aid increased the workload index of the imposed aid group (Fig. 6), and is likely to have affected the adaptive aid group in a similar way, although the workload level of the adaptive aid group would already have been high by the time the aid was rendered. An auditory aid was chosen to minimize resource competition with the detection tasks, but studies of multimodal displays show that auditory interruptions can also draw attention away from the primary task (Lu et al., 2013). The impact of the aid on workload may reflect the costs of resisting such pre-emption of attention.

The consequences of the aid for performance were complex, and costs and benefits depended on the task configuration. The most salient cost of providing the aid was increased misclassifications, perhaps reflecting an impulsive strategy shift associated with anxiety induced by high demand. Indeed, aid appeared potentially problematic in the single-task condition, in which it elicited impulsive responding which resulted in more false alarms without any overall improvement in accuracy in those who were already experiencing high workload (i.e., adaptive aid group). In contrast, the aid resulted in fewer false alarms in the imposed aid group, who presumably had sufficient resources to cope with the increase in workload from the interruption produced by the aid. In the dual-task condition, the aid produced a clearer benefit in substantially enhancing overall accuracy without inflating false alarm frequency, although there was some impairment in TDT. This second task probably diverted some resources to TDT, which suppressed false alarms in the CDT, but not enough to improve TDT performance. Nevertheless, the accuracy benefit of the aid was considerably larger when the aid was adaptive. Thus, on a cost-benefit basis, providing an automated aid may only be worthwhile in dual-task and adaptive conditions. In terms of attentional resource theory (Matthews et al., 2000), the aid reduces the need to allocate attention to change detection, but under low taskload the person already has sufficient resources for the task in any case.

Thus, the findings imply that the interplay between taskload, workload and performance is not straightforward, especially in a multi-tasking environment. Individual operators have the prerogative to allocate resources dynamically to different processes that underpin various tasks. Closed-loop regulation of workload contributes to effective human-robot teaming by maintaining workload within the moderate range in which people are generally effective in compensating for variation in cognitive demand (Hancock and Warm, 1989). However, operators still require additional skills for optimal prioritization of different task components and maintaining team situation awareness (Schuster and Jentsch, 2011). Thus, workload regulation may be necessary but not sufficient for optimizing teaming.

Findings also illustrate the challenges of applying closed-loop technology. In the single-task, SSS scenario, during which the aid should not be triggered, there was nevertheless considerable variance in values of the physiological workload index (see Fig. 5). Some of this variation may reflect random error, but it may also indicate workload associated with individual differences in task strategy (Hockey, 1997). For example, some individuals may be applying more effort than is necessary to maintain performance; others may incur workload by effortful suppression of distracting thoughts or other voluntary processing strategies (e.g, Zaccaro et al., 2014). Adaptive aid may then be triggered inappropriately. In the dual-task, SDD scenario, benefits of adaptive aiding were more apparent. Indeed, adaptive triggering of the aid may be more effective in enhancing performance accuracy because it delivers relief from attentional overload when most needed. However, although the experiment was designed to ensure a close match between baseline (control) and test (experimental) conditions, adaptive aid was triggered only in a minority of the present participants during dual-tasking.

Another challenge is that adaptive aiding may influence factors other than workload, such as stress, motivation, and situation awareness (see Smith and Hancock, 1995). In consequence, physiological workload measures may not map directly onto any specific mental processes as opposed to global brain states in that they are sensitive only to large changes in workload. Several peripheral measures, such as ocular and cardiac responses, are also more prone to be confounded by task demands and so may be imperfect indicators of cognitive workload. There can also be unforeseen issues and tradeoffs with adaptive closed-loop systems. For instance, Chen and Barnes (2014) suggest that in human-robot teaming, adaptive automation may help to maintain situation awareness but threaten the human's sense of delegation authority. Such tradeoffs may impact performance over and above mitigation of workload.

A shortcoming of the present aid was that it seemed to elevate workload. Optimizing benefits of the aid, especially under high taskload conditions, may require steps to minimize attentional preemption by the aid stimulus. There may be design solutions: e.g., use of tactile rather than auditory stimuli might reduce such effects (Lu et al., 2013). Alternatively, training and practice may allow operators to develop strategies for effective processing of the aid.

Further work to improve the algorithm used to integrate workload data from multiple sensors may also be necessary, using additional methods for modeling individual differences in workload response more precisely. These include data mining techniques such as cluster analyses, neural networks, machine learning algorithms etc. For instance, using cluster analysis, Johannes and Gaillard (2014) identified five autonomic response patterns (ARP) based on normalized eigendifference scores. These ARPs, which are influenced differently by coping styles, task demands, and individual states, were used in a scaling method to calculate an index of psychophysiological arousal that accounted for individual differences in autonomic responsivity (Johannes and Gaillard, 2014).

The present study which involved individualized workload modeling might also be applied to training contexts (Zaccaro et al., 2014). Interventions for excessive workload in this context may be adapting information presentation format, or training curricula to the trainee's learning pace (e.g., adaptive tutors). There are also recent developments of wearable technology that claim to assess state changes in real time. These include the *Spire* which is purported to monitor tension and stress from respiratory measures (Pilkington, 2014), and the *Exmocare* which supposedly monitors emotional state from cardiac activity, GSR, skin temperature and movement (Ostrovsky, 2008). All these yield rich data with which to develop models of how psychophysiological states vary within and across individuals.

5. Conclusion

Soldiers, as well as other human operators interacting with technology and working closely with robots will only become more of our future reality. The present work demonstrated the feasibility of using a closed-loop system based on physiological workload measures to drive adaptive robot aid. The applications of such a real-time, adaptive system are diverse. Our findings clearly indicate a need for more research as it is not advisable to adopt a “one-size-fits-all” approach with respect to workload measures. Future research on how these robots may work more collaboratively and communicate more intuitively with human users, must be driven by a deeper understanding of how individuals interact with changing taskloads, as well as the nature of the workload they experience while doing so.

Acknowledgement

This work was supported by the Army Research Laboratory under Grant W91CRB-08-D-0015. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied of the Army Research Laboratory of or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

Abich IV, J., 2013. Investigating the Universality and Comprehensive Ability of Measures to Assess the State of Workload. Unpublished Doctoral Dissertation,. University of Central Florida.

Abich IV, J., Reinerman-Jones, L., Matthews, G., 2017. Impact of three task demand factors on simulated unmanned system intelligence, surveillance, and reconnaissance operations. Ergonomics 60 (6), 791–801.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., Babiloni, F., 2014. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neurosci. Biobehav. Rev. 44, 58–75.

Byrne, E.A., Parasuraman, R., 1996. Psychophysiology and adaptive automation. Biol. Psychol. 42, 249–268.

Cain, B., 2007. A Review of the Mental Workload Literature. Defense Research and Development Canada, Toronto, Canada.

Carmody, M.A., Gluckman, J.P., 1993. Task-specific effects of automation and automation failure on performance, workload and situational awareness. In: Jensen, R.S., Neumeister, D. (Eds.), Proceedings of the Seventh International Symposium on Aviation Psy-chology. Ohio State University, Department of Aviation, Columbus, OH, pp. 167–171.

Chen, J.Y., Barnes, M.J., 2014. Human–agent teaming for multirobot control: a review of human factors issues. IEEE Trans. Human-Machine Syst. 44, 13–29.

da Vinci surgery, 2013. Changing the experience of surgery. Retrieved from. http://www.davincisurgery.com/.

DRC, 2013. DARPA robotics challenge 2013. Retrieved from. http://www.theroboticschallenge.org/about.

Endsley, M.R., Kiris, E.O., 1995. The out-of-the-loop performance problem and level of control in automation. Hum. Factors 37, 381–394.

Guizzo, E., 2010. World robot population reaches 8.6 million. IEEE Spectrum. Retrieved from. http://spectrum.ieee.org/automaton/robotics/industrial-robots/041410-world-robot-population.

Guizzo, E., Ackerman, E., 2012. How rethink robotics built its new baxter robot worker. IEEE spectrum. Retrieved from. http://spectrum.ieee.org/robotics/industrial-robots/rethink-robotics-baxter-robot-factory-worker.

Hancock, P.A., Chignell, M.H., 1987. Adaptive control in human-machine systems. In: Hancock, P.A. (Ed.), Human Factors Psychology, pp. 305–345. Amsterdam: North-Holland.

Hancock, P.A., Chignell, M.H., 1988. Mental workload dynamics in adaptive interface design. IEEE Trans. Syst. Man Cybern. 18 (4), 647–658.

Hancock, P.A., Meshkati, N., Robertson, M.M., 1985. Physiological reflections of mental workload. Aviat. Space, Environ. Med. 56, 1110–1114.

Hancock, P.A., Warm, J.S., 1989. A dynamic model of stress and sustained attention. Hum. Factors 31 (5), 519–537.

Hancock, P.A., Scallen, S.F., 1996. The future of function allocation. Ergon. Des. 4 (4), 24–29.

Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (Eds.), Human Mental Workload. Elsevier Science Publishers, North-Holland, pp. 139–184.

Hockey, G.R.J., 1997. Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework. Biol. Psychol. 45, 73–93.

Horrey, W.J., Lesch, M.F., Garabet, A., 2009. Dissociation between driving performance and drivers' subjective estimates of performance and workload in dual-task conditions. J. Saf. Res. 40, 7–12.

Johannes, B., Gaillard, A.W.K., 2014. A methodology to compensate for individual differences in psychophysiological assessment. Biopsychology 96, 77–85.

Kumar, V., Bekey, G., Zheng, Y., 2006. Industrial, personal and service robots. In: Bekey, G. (Ed.), Assessment of International Research and Development in Robotics. World Technology Evaluation Center, Lancaster, pp. 55–62.

Lu, S.A., Wickens, C.D., Prinet, J.C., Hutchins, S.D., Sarter, N., Sebok, A., 2013. Supporting interruption management and multimodal interface design: three meta-analyses of task performance as a function of interrupting task modality.

Hum. Factors 55 (4), 697–724.
Marshall, S.P., 2002. Index of Cognitive Activity: Measuring Cognitive Workload. Presented at the 7th IEEE Human Factors Meeting, Scottsdale, Arizona.
Matthews, G., 2016. Multidimensional profiling of task stress states for human factors: a brief review. Hum. Factors 58, 801–813.
Matthews, G., Davies, D.R., Westerman, S.J., Stammers, R.B., 2000. Human Performance: Cognition, Stress and Individual Differences. Psychology Press, London.
Matthews, G., Reinerman-Jones, L., Barber, D., Abich, J., 2015. The psychometrics of mental workload: multiple measures are sensitive but divergent. Hum. Factors 57, 125–143.
Maxwell, S.E., Delaney, H.D., 2004. Designing Experiments and Analyzing Data: a Model Comparison Perspective, second ed. Lawrence Erlbaum Associates, Mahwah, N.J.
Meshkati, N., Hancock, P.A., Rahimi, M., Dawes, S.M., 1995. Techniques of mental workload assessment. In: Wilson, J., Corlett, E.N. (Eds.), Evaluation of Human Work: a Practical Ergonomics Methodology, second ed. Taylor and Francis, London, pp. 605–627.
Meshkati, N., Loewenthal, A., 1988. An eclectic and critical review of four primary mental workload assessment methods: a guide for developing a comprehensive model. In: Hancock, P.A., Meshkati, N. (Eds.), Human Mental Workload, pp. 251–267. North-Holland: Amsterdam.
Moray, N., 1984. Mental workload. In: Proceedings of the International Conference on Occupational Ergonomics (Toronto, Canada).
Ostrovsky, G., 2008. Exmocare emotion detection technology: capitalizing on How R U. Retrieved from. http://www.digitaltrends.com/fitness-apparel-reviews/spire-activity-tracker-review/.
Parasuraman, R., Wickens, C.D., 2008. Humans: still vital after all these years of automation. Hum. Factors 50, 511–520.
Pilkington, E., 2014. Spire activity tracker review: every breath you take, every move you make, Spire will be watching you. Retrieved from. http://www.digitaltrends.com/fitness-apparel-reviews/spire-activity-tracker-review/.
Reinerman-Jones, L., Barber, D., Lackey, S., Nicholson, D., 2010. Developing methods for utilizing physiological measures. In: Tadeusz, M., Waldemar, K., Rice, V. (Eds.), Advances in Understanding Human Performance: Neuroergonomics, Human Factors Design, and Special Populations, pp. 102–110. Miami, Florida.
Roscoe, A.H., 1993. Heart rate as a psychophysiological measure for in-flight workload assessment. Ergonomics 36, 1055–1062.
Saxby, D.J., Matthews, G., Warm, J.S., Hitchcock, E.M., Neubauer, C., 2013. Active and passive fatigue in simulated driving: discriminating styles of workload regulation and their safety impacts. J. Exp. Psychol. Appl. 19, 287–300.
Schuster, D., Jentsch, F., 2011. Measurement of Situation Awareness in Human-robot Teams. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Human Factors and Ergonomics Society, Santa Monica, CA.
Smith, K., Hancock, P.A., 1995. Situation awareness is adaptive, externally-directed consciousness. Hum. Factors 37 (1), 137–148.
Szalma, J.L., Teo, G.W., 2012. Spatial and temporal task characteristics as stress: a test of the dynamic adaptability theory of stress, workload, and performance. Acta Psychol. 139, 471–485.
Tabachnick, B.G., Fidell, L.S., 2006. Using Multivariate Statistics, fifth ed. Harper Collins, New York, NY.
Takayama, L., Ju, W., Nass, C., 2008. Beyond dirty, dangerous and dull: what everyday people think robots should do. In: Proceedings of the Human-robot Interaction Conference: HRI 2008, pp. 25–32. Amsterdam, Netherlands.
Taylor, G.S., 2012. Comparing Types of Adaptive Automation within a Multi-tasking Environment. Unpublished Doctoral Dissertation. University of Central Florida.
Taylor, G.S., Reinerman-Jones, L.E., Szalma, J.L., Mouloua, M., Hancock, P.A., 2013. What to automate: addressing the multidimensionality of cognitive resources through system design. J. Cognit. Eng. Decis. Mak. 7, 311–329.
Teo, G., Reinerman-Jones, L., Matthews, G., 2016. Augmenting robot behaviors using physiological measures of workload state. In: Schmorrow, D.D., Fidopiastis, C.M. (Eds.), Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience. Springer International Publishing, Switzerland, pp. 404–415.
Warm, J.S., Tripp, L.D., Matthews, G., Helton, W.S., 2012. Cerebral hemodynamic indices of operator fatigue in vigilance. In: Matthews, G., Desmond, P.A., Neubauer, C., Hancock, P.A. (Eds.), Handbook of Operator Fatigue. Ashgate Press, Aldershot, UK, pp. 197–207.
Wickens, C.D., 2008. Multiple resources and mental workload. Hum. Factors 50, 449–455.
Wilson, J., 2016. Robotics: technologies and global markets. Market research reports. Retrieved from. https://www.bccresearch.com/market-research/engineering/robotics-technologies-markets-report-eng001e.html.
Young, M.S., Brookhuis, K.A., Wickens, C.D., Hancock, P.A., 2015. State of science: mental workload in ergonomics. Ergonomics 58, 1–17.
Young, M.S., Stanton, N.A., 2007. What's skill got to do with it? Vehicle automation and driver mental workload. Ergonomics 50, 1324–1339.
Zaccaro, S.J., Weis, E.J., Chen, T.R., Matthews, M.D., 2014. Situational load and personal attributes: implications for cognitive readiness, adaptive readiness, and training. In: Teaching and Measuring Cognitive Readiness. Springer, US, pp. 93–115.