Workload as a Performance Shaping Factor for Human Performance Models

Organizer

Brian F. Gore, Ph.D.
SJSURF/NASA Ames Research Center
HSI Division, MS 262-4
Moffett Field, CA 95035
650-604-2542
brian.f.gore@nasa.gov

Although reliable techniques exist for measuring operator workload (Gawron, 2008), surprisingly little attention has been directed toward the question of how workload affects performance in extended missions, particularly in extreme environments with expert operators such as those encountered in space operations. This gap in empirical research leaves a corresponding gap in computational workload models.

Performance shaping factors (PSFs), including both internal moderators (e.g., intelligence, expertise, personality, emotion, attitudes) and external moderators (e.g., physiological stressors such as fatigue and time stress) impact human performance in a variety of important ways. For example, fatigue and stress have been found to be precursors to operator errors. In aviation, it has been estimated that flight crews' alertness levels are degraded approximately 15% of the time that they are on duty leaving them vulnerable to error. In addition, excessive time on task has been found to negatively impact a human operator's vigilance, and an inverse relationship has been found between hours of wakefulness and performance on a critical task (Mallis, Mejdal, Nguyen, & Dinges, 2004). To accurately represent human behavior computationally, many aspects internal to an operator that might impact his/her performance capability needs to be accurately represented. This fosters the need for models of erroneous performance and PSFs in many of the human performance models being developed today. These PSFs need to accurately represent workload's impact.

This symposium moderated by Dr. Brian Gore from SJSURF / NASA Ames Research Center will be comprised of Dr. Christopher Wickens from Alion Science and Technology, Mr. Joe Armstrong from CAE Inc, Dr. Andrew Belyavin from *QinetiQ Ltd*, and Dr. Robert McCann from NASA Ames Research Center. Dr. Gore will highlight the motivation behind some recent workload requirements from the Space Human Factors Engineering (SHFE) project, Dr. Wickens will discuss issues that deal with human performance, task performance and workload thresholds, Mr. Armstrong will discuss efforts to incorporate human factors and modeling and simulation domains in a novel way, Dr. Belyavin will discuss validating models of the effect of moderators on human behavior, Dr. McCann will discuss recent work where he has introduced behavioral variability to HPMs using benchmarks to drive operator models, and Dr. Gore will conclude the session by providing an update on the manner that MIDAS v5 predicts workload as applied in a NextGen application.

References

Gawron, V. J. (2008). Human performance, workload, and situational awareness measures handbook (2nd ed.). Boca Raton, Florida: CRC Press, Taylor & Francis Group.

Mallis, M.M., Mejdal, S., Nguyen, T.T. &.Dinges, D.T. (2004). Summary of the key features of seven biomathematical models of human fatigue and performance. *Aviation, space, and environmental medicine*, Vol. 75, No. 3, Sect. 2, pp. A4-A14, March.

The Search for Redlines of Workload

Christopher D, Wickens, Ph.D.
Alion Science and Technology
cwickens@alionscience.com

The relationship between workload and performance has often been represented as an inverted U, with poor performance resulting when workload is quite low (often coupled with sleep disruption) and when it is quite high. This presentation will focus on the needs for modeling in the latter, high workload range. Within this range, designers and mission planners often speak of a "red line" of workload, by asking "how much workload is too much", or, at what level of task demands does performance begin to suffer because human resources are inadequate. A second question asked is, "given that the red line is crossed, how does performance break down? Are the strategies of coping with task overload at all predictable, in general, and given the context of overload?" This presentation describes how models have addressed laying out a scale upon which the redline might be placed, defining the location (and variability) of that redline, and implementing strategies for task overload management.

The Convergence of Human Behaviour Representation and Synthetic Environments

Joe Armstrong, MA, CMSP
jarmstrong@cae.com

In defence simulation, unless every entity (whether blue force, red force or civilian) is controlled by a human operator, the corresponding decision-making and behaviours of human entities must be simulated. While the cognitive science domain has developed sophisticated models representing a range of human activity and cognition instantiated in a range of tools (e.g. ACT-R, SOAR, IPME, IMPRINT) these efforts have generally existed in isolation, are highly specialized, and limited in scope. By not informing the development of defence simulation activities in a significant fashion, there remains a significant gap between the human simulation activities in the cognitive science and Human Factors (HF) domain and the ability for the community to meet the requirements for enhanced behaviours to support Modelling and Simulation applications (M&S).

One significant barrier that exists between the cognitive science domain and defence simulation relates to mismatches in conceptual abstraction. Traditionally, Human Factors (HF) combines academic level research with standardized knowledge capture from subject matter experts to understand and predict human performance. The cognitive science requirements to describe human behaviour are naturally phrased in abstract concepts, while the defence M&S requirements to model real-world events lead to more concrete constructs. HF concepts used to describe how human agents interact with the environment can become lost in translation when they are not explicitly and consistently defined for M&S. Consequently, M&S must filter this information to transform abstract concepts into the models that result in simulation.

This observation has motivated the development of techniques to span, or translate, levels of conceptual abstraction to integrate models of different resolutions into federated simulations. This is expected to bring the flexibility of more abstract formulations into the specificity of more concrete applications. Our HF and M&S teams are collaborating under a common research and development initiative to gain a shared understanding and to develop a standardized language. As such, the HF and M&S domains work together to assess captured domain knowledge to select information that is both representative of the problem space, but also technically feasible for simulation. This process allows for more efficient construction of explicit hierarchical databases (i.e., ontologies) built from the HF products, which feed the eventual model. Compromising a degree of domain specific practice in exchange for enhanced shared understanding facilitates a more rapid, realistic and reusable data integration for the human modelling community.

Validating Models of the Effect of Moderators on Human Behaviour

Andrew Belyavin, PhD
ajbelyavin@QinetiQ.com

The key to modelling the effect of many moderators on human behaviour is to break the overall model into two parts: a model of the evolution of internal state and a model of the impact of internal state on behaviour and performance. Examples of this type of model are provided by representations of the impact of the thermal environment or sleep loss on behaviour. For example, an overall model of the effect of the thermal environment on behaviour comprises a model of the impact of the environment on thermal state and model of the effect of thermal state on behaviour. In an earlier paper we have described an approach to the validation of a whole body thermal model and its prediction of human thermal state (Belyavin & Cain, 2009). Thermal state is a good analogue of many other states such as fatigue in that the observations of state form a continuous sequence of values, which can be determined by experiment. It is possible to compare observed and modelled interval data of this kind using standard statistical tools such as Analysis of Variance (ANOVA) or linear regression (Belyavin & Cain, 2009). Validating models of human behaviour is a more demanding process since the outcome measures are often a stochastic sequence of discrete responses. We have explored two possible approaches to validating models of this kind. The first approach exploits tasks which provide a sequence of low level decisions such as a tracking task to assess the effectiveness of a model (Belyavin & Cain, 2010) and the second approach uses indirect effects such as the effectiveness of training (Cain, Magee, & Belyavin, 2011) to validate the application of a model.

A more complex moderator than either sleep loss or the thermal environment is the impact of continuous working on performance – "time-on-task" – in that a model of the effect of the moderator must include two effects: the control of the task flow introduced by the operator(s) to manage workload and the impact of any continuous work on performance (Belyavin & Spencer, 2004). A model of the impact of "time-on-task" will be outlined and the challenge of validating the model in operational settings will be described.

References

Belyavin, A, & Cain, B. (2009). *An example of validating models of continuous processes.* Proceedings of 18th Conference on Behavioral Representation in Modeling and Simulation. Sundance Resort, Utah. 30 Mar – 2 Apr. 8.

Belyavin A,, & Cain B. (2010). *Validation of human behavioural models.* Poster presented at NATO Symposium HFM 202. Amsterdam.

Cain B., Magee, L., Belyavin, A. (in press, March 2011). Validation of a prototype training simulation employing HBR.

Belyavin, A.J., & Spencer, M.B.. (2004). *Modelling Performance And Alertness: The QinetiQ Approach.* Aviat Space Environ Med 2004; 75(3, Suppl.):A93–103.

Understanding and Modeling Multitasking Oculomotor Behavior by Aerospace Vehicle Operators

Robert S. McCann, Ph.D.
NASA Ames Research Center
robert.s.mccann@nasa.gov

Should a crewed spacecraft experience a systems malfunction during a dynamic flight phase, operators must process multiple sources and forms of visual information to understand and respond to the malfunction while continuing to monitor critical flight-related parameters on their primary flight display. Given the limited processing resources of the visual system, the only way to meet these multitasking requirements is to fixate serially on the relevant information sources. Substantial variability in individual fixation latencies and in the length (number) of task-related fixation sequences yields a wide range of malfunction-related task completion times. Existing computational models of human multi-tasking behavior have only limited capability to simulate oculomotor multi-tasking behavior, however, rendering them too deterministic to generate realistic distributions of task finishing times. We addressed this predictive deficiency by augmenting the modeling capabilities of "Apex", a state-of-the-art multitasking human performance model, with a Human Oculomotor Performance (HOP) module that selects individual fixation latencies, and builds fixation sequences, via stochastic processes. Simulations of human operator performance with the Apex/HOP hybrid yielded fault diagnosis latencies that covered the full range of finishing times obtained from human-in-the-loop testing. However, task completion times were slower overall for the model than for humans, and detailed analyses of human eye movements revealed several important differences with the model. These differences helped identify specific aspects of the model that require either modification or additional development, while advancing our understanding of human information acquisition and information processing behavior in complex real-world environments.

Workload as a Performance Shaping Factor in MIDAS v5

Brian F. Gore, Ph.D.
SJSU/NASA Ames Research Center
brian.f.gore@nasa.gov

The challenges associated with the measurement and management of workload from an empirical perspective have led to many different conceptualizations on the degree to which workload should influence an operator's performance. There is little question that workload does impact nominal performance but there is less agreement on precisely how workload influences performance. Some individuals thrive under periods of high task load while others fail under periods of low task load and vice versa. Representing this divergent empirical performance computationally is needed so that accurate representations of human-system interactions are generated by model analysts. The Man-machine Integration Design and Analysis System (MIDAS) possesses two distinct workload characterizations. The first is an unconstrained representation where MIDAS completes the coded tasks and outputs workload without thresholds to limit task performance. This mode of operation allows model analysts to see where the model predicts workload spikes. The second is an implementation where the MIDAS model completes the coded tasks using thresholds based on the Multiple Resource Theory (MRT). This mode of operation combines a conflict matrix and task degradation functions. The process undertaken to develop the workload model and to verify its operation in a recent NASA-FAA MIDAS v5 validation effort will be detailed in this presentation. The presentation will end with an outline of research needed to accurately model workload.