

Dynamic “What-if” Modeling Simulation

Eric M. Moyer

Sandia National Laboratories 1515 Eubank, Mailstop 0152, Albuquerque, NM, 87123

emmoyer@sandia.gov

Dynamic modeling and simulation will be used to provide an understanding of the interactions between various complex systems. This dynamic model is based on an enterprise architecture framework whereby complex, dynamic and non-linear interactions, particularly those involving the human, can be understood and analyzed. Our modeling approach will include a synthesis of top-down and bottom-up strategies. The top-down portion will analyze high-level, mandated guidance and trace its tenants down to individually identifiable activities at the worker-level. We will then model these activities through the provision of a discrete event task model emphasizing research-based human performance and cognitive workload principles (bottom-up). These principles are based on accepted theories of the interaction between cognitive workload and human error. Synthesizing these two approaches will demonstrate both the impact and effect of high-level mandated activities and aid analysts in their understanding of how, why and when these impacts help or possibly hinder humans at the worker level. Benefits of using this model, namely the ability to predict “what if” scenarios in real time will be discussed. The model will be tested across multiple domains to demonstrate the potential modeling approach and its application in future hazard analyses.

1. Introduction

When developing a new system, it is the goal of the designing engineers to use as much objective evidence as possible for the basis of the new design. Requirements are not pulled from thin air, but ride on a paradigm of known and quantified performance parameters. Requirement generation for the mechanical aspects of a new system can often be modeled using known performance capabilities of legacy systems and fine tuning them to project the performance of the new system. These known performance capabilities, along with the newly modeled capabilities are a crucial platform in understanding the scope of the new system, and by rights should bear the brunt of early efforts.

Consequently, and in most cases, human machine interaction (HMI) decisions are often made later in the design process. In some ways this order of operation makes sense. As an example, if we consider an aircraft design, the early focus would be on the aero-mechanical performance requirements of the proposed aircraft, such as how fast it can go, how far it can go, how much weight it can carry, etc., since these considerations are most likely the primary performance requirements of the stake holders. Even with systems where the HMI elements are communicated and considered early, their actual design cannot truly begin until the mechanical envelope in which the human will reside is determined. Also, many engineers prefer not to analyze elements of the system that include human-in-the-loop decisions too early, since it's perceived that these events are more nondeterministic than their mechanical counterparts. Therefore, the impact of HMI requirements, while often captured early, are often not able to be explored until the mechanical system is somewhat mature. This can be problematic when working to ensure HMI design considerations are considered early, and can sometimes lead to delays in performing hazard analyses, which could lead to late-arriving design decisions. Therefore, how can human factors analysts provide objective insight on human performance considerations and subsequently safety concerns early in the system development process to affect not only preliminary designs, but possibly even requirements?

2. High Level Guidance

System designs often draw requirements not only from stakeholders, but also from mandated guidance such as industry specifications and standards. Mandated requirements can push design decisions early in the life-cycle, and it's important to understand these at the earliest possible stages. Two such standards are the Department of Defense (DOD) Design Criteria Standard, Human Engineering, or MIL-STD-1472G and the DOD Standard Practice System Safety, or MIL-STD-882E. MIL-STD-1472G is comprised of design criteria intended to be included as individual requirements within system specifications aimed at ensuring best-practice human engineering considerations are included as part of system designs. Criteria are meant to be included individually as requirements themselves to accommodate human capabilities across a multitude of possible system interaction scenarios.

MIL-STD-882E presents the DOD approach to analyzing and mitigating hazards and reducing risk overall for systems, products, equipment and infrastructure. These approaches allow systems analysts methods through which system design decisions can be analyzed according to system risk. These two documents are just examples. It isn't easy to understand the impact mandated requirements can make on a system during the early design process, especially if those requirements involve human performance without some form of reference. The relationship between high-level requirements/mandated guidance and worker-level activities is displayed below in Figure 1.

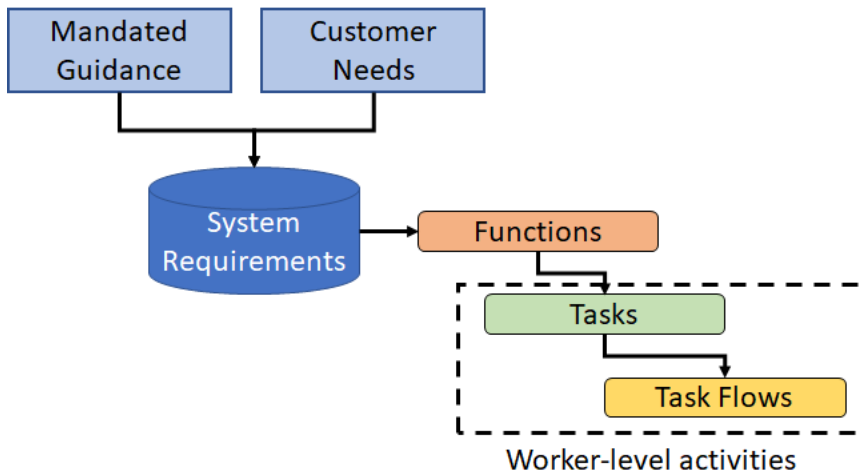


Figure 1: Relationship Between High-level Requirements and Worker-level Activities

3. Worker-Level Activities

High-level and early design decisions create the foundation upon which each subsequent design is based. These decisions must include all instances where humans are involved in the execution of system functions. These instances are often captured through the creation of a task list. It's from this list that system designers understand the initial details surrounding how humans will interact with all parts of the system, be they hardware or software. This task list is born from a process called task analysis. Which is a comprehensive process that seeks to understand all activities that occur on the part of a discrete individual or team employing a defined set of tools or methods, surrounding the accomplishment of a specific goal (Kirwan & Ainsworth, 1992). Task analyses are performed to ensure that engineers and analysts completely understand the scope of activities, techniques, tools and methods for an identified task or set of tasks. Once completed, the task analysis can be used as the basis for additional analyses to define the performance of the humans involved in the system (Hackos & Redish, 1998). Task analyses consist of a functional analysis, task inventory and task flows. A functional analysis aims to determine the functional steps of the scenarios of use to achieve a goal with a system, device, or process under specific conditions (Alexander & Maiden, 2004). The generation of task inventories and task flows are a foundational approach that provides the basis of analysis for human performance evaluations. After the functions are defined, they are then analyzed and broken down into individual tasks, from which task inventories and flows were assembled. This can be accomplished by interviewing subject matter experts who are familiar with the domain of operations, and refining their input based on the anticipated activities of the new system.

For example, if the system under development is based on an existing system, the users of the existing system will be able to provide a starting point for understanding the tasks of the new system. Their knowledge, combined with the anticipated user interface design elements will allow for the formulation of anticipated functions and tasks of the new system. These tasks and task flows can then be used as basis for dynamic models of human performance for the new system.

4. Workload and Human Performance

Workload refers to "... a mental construct that reflects the mental strain resulting from performing a task under specific environmental and operational conditions, coupled with the capability of the operator to respond to those demands." (Cain, 2007, p. 4-3). The image below in Figure 2, from Yerkes and Dodson (1908), depicts the relationship between performance and engagement/stress (labelled arousal, in 1908 terms) which has come to be accepted as the basic model for understanding human performance in relation to stress. It depicts the relationship between stress and performance, in that as a task or set of tasks require more effort, cognitively or otherwise, the associated workload to complete those tasks rises, and as it crosses some threshold of tolerance of the individual their performance starts to wane, and their stress level rises. This is essentially the basis for the current theory concerning workload as a predictor of individual performance for a task or set of tasks (Paas, 1992). While the concept of high workload or workload overload is often discussed, the concept of low workload, or workload underload is also of importance. Workload underload is when the individual is essentially bored with a task and therefore is not engaged in it, which would also result in a detriment to performance. Measuring workload allows for better understanding of the performance demands of given tasks, which improves prediction of operator and system performance (Cain, 2007).

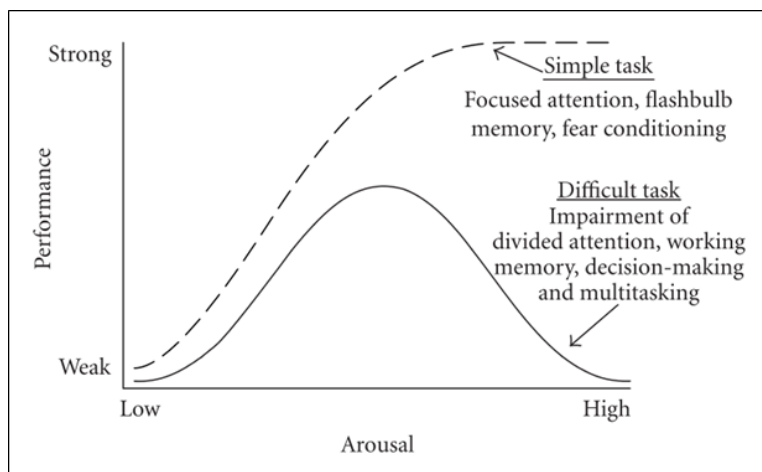


Figure 2: Yerkes and Dodson (1908) Arousal/Engagement/Stress Curve

5. Workload Modeling and What-If Scenarios

Workload models can be constructed and analyzed in many ways, however, certain popular workload theories lend themselves to discrete event simulation models. The Improved Performance Research Integration Tool (IMPRINT) is a discrete event, simulation and modeling software tool that allows the user to input quantifiable system parameters and task flows (Mitchell, 2000, 2003). The simulated task flows are run over a certain amount of time to produce an estimated measure of workload. For each task in the task network, the software computes the time-to-complete the task flow, along with overall workload. In IMPRINT, the workload measures are based on the VACP (visual, auditory, cognitive, and psychomotor) (McCracken & Aldrich, 1984) theory and the Multiple Resource Theory (MRT) of workload (Wickens and Yeh, 1986). These theories postulate that individuals possess channels of capacity from which resources can be utilized to complete differing types of tasks. Tasks draw resources from different channels and this concept formulates a measure of the difficulty of the task in question, with more difficult tasks drawing more resources from more channels. Rating scales have been developed that capture common scores for different types of activities (Szabo and Bierbaum, 1986) and have been in use for some time. Table 1 displays the rating scale values for the cognitive workload channel of the VACP and MRT theories.

Table 1: Cognitive Channel Workload Ratings

Workload Score	Activity Type
0.0	No Cognitive Activity
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

The core of the model revolves around the resource cost values assigned to each task and whether an individual task performer's resource allocation exceeds an established limit. Normally this value is held at 60, and it accounts for parallel task execution while incorporating MRT (Wickens and Yeh, 1986) modifiers for tasks similar and dis-similar. This threshold is used as an indicator that an individual could be overloaded and marks a point in the task flow that should be examined more closely to cull out possible task flow changes, alternative system designs or crew task reassignments to avoid performance decrements. Analogously, while no threshold value is identified within MRT, situations of underload should also be examined since they too can lead to performance decrements (Swain, 1964). Additionally, task performance times can also be predicted and incorporated into the model by leveraging research-based human performance time estimates for activities such as speech listening rate (Miller & Licklider, 1950); hand movement rates (Welford, 1968); human-computer interaction and decision-making activities (Card et al., 1983); and fixation and target-finding times (Houtmans & Sanders, 1984). These estimates are used to generate models of human performance times for task flow activities, which are then compiled to compute overall task-performance times. An example of one of these models is illustrated below.

Choice reaction time (Card et al., 1983) – model used to represent an individual's selection of a correct option as a function of the number of alternatives options available. $P1$ = Number of possible alternatives.

$$15 * \log_2(P1+1) \quad (1)$$

By utilizing IMPRINT (or similar modeling tool) an analyst can build tasks and task flows, assign them workload resource channel values and modeled task execution times, execute those task flows temporally, and then observe the level of workload an individual would experience at any given time based on the tasks they are performing; thus, generating a model of workload for specific activities. An example of an IMPRINT model output is provided below in Figure 3.

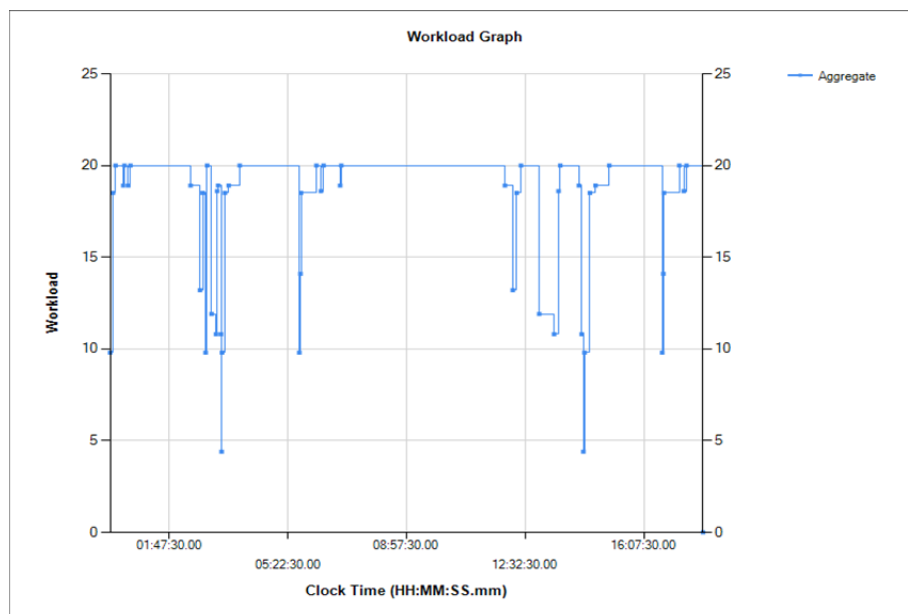


Figure 3: IMPRINT Workload Model Example

The model output displayed, indicates aggregated workload values, over time, for all utilized channels from the VACP theory, while also incorporating the concept of channel interference as postulated by MRT. These values are what are measured against the theoretical resource limit of 60, and can be directly compared across models to determine differences. This process can be repeated over multiple instances, in “what-if” types of analyses, where design parameters or even mechanical and physical elements can be changed and the resulting models created and compared. These models provide a glimpse into how these changes might affect HMI performance without the need for physical mock-ups or test-beds. Once identified, these instances can be further analyzed in terms of driving requirements or design constraints, which could theoretically be changed to accommodate human performance concerns demonstrated through these models. Workload models can become integrated in the end-to-end system analysis with results providing a feedback loop to requirements and design changes as displayed in Figure 4.

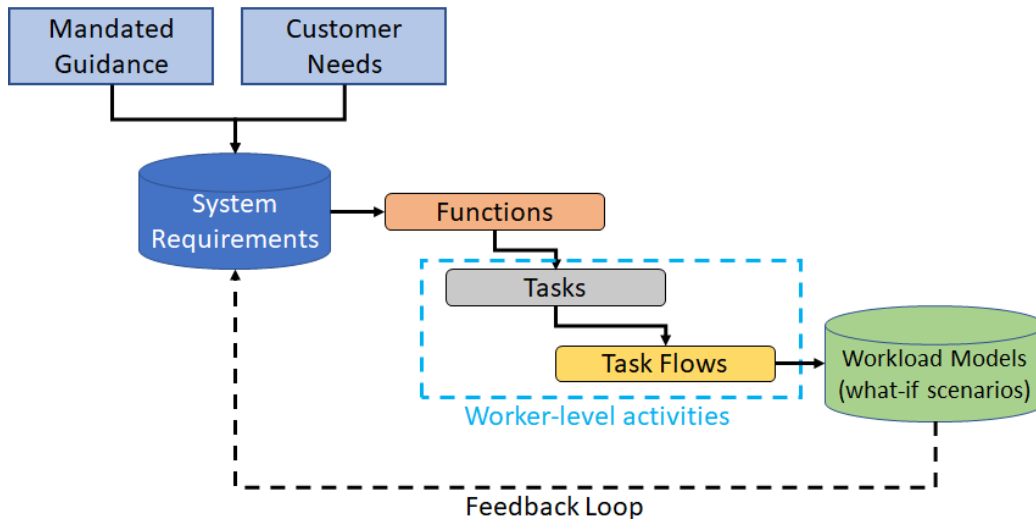


Figure 4: Workload Model Feedback Loop

6. Conclusions

By modeling worker-level activities through this method, a better picture of human-in-the-loop activities can be constructed early in the design cycle, and traced back to high-level requirements. It is not conclusive that indications of high workload for tasks occurring in models such as those discussed in this paper correlate with higher incidents of human performance failures (Hancock & Warm, 1989). However, designing systems with the goal of streamlining the HMI would pose little doubt to minimizing situations where the human user would be subjected to high-workload conditions for which many would argue are more “stressing” for the individual than low or moderate workload conditions (Paas, 1992). This type of modeling can be used for any system, from chemical plant control centers, to aircraft cockpits; wherever a human interfaces with a machine to perform prescribed activities as part of a system.

References

- Alexander, I. F., and Maiden, N., (Eds.), 2004, Scenarios, stories, use cases through the systems development life cycle, Wiley, New York, NY, USA.
- Cain, B. (2007) A Review of the Mental Workload Literature, Report Contract No.: RTO-TR-HFM-121-Part-II, Defence Research and Development Canada Toronto Human System Integration Section, Toronto, CA, 4-1-4-32.
- Card, S. K., Moran, T. P., and Newell, A., 1983, The Psychology of Human-Computer Interaction, Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, USA.
- Hackos, J.T., Redish, J.C., 1998, User and Task Analysis for Interface Design, John Wiley & Sons, Inc., New York, NY, USA.
- Hancock, P. A., and Warm, J. S., 1989, A dynamic model of stress and sustained attention, Human Factors, 31, 519–537.
- Houtmans, M. J. M., and Sanders, A. F., 1984, Perception of Signals Presented in the Periphery of the Visual Field, Acta Psychologica, 55, 143–155.

- McCracken, J. H. and T. B. Aldrich., 1984, Analysis of selected LHX mission functions: Implications for operator workload and system automation goals (Technical Note ASI479-024-84), Army Research Institute Aviation Research and Development Activity, Fort Rucker, AL, USA.
- Miller, G. A. and Licklider, J. C. P., 1950, The Intelligibility of Interrupted Speech, *Journal of the Acoustic Society of America*, 22, 167–173.
- Mitchell, D. K., 2000, Mental Workload and ARL Workload Modeling Tools (ARL-TN-161), Aberdeen Proving Ground, MD, USA.
- Mitchell, D. K., 2003, Advanced Improved Performance Research Integration Tool (IMPRINT) Vetronics Technology Test Bed Model Development (United States, Army Research Laboratory, Human Research & Engineering Directorate), Army Research Lab, Aberdeen Proving Ground, MD, USA.
- Paas, F., 1992, Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach, *Journal of Educational Psychology*, 84, 429–434.
- Swain, A.D., 1964, THERP (Report SC-R-64-1338) Sandia Corp., Albuquerque, NM, USA.
- Szabo, S. M. and Bierbaum, C. R., 1986, A comprehensive task analysis of the AH-64 mission with crew workload estimates and preliminary decision rules for developing and AH-64 workload prediction model (Technical Report ASI678-204-86[B], Vol I, II, III, IV), Anacapa Sciences, Inc., Fort Rucker, AL, USA.
- Welford, A. T., 1968, *Fundamentals of Skill*, Methuen, London, UK.
- Wickens, C. D., and Yeh, Y. Y., 1986, A multiple resource model of workload prediction and assessment. In *Proc. IEEE Conf. Systems, Man, Cybernetics*, 1044–1048.
- Yerkes, R., Dodson, J., 1908, The Relation of Strength of Stimulus to Rapidity of Habit-Formation *Journal of Comparative Neurology and Psychology*, 18, 459–482.