2018 NDIA GROUND VEHICLE SYSTEMS ENGINEERING AND TECHNOLOGY SYMPOSIUM
MODELING & SIMULATION, TESTING AND VALIDATION (MSTV) TECHNICAL SESSION
AUGUST 7-9, 2018 - NOVI, MICHIGAN

HUMAN WORKLOAD MODELING FOR AUTONOMOUS GROUND VEHICLES

Vlad L. Pop, Ph.D.
W. Stuart Michelson
Human Systems Engineering
Georgia Tech Research Institute
Atlanta, GA

ABSTRACT

Accurate models of operator workload in highly automated ground vehicles could inform interface design decisions, predict performance impacts of new systems, and evaluate existing systems. This paper summarizes an existing methodology for modeling human operator workload, demonstrates its application to automated ground vehicles, and discusses its value in development, certification, and acquisition of autonomous military ground systems.

INTRODUCTION

Autonomous Ground Systems (AGS) play a significant role in the DoD's Third Offset Strategy. As technology matures and more automation-enabled vehicles are fielded the role of automation in these ground systems becomes more complex. Automation has the potential to decrease operator workload, increase efficiency, and maintain high levels of safety, as we have seen in aviation [1]. However, automation can also introduce new cognitive demands in the form of knowledge requirements, data management tasks, and attentional demands [2]. A large volume of human factors research has focused on understanding how operators interact with automation [3]. This research has demonstrated that introducing new automation without regard for human operators can lead to unforeseen problems that risk the potential gains of new technologies [1]. In order to support the step increase in performance required for the Third Offset Strategy it is critical to assess the impact on human operators as AGS are implemented into the sensor, C4I, and support grids.

Previous research we have worked on [4] as well as other studies [5-6] have demonstrated that one of the major consequences of automation is changes to operator workload. As new automation is implemented operator workload changes depending on how much monitoring of the automation is required, how much cross-checking of the automation is required, how much crew coordination is required, and how much time or effort for task management is required [4]. System designers can optimize operator workload by ensuring that automation is implemented where it is most beneficial to the operator, reduces complexity, and does not overwhelm the operator [7]. In order to achieve optimal workload in AGS designers must be able to assess the operator workload levels associated with different configurations of automation and human operators involved in the performance of missions.

The purpose of this paper is to summarize an existing methodology for modeling human operator workload, demonstrate its application to automated ground vehicles, and discuss the value for development, certification, and acquisition of autonomous military ground systems.

WORKLOAD MODELING

As military systems become increasingly more complex they require system operators to process increasingly larger amounts of information. In order to understand the performance impacts of these complex systems the Human Research and Engineering Directorate of the U.S. Army Research Laboratory (ARL) has developed, applied, and validated the Improved Performance Research Integration Tool (IMPRINT) [8].

IMPRINT was originally developed to assess the mental workload associated with different configurations of soldiers and complex equipment [7]. It has since been used in a variety of systems and demonstrated to be useful for identifying peak levels of workload that indicate which tasks should be reallocated, redesigned, or automated [9]. The tool has been successfully used to determine function allocations in U.S. Navy destroyers [10], determine the number of operators needed in Special Operations command stations [11], determine the crew size needed for the U.S. Army automated artillery system [12], determine the performance effects of the U.S. Army Land Warrior integrated fighting system [13], and most recently used by the authors of this paper to determine function allocation among crew members for new capabilities in the C-130H.

IMPRINT models human performance using a task network architecture approach. Human behavior within complex military systems is organized by the missions that operators perform. Each mission is then decomposed into smaller elements according to the tasks that must be accomplished. Tasks are continued to be broken down into subtasks until all human-system interaction is described as a closed-loop function. Once all behavior is broken down into small elements tasks and subtasks are linked together and organized according to system task sequence to build the structure of the task network model [14].

Once the structure of the task network model is built IMPRINT can be used to evaluate system performance using a detailed analysis of operator workload. Workload is analyzed at each subtask, for each operator, for each action, with each specific equipment interface. IMPRINT uses a workload evaluation method based on Multiple Resource Theory [15]. Workload is evaluated for each Visual, Auditory, Cognitive, and Psychomotor (VACP) dimension [16].

Validation studies have shown that the VACP method has good predictive validity, providing workload measures that correlate with subjective workload ratings from real operators, and predicting the same performance differences observed in real systems [7].

The VACP scale, included in Table 1, contains different 7.0 point interval scales with verbal anchors for each mental resource. Each task within the task network model is given separate visual, auditory, cognitive, and psychomotor ratings depending on the demands they place on each component. For instance, if a task requires operators to identify whether a system is ready by detecting a light, it is given a visual workload rating of 1.0, but if a task requires operators to identify whether a system is ready by reading, it is given a visual workload rating of 5.9.

Once all elements within the task network have workload ratings the model is ready to run. The model structure identifies the tasks that must be accomplished for each mission, the sequence of subtasks that are performed, which operators perform them, and the workload demands imposed. The workload profile for each mission can then be analyzed to identify aspects of the system design that are complex, overwhelming, or degrade performance.

IMPRINT has been used throughout the lifecycle of a wide array of complex military systems and may be a valuable tool for evaluating AGS.

Visual	
1.0	Visually Register/Detect (detect occurrence of image)
3.7	Visually Discriminate (detect visual differences)
4.0	Visually Inspect/Check (discrete inspection/static condition)
5.0	Visually Locate/Align (selective orientation)
5.4	Visually Track/Follow (maintain orientation)
5.9	Visually Read (symbol)
7.0	Visually Scan/Search/Monitor (continuous/serial inspection)

Auditory	
1.0	Detect/Register Sound (detect occurrence of sound)
2.0	Orient to Sound (general orientation/attention)
4.2	Orient to Sound (selective orientation/attention)
4.3	Verify Auditory Feedback (detect anticipated sound)
4.9	Interpret Semantic Content (speech)
6.6	Discriminate Sound Characteristics (detect auditory difference
7.0	Interpret Sound Patterns (pulse rates, etc.)

Cognitive	
1.0	Automatic (simple association)
1.2	Alternative Selection
3.7	Sign/Signal Recognition
4.6	Evaluation/Judgment (consider single aspect)
5.3	Encoding/Decoding, Recall
6.8	Evaluation/Judgment (consider several aspects)
7.0	Estimation, Calculation, Conversion

Psychomotor	
1.0	Speech
2.2	Discrete Actuation (button, toggle, trigger)
2.6	Continuous Adjustive (flight control, sensor control)
4.6	Manipulative
5.8	Discrete Adjustive (rotary, thumbwheel, lever position)
6.5	Symbolic Production (writing)
7.0	Serial Discrete Manipulation (keyboard entries)

Table 1: The Visual Auditory Cognitive Psychomotor (VACP) Workload Scale.

APPLICATION TO GROUND VEHICLES

The application of IMPRINT to automated ground vehicles is discussed using a detailed example. In order to keep this paper Distribution A and releasable to the public, the example is focused on evaluating the capabilities of automated systems that are currently commercially available in passenger vehicles. Nevertheless, all the methods, analyses, and applications discussed are the same across AGS.

In this example we apply IMPRINT to compare manual driving, Tesla Autopilot, and Cadillac Super Cruise™ for the task of maintaining a lane during Interstate highway driving. We begin with a task network approach and decompose the task of using automation into the following subtasks: Manually Driving, Engaging Automation, Confirming Engagement, “Driving” with Automation, Automation Disengagement, and Manual Disengagement.

Once the task is decomposed we apply the VACP workload scale to each subtask. We begin with Manual Driving. Manual Driving requires the driver to continuously scan the road and monitor the position of the vehicle in relation to the lane markings as well as other vehicles on the road, thus it receives a visual workload of 7.0 on the VACP scale. Manual Driving can be performed without necessarily requiring auditory activity, thus is receives a VACP scale rating of 0. For cognitive activity, Manual Driving requires the driver to evaluate and judge several aspects, including own speed, speed of other vehicles, and the intention of other drivers, thus is receives a cognitive workload of 6.8. Finally, for psychomotor activity, Manual Driving requires the driver to continuously adjust the steering wheel and pedals, thus it receives a psychomotor rating of 2.6 on the VACP scale.

Once the baseline workload for manual driving is assed in each VACP dimension we move on to adding the automation subtasks. We begin with Engaging the Automation. For visual workload, both Tesla Autopilot and Cadillac Super Cruise™ require the driver to visually detect a steering wheel icon on the dash (1.0 VACP rating), while also still manually driving (7.0 VACP rating), thus the total visual workload is 8.0 for both systems. No auditory activity is required for engagement, so auditory workload remains at 0. For cognitive workload, the driver had to make additional decisions about the automation in addition to the evaluations and judgements while manually driving (6.8 VACP rating). With Tesla Autopilot the driver must make a simple association between the steering wheel icon on the dash and engaging the automation (1.0 VACP rating), thus the total cognitive workload was 7.8. With Cadillac Super Cruise™ the driver must make an alternative selection, first engaging the radar cruise control, then engaging Super Cruise, thus a 1.2 VACP rating is given in addition to the 6.8 VACP rating for manual driving, for a total of 8.0 VACP rating. For psychomotor workload, both Tesla Autopilot and Cadillac Super Cruise™ require the driver to perform a discrete actuation (2.2 VACP rating) via a button or knob to activate the automation while also continuing to adjust the steering wheel and pedals (2.6 VACP rating), thus both systems receive a total psychomotor workload rating of 4.8.

Once both systems are evaluated in all VACP dimensions for the first subtask of Engaging Automation the same process is applied to the next subtask of Confirming Engagement, and then so on to "Driving" with Automation, Automation Disengagement, and Manual Disengagement. The end result of all VACP ratings, for all subtasks, for each system is included in Table 2. The results are then graphed to view workload for each dimension and in total, as demonstrated in Figure 1.

Looking at Figure 1 we can first see some trends across both systems. In general both Tesla Autopilot and Cadillac Super Cruise™ briefly increase driver workload during engagement and disengagement tasks. This is to be expected since the automation must be engaged while the driver is still also maintaining manual control of the vehicle. On the other hand, once automation is engaged, both Tesla Autopilot and Cadillac Super Cruise™ decrease driver workload in comparison to manual driving. Overall this type of analysis shows where the addition of automation is beneficial to operator workload and where it is not; however, the analysis also reveals some differences between the two systems.

Looking at Figure 1 again, we see that Tesla Autopilot induces more workload on the operator during engagement and disengagement than Cadillac Super Cruise™ does. This difference in workload stems from each system's design. With Tesla Autopilot the driver must confirm that the system is engaged by detecting a visual difference in the color of the steering wheel icon on the right of the speedometer, resulting in a visual workload rating of 3.7. With Cadillac Super Cruise™ the driver must confirm that the system is engaged by visually registering that the top of the steering wheel has now been illuminated green, resulting in a visual workload rating of 1.0.

Figure 1 also shows a difference in workload while "Driving" with Automation across the two systems. Although both systems decreased workload levels in comparison to manual driving, Tesla Autopilot requires drivers to maintain their hands on the steering wheel, thus imposing more psychomotor workload on drivers than the Cadillac Super Cruise™ system which is a hands free system.

DISCUSSION

Our goal for this paper was to summarize the IMPRINT approach to human workload modeling and discuss a detailed example of how it can be applied to inform design decisions in AGS. Our example compared manual driving, Tesla Autopilot, and Cadillac Super Cruise™ while maintaining a lane on an Interstate highway. The results demonstrated that IMPRINT can be used to evaluate the performance impacts of new ground vehicle automation in several ways. First, this methodology can be used to evaluate system design decisions. In our example we found greater visual workload with the system indicating that the automation is engaged by changing the color of an icon versus the additional presence of light or image. Second, this methodology can be used to identify when the addition of automation benefits operators and when it does not. In our example we found increased workload during automation engagement and automation disengagement but decreased workload while automation was in use. Finally, this methodology can be used to make objective comparisons across systems. For example we could quantify the difference in

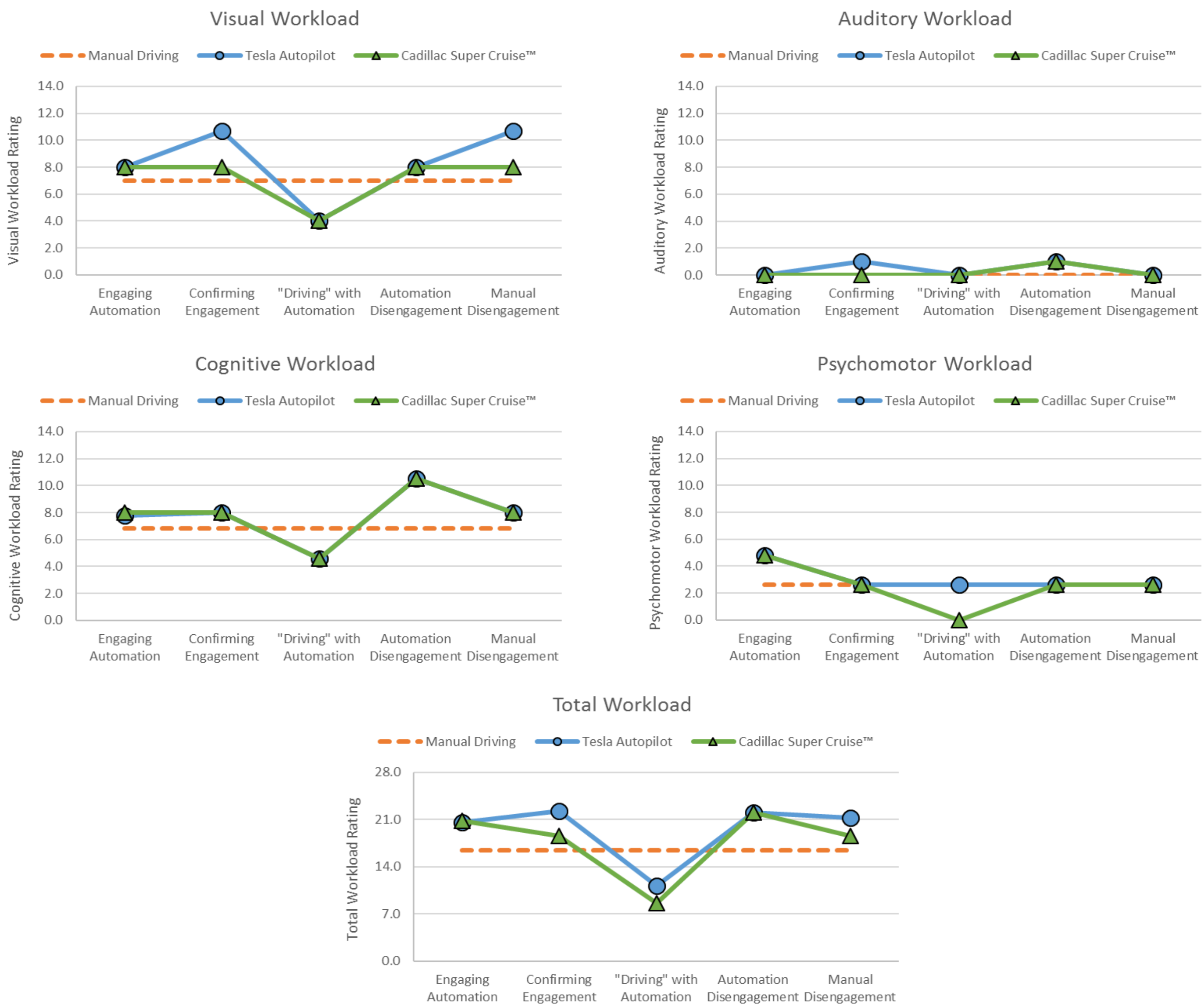

Figure 1: Graphical VACP Workload Ratings for Each Automation Subtask

	Tesla Autopilot					Cadillac Super Cruise™				
Task	Vis.	Aud.	Cog.	Psy.	Tot.	Vis.	Aud.	Cog.	Psy.	Tot.
Engaging Automation	8.0	0.0	7.8	4.8	20.6	8.0	0.0	8.0	4.8	20.8
Confirming Engagement	10.7	1.0	8.0	2.6	22.3	8.0	0.0	8.0	2.6	18.6
"Driving" with Automation	4.0	0.0	4.6	2.6	11.2	4.0	0.0	4.6	0.0	8.6
Automation Disengagement	8.0	1.0	10.5	2.6	22.1	8.0	1.0	10.5	2.6	22.1
Manual Disengagement	10.7	0.0	8.0	2.6	21.3	8.0	0.0	8.0	2.6	18.6

Table 2: VACP Workload Ratings for Each Automation Subtask.

operator workload between the hands-on and hands-off automated driving system.

The work described in this paper highlights how IMPRINT can be used to inform design decisions in AGS, but this is only the beginning of this application. Our example was purposefully simple and only had drivers engage the vehicle automation while performing a single baseline task of maintaining a lane on the highway in order to show how the model computes workload across concurrent tasks by Visual, Auditory, Cognitive, and Psychomotor dimensions. In reality the value of IMPRINT is the ability to model the potential overlap between the wide varieties of driving tasks that occur for different types of driving missions, in different environmental conditions. The application of IMPRINT to AGS should model VACP workload demands of automation tasks across the variety of secondary automotive tasks used in National Highway Traffic Safety Administration research, including vehicle device oriented tasks such as manually tuning the radio, navigation, communication, and entertainment, portable device tasks including cell phones and tablets, and non-device oriented tasks including eating, drinking, grooming, and attending to passengers [17].

The value of IMPRINT in the AGS domain spans beyond system design. IMPRINT is utilized throughout the lifecycle of a wide array of complex military systems. As AGS are developed and actual performance data becomes available the task network model within IMPRINT can be augmented with performance data for each individual task for mean time, standard deviation, distribution curve, and completion rate [18]. As a result, these types of expanded task network models have demonstrated to provide useful, valid, and accurate predictions of the situations and circumstances in which human errors will occur within a system [19].

A wide variety of previous work in other transportation domains has demonstrated that accurate quantitative models of operator workload in highly automated systems can successfully inform user interface design decisions, predict performance impacts of new systems, and evaluate existing systems. IMPRINT shows great potential for valuable applications in the AGS domain and we hope to see it further developed.

REFERENCES

[1] Durso, F. T., Feigh, K., Fischer, U., Morrow, D., Mosier, K., Pop, V. L., Sullivan, K., Blosch, J., & Wilson, J. (2011). Automation input features from the modern cockpit: toward a human-automation relationship taxonomy. Washington, DC.

[2] (Amalberti, R., & Sarter, N. B. (2000). Cognitive engineering in the aviation domain - Opportunities and challenges. In N. B. Sarter and R. Amalberti (Eds.), Cognitive engineering in the aviation domain. Mahwah, NJ: Lawrence Erlbaum Associates. (pp. 1-9).

[3] Sheridan, T. B., & Parasuraman, R. (2006). Human-automation interaction. Review of human factors and ergonomics (Vol. 1, pp. 89–129). Santa Monica, CA: Human Factors and Ergonomics Society.

[4] Durso, F. T., Stearman, E. J., Morrow, D. G., Mosier, K. L., Fischer, U., Pop, V. L., & Feigh, K. M. (2015). Exploring relationships of human automation interaction consequences on pilots: Uncovering Subsystems. Human Factors, 57(3), p. 397-406.

[5] Durso, F. T., & Alexander, A. (2010). Managing workload, performance, and situation awareness in aviation systems. In E. Salas & D. Maurino (Eds.) Human factors in aviation (pp. 217–247). London: Elsevier.

[6] Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. Journal of Cognitive Engineering and Decision Making, 2, 140–160.

[7] Mitchell, D.K. (2000). Mental workload and ARL workload modeling tools. Aberdeen Proving Ground, MD, US Army Research Laboratory, Human Research & Engineering Directorate: 35.

[8] Improved Performance Research Integration Tool, IMPRINT (2009). [Computer Software]. Available from the U.S. Army Research Laboratory website at: http://www.arl.army.mil/www/default.cfm?page=3200

[9] Cain, B. (2007). A review of the mental workload literature (Report Number RTO-TR HFM-121-Part-II). Toronto, Canada: Defence Research and Development, Human System Integration Section.

[10] Archer, R.D., & Lockett, J.F. III (1997). WinCrew - a tool for analyzing performance, mental workload and function allocation among operators. Proceedings of the First International Conference on Allocation of Functions, Galway, Ireland.

[11] Malkin, F.J., Allender, L.E., Kelley, T.D., O'Brien, P., & Graybill, S. (1997). Joint base station variant 1 MOS-workload-skill requirements analysis (Document Number ARL-TR-1441). Aberdeen Proving Grounds, MD: Army Research Laboratory.

[12] Beideman, L.R., Munro, I., & Allender, L.E. (2001). IMPRINT modeling for selected crusader research issues. Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

[13] Adkins, R., Murphy, W., Hemenway, M., Archer, R., & Bayless (1996). HARDMAN III analysis of the land warrior system (Document Number ARL-CR-291). Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

[14] Leiden, K., Laughery, K. R., Keller, J. W., French, J. W., Warwick, W. & Wood, S.D. (2001). A Review of Human Performance Models for the Prediction of Human Error. Boulder, CO: Micro Analysis and Design, Inc.

[15] Wickens, CD. (1991). Processing resources and attention. In D.L. Damos (Ed.), Multiple Task Performance (pp. 3-34). Washington, DC: Taylor & Francis.

[16] McCrasken, J. H., & Aldrich, T. B. (1984). Analysis of selected LHX mission functions: Implications for operator workload and system automation goals (Technical Note ASI479-024-84). Fort Rucker, AL: Army Research Institute Aviation Research and Development Activity.

[17] Angell, L., Auflick, J., Austria, P.A., Kochhar, D., Tijerina, L., Biever, W., Diptiman, T., Hogsett, J., & Kiger, S. (2006). Driver Workload Metrics Task 2 Final Report. National Highway Traffic Safety Administration. Report No. DOT HS 810 635.

[18] Wong, D. T. (2010). Validating Human Performance Models of the Future. Orion crew exploration vehicle, 54th Annual Meeting Human Factors and Ergonomics Society, San Francisco, CA, pp. 1002–1006.

[19] Pop, V. L. (2015). Using task network modeling to predict human error. Georgia Institute of Technology, Atlanta, GA.