Human Factors – rok 2021, roč. 63

Číslo 5 (August)

SPECIAL SECTION: MEASURING COGNITIVE WORKLOAD IN HUMAN FACTORS

Per Ø. Braarud, Terje Bodal, John E. Hulsund, Michael N. Louka, Christer Nihlwing, Espen Nystad, Håkan Svengren, Emil Wingstedt. *An Investigation of Speech Features, Plant System Alarms, and Operator–System Interaction for the Classification of Operator Cognitive Workload During Dynamic Work*. pp. 736–756.

Objective: To investigate speech features, human–machine alarms, and operator–system interaction for the estimation of cognitive workload in full-scale realistic simulated scenarios. **Background:** Theories and models of cognitive workload are critical for the design and evaluation of human–machine systems. Unfortunately, there are very few nonintrusive cognitive workload measures available for realistic dynamic human–machine interaction. **Method:** The study was conducted in a full-scope control room research simulator of an advanced nuclear reactor. Six crews, each consisting of three operators, participated in 12 scenarios. The operators rated their workload every second minute. Machine learning algorithms were trained to estimate operators' workload based on crew communication, operator–system interaction, and system alarms. **Results:** Random Forest (RF) utilizing speech and system features achieved an accuracy of 67% on test data. Utilizing speech features only, the accuracy achieved was 63%. The most important speech features were pitch, amplitude, and articulation rate. A 61% accuracy was achieved when alarms and operator–system interaction features were used. The most important features were the number of alarms and amount of operator–system interaction. Accuracy for algorithms trained for each operator ranged from 39% to 98%, with an average of 72%. For a majority of analyses performed, RF and extreme gradient boosting (XGB) outperformed other algorithms. **Conclusion:** The results demonstrate that the features investigated and machine learning models developed provide a potential for the dynamic nonintrusive measurement of cognitive workload. **Application:** The approach presented can be developed for nonintrusive workload measurement in real-world human–machine applications, simulator-based training, and research.

- **Keywords:** cognitive workload, dynamic measurement, nonintrusive, machine learning, simulator research

Lauren R. Kennedy-Metz, Roger D. Dias, Rithy Srey, Geoffrey C. Rance, Heather M. Conboy, Miguel E. Haime, Jacquelyn A. Quin, Steven J. Yule, Marco A. Zenati. *Analysis of Dynamic Changes in Cognitive Workload During Cardiac Surgery Perfusionists' Interactions With the Cardiopulmonary Bypass Pump*. pp. 757–771.

Objective: This novel preliminary study sought to capture dynamic changes in heart rate variability (HRV) as a proxy for cognitive workload among perfusionists while operating the cardiopulmonary bypass (CPB) pump during real-life cardiac surgery. **Background:** Estimations of operators' cognitive workload states in naturalistic settings have been derived using noninvasive psychophysiological measures. Effective CPB pump operation by perfusionists is critical in maintaining the patient's homeostasis during open-heart surgery. Investigation into dynamic cognitive workload fluctuations, and their relationship with performance, is lacking in the literature. **Method:** HRV and self-reported cognitive workload were collected from three Board-certified cardiac perfusionists (N = 23 cases). Five HRV components were analyzed in consecutive nonoverlapping 1-min windows from skin incision through sternal closure. Cases were annotated according to predetermined phases: prebypass, three phases during bypass, and postbypass. Values from all 1min time windows within each phase were averaged. **Results:** Cognitive workload was at its highest during the time between initiating bypass and clamping the aorta (preclamp phase during bypass), and decreased over the course of the bypass period. **Conclusion:** We identified dynamic, temporal fluctuations in HRV among perfusionists during cardiac surgery corresponding to subjective reports of cognitive workload. Not only does cognitive workload differ for perfusionists during bypass compared with pre- and postbypass phases, but differences in HRV were also detected within the three bypass phases. **Application:** These preliminary findings suggest the preclamp phase of CPB pump interaction corresponds to higher cognitive workload, which may point to an area warranting further exploration using passive measurement.

- **Keywords:** physiological measurement, patient safety, physiological psychology, wearable devices, surgical care and procedural technologies

Shiyan Yang, Jonny Kuo, Michael G. Lenné, Michael Fitzharris, Timothy Horberry, Kyle Blay, Darren Wood, Christine Mulvihill, Carine Truche. *The Impacts of Temporal Variation and Individual Differences in Driver Cognitive Workload on ECG-Based Detection*. pp. 772–787.

Objective: This paper aimed to investigate the robustness of driver cognitive workload detection based on electrocardiogram (ECG) when considering temporal variation and individual differences in cognitive workload. **Background:** Cognitive workload is a critical component to be monitored for error prevention in human–machine systems. It may fluctuate instantaneously over time even in the same tasks and differ across individuals. **Method:** A driving simulation study was conducted to classify driver cognitive workload underlying four experimental conditions (baseline, N-back, texting, and N-back + texting distraction) in two repeated 1-hr blocks. Heart rate (HR) and heart rate variability (HRV) were compared among the experimental conditions and between the blocks. Random forests were built on HR and HRV to classify cognitive workload in different blocks and for different individuals. **Results:** HR and HRV were significantly different between repeated blocks in the study, demonstrating the time-induced variation in cognitive workload. The performance of cognitive workload classification across blocks and across individuals was significantly improved after normalizing HR and HRV in each block by the corresponding baseline. **Conclusion:** The temporal variation and individual differences in cognitive workload affects ECG-based cognitive workload detection. But normalization approaches

relying on the choice of appropriate baselines help compensate for the effects of temporal variation and individual differences. **Application:** The findings provide insight into the value and limitations of ECG-based driver cognitive workload monitoring during prolonged driving for individual drivers.

- **Keywords:** driver cognitive workload, temporal variation, individual differences, heart rate variability, multiclass classification, baseline normalization

Reilly J. Innes, Zachary L. Howard, Alexander Thorpe, Ami Eidels, Scott D. Brown. *The Effects of Increased Visual Information on Cognitive Workload in a Helicopter Simulator*. pp. 788–803.

Objective: To test the effects of enhanced display information ("symbology") on cognitive workload in a simulated helicopter environment, using the detection response task (DRT). **Background:** Workload in highly demanding environments can be influenced by the amount of information given to the operator and consequently it is important to limit potential overload. **Methods:** Participants (highly trained military pilots) completed simulated helicopter flights, which varied in visual conditions and the amount of information given. During these flights, participants also completed a DRT as a measure of cognitive workload. **Results:** With more visual information available, pilots' landing accuracy was improved across environmental conditions. The DRT is sensitive to changes in cognitive workload, with workload differences shown between environmental conditions. Increasing symbology appeared to have a minor effect on workload, with an interaction effect of symbology and environmental condition showing that symbology appeared to moderate workload. **Conclusion:** The DRT is a useful workload measure in simulated helicopter settings. The level of symbology-moderated pilot workload. The increased level of symbology appeared to assist pilots' flight behavior and landing ability. Results indicate that increased symbology has benefits in more difficult scenarios. **Applications:** The DRT is an easily implemented and effective measure of cognitive workload in a variety of settings. In the current experiment, the DRT captures the increased workload induced by varying the environmental conditions, and provides evidence for the use of increased symbology to assist pilots.

- **Keywords:** cognitive workload, detection response task, helicopter simulation, heads-up display

Francesco N. Biondi, Balakumar Balasingam, Prathamesh Ayare. *On the Cost of Detection Response Task Performance on Cognitive Load*. pp. 804–812.

Objective: This study investigates the cost of detection response task performance on cognitive load. **Background:** Measuring system operator's cognitive load is a foremost challenge in human factors and ergonomics. The detection response task is a standardized measure of cognitive load. It is hypothesized that, given its simple reaction time structure, it has no cost on cognitive load. We set out to test this hypothesis by utilizing pupil diameter as an alternative metric of cognitive load. **Method:** Twenty-eight volunteers completed one of four experimental tasks with increasing levels of cognitive demand (control, 0-back, 1-back, and 2-back) with or without concurrent DRT performance. Pupil diameter was selected as nonintrusive metric of cognitive load. Self-reported workload was also recorded. **Results:** A significant main effect of DRT presence was found for pupil diameter and self-reported workload. Larger pupil diameter was found when the n-back task was performed concurrently with the DRT, compared to no-DRT conditions. Consistent results were found for mental workload ratings and n-back performance. **Conclusion:** Results indicate that DRT performance produced an added cost on cognitive load. The magnitude of the change in pupil diameter was comparable to that observed when transitioning from a condition of low task load to one where the 2-back was performed. The significant increase in cognitive load accompanying DRT

performance was also reflected in higher self-reported workload. **Application:** DRT is a valuable tool to measure operator's cognitive load. However, these results advise caution when discounting it as cost-free metric with no added burden on operator's cognitive resources.

- **Keywords:** cognitive workload, detection response task, pupil diameter, cognitive cost

Francesco N. Biondi, Angela Cacanindin, Caitlyn Douglas, Joel Cort. *Overloaded and at Work: Investigating the Effect of Cognitive Workload on Assembly Task Performance*. pp. 813–820.

Objective: This study investigates the effect of cognitive overload on assembly task performance and muscle activity. **Background:** Understanding an operator's cognitive workload is an important component in assessing human–machine interaction. However, little evidence is available on the effect that cognitive overload has on task performance and muscle activity when completing manufacturing tasks. **Method:** Twenty-two volunteers completed an assembly task while performing a secondary cognitive task with increasing levels of demand (n-back). Performance in the assembly task (completion times, accuracy), muscle activity recorded as integrated electromyography (EMG), and self-reported workload were measured. **Results:** Results show that the increasing cognitive demand imposed by the n-back task resulted in impaired assembly task performance, overall greater muscle activity, and higher self-reported workload. Relative to the control condition, performing the 2-back task resulted in longer assembly task completion times (+10 s on average) and greater integrated EMG for flexor carpi ulnaris, triceps brachii, biceps brachii, anterior deltoid, and pectoralis major. **Conclusion:** This study demonstrates that working under high cognitive load not only results in greater muscle activity, but also affects assembly task completion times, which may have a direct effect on manufacturing cycle times. **Application:** Results are applicable to the assessment of the effects of high cognitive workload in manufacturing.

- **Keywords:** cognitive workload, assembly task, multitasking, cognitive ergonomics, muscle activity

Tobias Grundgeiger, Jörn Hurtienne, Oliver Happel. *Why and How to Approach User Experience in Safety-Critical Domains: The Example of Health Care*. pp. 821–832.

Objective: To highlight the importance of the personal experience of users who interact with technology in safety-critical domains and summarize three interaction concepts and the associated theories that provide the means for addressing user experience. **Background:** In health care, the dominant concepts of interaction are based on theories arising from classic cognitive psychology. These concepts focus mainly on safety and efficiency, with too little consideration being given to user experience. **Method:** Users in complex socio-technical and safety-critical domains such as health care interact with many technological devices. Enhancing the user experience could improve the design of technology, enhance the well-being of staff, and contribute to modern safety management. We summarize concepts of "interaction" based on modern theories of human–computer interaction, which include the personal experience of users as an important construct. **Results and Conclusion:** Activity theory, embodiment, and interaction as experience provide a theoretical foundation for considering user experience in safety-critical domains. Using an example from anesthesiology, we demonstrate how each theory provides a unique but complementary view on experience. Finally, the methodological possibilities for considering personal experience in design and evaluations vary among the theories. **Application:** Considering user experience in health care and potentially other safety-critical domains can provide an additional means of optimizing interaction with technology, contributing to the well-being of staff, and improving safety.

- **Keywords:** activity theory, concepts of interaction, safety-critical domains, embodiment, eudaimonia, interaction as experience, user experience

Elizabeth L. Fox, Joseph W. Houpt, Pamela S. Tsang. *Derivation and Demonstration of a New Metric for Multitasking Performance*. pp. 833–853.

Objective: We proposed and demonstrate a theory-driven, quantitative, individual-level estimate of the degree to which cognitive processes are degraded or enhanced when multiple tasks are simultaneously completed. **Background:** To evaluate multitasking, we used a performance-based cognitive model to predict efficient performance. The model controls for single-task performance at the individual level and does not depend on parametric assumptions, such as normality, which do not apply to many performance evaluations. **Methods:** Twenty participants attempted to maintain their isolated task performance in combination for three dual-task and one triple-task scenarios. We utilized a computational model of multiple resource theory to form hypotheses for how performance in each environment would compare, relative to the other multitask contexts. We assessed if and to what extent multitask performance diverged from the model of efficient multitasking in each combination of tasks across multiple sessions. **Results:** Across the two sessions, we found variable individual task performances but consistent patterns of multitask efficiency such that deficits were evident in all task combinations. All participants exhibited decrements in performing the triple-task condition. **Conclusions:** We demonstrate a modeling framework that characterizes multitasking efficiency with a single score. Because it controls for single-task differences and makes no parametric assumptions, the measure enables researchers and system designers to directly compare efficiency across various individuals and complex situations. **Application:** Multitask efficiency scores offer practical implications for the design of adaptive automation and training regimes. Furthermore, a system may be tailored for individuals or suggest task combinations that support productivity and minimize performance costs.

- **Keywords:** multitasking, efficiency, models of performance, multiple resource theory, multiple-attribute task battery

AUTOMATION, EXPERT SYSTEMS

Robert S. Gutzwiller, John Reeder. *Dancing With Algorithms: Interaction Creates Greater Preference and Trust in Machine-Learned Behavior*. pp. 854–867.

Objective: We examined a method of machine learning (ML) to evaluate its potential to develop more trustworthy control of unmanned vehicle area search behaviors. **Background:** ML typically lacks interaction with the user. Novel interactive machine learning (IML) techniques incorporate user feedback, enabling observation of emerging ML behaviors, and human collaboration during ML of a task. This may enable trust and recognition of these algorithms. **Method:** Participants judged and selected behaviors in a low and a high interaction condition (IML) over the course of behavior evolution using ML. User trust in the outputs, as well as preference, and ability to discriminate and recognize the behaviors were measured. **Results:** Compared to noninteractive techniques, IML behaviors were more trusted and preferred, as well as recognizable, separate from non-IML behaviors, and approached similar performance as pure ML models. **Conclusion:** IML shows promise for creating behaviors by involving the user; this is the first extension of this technique for vehicle behavior model development targeting user satisfaction and is unique in its multifaceted evaluation of how users perceived, trusted, and implemented these learned controllers. **Application:** There are

many contexts where the brittleness of ML cannot be trusted, but the advantage of ML over traditional programmed behaviors may be large, as in some military operations where they could be scaled. IML in this early form appears to generate satisfactory behaviors without sacrificing performance, use, or trust in the behavior, but more work is necessary.

- **Keywords:** automated agents, human–automation interaction, machine learning, trust in automation, human–systems integration, uninhabited aerial vehicles

BIOMECHANICS, ANTHROPOMETRY, WORK PHYSIOLOGY

Woojin Yoon, Seobin Choi, Hyeseon Han, Gwanseob Shin. *Neck Muscular Load When Using a Smartphone While Sitting, Standing, and Walking*. pp. 868–879.

Objective: Myoelectric activity of neck extensor muscles and head kinematic variables, when using a smartphone for one-handed browsing and two-handed texting while sitting, standing, and walking, were evaluated to compare the neck muscular load during these tasks and across the posture conditions. **Background:** There has been limited research on the relation between head-down postures and the muscular load on the neck of smartphone users. **Methods:** Twenty-one asymptomatic young users were asked to perform one-handed browsing and two-handed texting tasks in each of the posture conditions, and the myoelectric activities of the neck extensor muscles, head kinematic variables, and upper back posture were quantified. **Results:** The muscle activation level when using a phone during walking was 21.2% and 41.7% higher than that of sitting and standing on average ($p < .01$). Head vertical and angular accelerations were also significantly greater ($p < .01$) for walking than for sitting and standing conditions. Between the two conducted tasks, participants flexed their heads more significantly ($p < .01$) with higher activation of the neck extensor muscles ($p < .01$) when texting as compared to when browsing. **Conclusion:** Results indicate that two-handed texting while walking would be the most physically demanding scenario for neck musculature, and it might be attributable to the dynamics of the head while walking with the head facing downwards. **Application:** These findings can be used to better understand the potential relation between smartphone use and the occurrence of neck musculoskeletal problems and to inform the users of the ergonomic risks of using smartphones while walking.

- **Keywords:** EMG, text neck, smartphone, Smombie, walking

COGNITION

Yasunori Kinosada, Takashi Kobayashi, Kazumitsu Shinohara. *Trusting Other Vehicles' Automatic Emergency Braking Decreases Self-Protective Driving*. pp. 880–895.

Objective: We focused on drivers in close proximity to vehicles with advanced driver assistance systems (ADAS). We examined whether the belief that an approaching vehicle is equipped with automatic emergency braking (AEB) influences behavior of those drivers. **Background:** In addition to benefits of ADAS, previous studies have demonstrated negative behavioral adaptation, that is, behavioral changes after introduction of ADAS, by its users. However, little is known about whether negative behavioral adaptation can occur for nonusers in close proximity to vehicles with ADAS. **Method:** Experienced (Experiment 1) and novice (Experiment 2) drivers drove a

simulator vehicle without ADAS and tried to pass through intersections. We manipulated participants' belief about whether an approaching vehicle had AEB and time-to-arrival of the approaching vehicle. Participants kept constant speed or pressed the brake pedal before entering each intersection. In Experiment 2, participants rated their trust in AEB by a questionnaire after driving. **Results:** In both experiments, belief about the approaching vehicle's AEB did not influence braking probability; however, belief delayed initiation of braking. The effect of belief on braking latency was only observed when trust in AEB was higher in Experiment 2. **Conclusion:** Negative behavioral adaptation can occur for nonusers in close proximity to users of AEB, and trust in AEB plays an important role. **Application:** When evaluating the effect of ADAS, the possible behavioral change of surrounding nonusers as well as users should be taken into account. To establish consumers' trust accurately, advertisements (e.g., TV commercials) must carefully consider their messages.

- **Keywords:** human–automation interaction, trust in automation, driver behavior, decision making, expert–novice differences

METHODS AND SKILLS

Reilly J. Innes, Nathan J. Evans, Zachary L. Howard, Ami Eidels, Scott D. Brown. *A Broader Application of the Detection Response Task to Cognitive Tasks and Online Environments*. pp. 896–909.

Objective: The present research applied a well-established measure of cognitive workload in driving literature to an in-lab paradigm. We then extended this by comparing the in-lab version of the task to an online version. **Background:** The accurate and objective measurement of cognitive workload is important in many aspects of psychological research. The detection response task (DRT) is a well-validated method for measuring cognitive workload that has been used extensively in applied tasks, for example, to investigate the effects of phone usage or passenger conversation on driving, but has been used sparingly outside of this field. **Method:** The study investigated whether the DRT could be used to measure cognitive workload in tasks more commonly used in experimental cognitive psychology and whether this application could be extended to online environments. We had participants perform a multiple object tracking (MOT) task while simultaneously performing a DRT. We manipulated the cognitive load of the MOT task by changing the number of dots to be tracked. **Results:** Measurements from the DRT were sensitive to changes in the cognitive load, establishing the efficacy of the DRT for experimental cognitive tasks in lab-based situations. This sensitivity continued when applied to an online environment (our code for the online DRT implementation is freely available at https://osf.io/dc39s/), though to a reduced extent compared to the in-lab situation. **Conclusion:** The MOT task provides an effective manipulation of cognitive workload. The DRT is sensitive to changes in workload across a range of settings and is suitable to use outside of driving scenarios, as well as via online delivery. **Application:** Methodology shows how the DRT could be used to measure sources of cognitive workload in a range of human factors contexts.

- **Keywords:** cognitive workload, multiple object tracking, detection response task

TEAMS AND GROUPS

Surabhi Pasarakonda, Gudela Grote, Jan B. Schmutz, Jasmina Bogdanovic, Merlin Guggenheim, Tanja Manser. *A Strategic Core Role*

***Perspective on Team Coordination: Benefits of Centralized Leadership for Managing Task Complexity in the Operating Room*. pp. 910–925.**

Objective: We examine whether surgical teams can handle changes in task requirements better when their formal leader and strategic core role holder—that is, the main surgeon—is central to team coordination. **Background:** Evidence regarding the benefits of shared leadership for managing complex tasks is divided. We tested whether a strategic core role holder's centrality in team coordination helps teams to handle different types of task complexity. **Method:** We observed coordination as specific leadership behavior in 30 surgical teams during real-life operations. To assess the strategic core role holder's coordination centrality, we conducted social network analyses. Task complexity (i.e., surgical difficulty and unexpected events) and surgical goal attainment were rated in a questionnaire. **Results:** In the critical operation phase, surgical difficulty impaired goal attainment when the strategic core role holder's coordination centrality was low, while this effect was nonsignificant when his/her coordination centrality was high. Unexpected events had a negative effect on surgical goal attainment. However, coordination centrality of the strategic core role holder could not help manage unexpected events. **Conclusion:** The results indicate that shared leadership is not beneficial when teams face surgical difficulty during the critical operation phase. In this situation, team coordination should rather be centralized around the strategic core role holder. Contrarily, when unexpected events occur, centralizing team coordination around a single leader does not seem to be beneficial for goal attainment. **Application:** Leaders and team members should be aware of the importance of distributing leadership differently when it comes to managing different types of task complexity.

- **Keywords:** strategic core role holder, team coordination, surgical teamwork, task complexity, social network analysis