Simulation-based Assessment of Performance-Workload Tradeoffs for System Design Evaluation

Christina F. Rusnock
Department of Systems Engineering and Management
Air Force Institute of Technology
Wright-Patterson AFB, OH 45433 USA

Christopher D. Geiger
Department of Industrial Engineering and Management Systems
University of Central Florida
Orlando, FL 32816 USA

Abstract

Ideally, when selecting a system design or re-design, the selected option should balance workload and performance. However, previous research suggests that the relationship between workload and performance is nonlinear, thus achieving one, does not imply achieving the other. To ensure that a system will meet performance and workload goals, system designers should investigate the expected workload and performance tradeoffs of the potential system designs. This study uses discrete event simulation to demonstrate how to perform workload-performance tradeoffs when designing or re-designing a system, using a baseline system and three alternative designs. By examining the relationship between workload and performance, the study demonstrates how to identify the impacts that design alternatives can have on workload and performance. The study finds that four types of decision circumstances can occur: 1) a single preferred alternative is identified, 2) multiple preferred alternatives are identified, 3) no preferred alternative is identified, and 4) alternatives achieve conflicting goals. Each of these circumstances has unique implications for the design selection decision.

Keywords
workload, performance, human performance modeling, discrete event simulation

1. Introduction

1.1 Background

Theories of cognitive workload, whether structural bottleneck [1], single capacity [2, 3], or multiple resource theories [4, 5], provide explanation for declines in performance as task load increases. In seeking to explain overload and decrements of performance in multi-tasking situations, these theories imply a linear, negative relationship between workload and performance [6].

However, underload is also recognized to impact performance negatively [7]. This suggests that workload and performance are better characterized by a concave, rather than linear, relationship; with both low and high workload associated with low performance, and peak performance occurring at some "middle" range of workload. Thus, achieving optimal human performance cannot simply be done with the reduction of workload, but rather it requires balancing workload to ensure that neither overload nor underload occurs.

In addition to performance goals, managing workload can produce other valuable outcomes. Reducing workload to a manageable level can increase safety, reduce stress, and reduce fatigue. Increasing workload above an underload state, can aid in reducing boredom, reducing distraction, and maintaining operator skill sets [8].
While the workload-performance continuum can be a concave function, the specific shape of the function will vary from human-computer system to human-computer system. To ensure that a desired system meets performance and workload goals, system designers should investigate the expected workload and performance tradeoffs of the potential system designs.

1.2 Purpose
The purpose of this paper is to demonstrate how to perform workload-performance tradeoffs when designing or redesigning a system. Characterizing the relationship between workload and performance enables system designers to effectively 1) predict impact of system designs on workload and performance and 2) make design decisions based on workload and performance goals.

Since this research focuses on evaluating potential system designs, the system designs are in the conceptual stages of development. In addition, this research uses discrete-event simulation to help to characterize the relationship between workload and performance in order to evaluate the workload-performance tradeoffs of an existing system and other potential system designs.

2. Method

2.1 Baseline System Overview
The case study involves an operator responsible for multiple autonomously-navigated unmanned vehicles performing surveillance operations. The operator receives inputs from these unmanned vehicles onto a visual display, the Operator Control Unit (Figure 1). The OCU is comprised of a computer monitor and a computer mouse. The operator monitors the inputs, evaluates the information, and then responds using the computer mouse, as appropriate.

![Figure 1: Operator Control Unit](image)

2.2 Baseline System Tasks
The operator is responsible for two separate tasks. The first is a threat detection task in which the operator is responsible for distinguishing between friendly and enemy forces. In this case, an unmanned ground vehicle is conducting surveillance through a hostile urban environment. As the vehicle travels through the city streets, a live-camera feed is displayed for the operator, depicting various actors. The operator is responsible for distinguishing between threats (enemy soldiers and armed civilians) and non-threats (friendly soldiers and friendly civilians), as shown in the examples in Figure 2. When a threat is detected, the operator indicates this by using the mouse to select the Threat Detect button on the left side of the screen, and then uses the mouse to select the identified threat (Figure 3).
The second task is a change detection task. In this task, multiple unmanned air vehicles (UAVs) are performing surveillance over the hostile village, tracking numerous points of interest. During the mission, the UAVs provide information on the locations of the points of interest. The operator perceives these updates as appearances, disappearance, and movements of the icons on the lower half of the screen, representing the surveillance map (Figure 4). As the operator observes changes, the operator identifies the type of change that occurred, by using the mouse to select the button directly above the surveillance map that corresponds with the type of change that occurred (i.e., Appeared, Disappeared, Movement). Each task is modeled using discrete-event simulation.
2.3 Experimental Design
These tasks are combined into four scenarios, each consisting of three task loads, which constitute 12 variants. In Scenario 1, only the change detection task is performed at a low, medium, and high event rate. In Scenario 3, only the threat detection task is performed at a low, medium, and high event rate. Scenarios 2 and 4 are dual-task scenarios, in which both tasks are performed simultaneously. In Scenario 2, the change detection task varies between low, medium, and high event rates, with the threat detection held constant at the medium rate. In Scenario 4, the threat detection task varies between low, medium, and high event rates, with the change detection held constant at the medium rate. Table 1 summarizes the event rates for the 12 variants.

Table 1: System Variant Event Rates (“CD” means Change Detection and “TD” means Threat Detection)

<table>
<thead>
<tr>
<th>Task Load</th>
<th>Scenario 1</th>
<th>Scenario 2</th>
<th>Scenario 3</th>
<th>Scenario 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Low</td>
<td>CD: 6 changes/min</td>
<td>CD: 6 changes/min</td>
<td>CD: none</td>
<td>CD: 12 changes/min</td>
</tr>
<tr>
<td></td>
<td>TD: none</td>
<td>TD: 28 actors/min</td>
<td>TD: 14 actors/min</td>
<td>TD: 14 actors/min</td>
</tr>
<tr>
<td>B: Med</td>
<td>CD: 12 changes/min</td>
<td>CD: 12 changes/min</td>
<td>CD: none</td>
<td>CD: 12 changes/min</td>
</tr>
<tr>
<td></td>
<td>TD: none</td>
<td>TD: 28 actors/min</td>
<td>TD: 28 actors/min</td>
<td>TD: 28 actors/min</td>
</tr>
<tr>
<td>C: High</td>
<td>CD: 24 changes/min</td>
<td>CD: 24 changes/min</td>
<td>CD: none</td>
<td>CD: 12 changes/min</td>
</tr>
<tr>
<td></td>
<td>TD: none</td>
<td>TD: 28 actors/min</td>
<td>TD: 56 actors/min</td>
<td>TD: 56 actors/min</td>
</tr>
</tbody>
</table>

The experimental design for this study is a 4x4, consisting of four human-computer system designs (i.e., the baseline, plus three alternative designs) and the four scenarios described above.

2.4 Procedure
This study demonstrates how to perform workload-performance analysis for a baseline system that is being considered for redesign. A human-in-the-loop experiment of the tasks described above is used to develop a Baseline discrete-event simulation model of the system, for use in evaluating the system and system alternatives.

The human-in-the-loop laboratory study is part of the Physiologically-based Robot Interaction as Multimodal Exchanges Phase 2 (PRIME 2) study. PRIME is an ongoing research project for the U.S. Army Research Development, and Engineering Command (RDECOM) as part of the Human Robot Interaction Analysis for Training Simulations & Operational Neuroscience that is performed by the Applied Cognition and Training in Immersive Virtual Environments Laboratory in the Institute for Simulation & Training at the University of Central Florida. In the human-in-the-loop study, 150 participants (who represent the human operator in the human-computer system) perform all four scenarios at each of the three task loads in a randomized order. A baseline discrete-event simulation model of the system and task is constructed from a detailed task analysis of each task and using task time data from the human-in-the-loop study. All discrete-event simulation models in this study are created using the Improved Performance Research Integration Tool (IMPRINT). Each model is simulated for 10 replications. Based on the low variability, it is determined that additional replications are unnecessary. The workload values from the baseline model are validated using subjective and physiological workload data from the human-in-the-loop studies. See Rusnock and Geiger [9] for more details on the baseline model creation and validation.

2.5 Alternative System Designs
In addition to evaluating the workload and performance for the baseline human-computer system, this study also includes analyses of three re-designs of the baseline system.

Alternative Design 1: Left-Handed Keyboard. The first re-design alternative is the addition of a left-handed keyboard for performing the change detection task. With this configuration, the left-hand rests on the keyboard, while the right hand remains on the mouse. For the change detection task, when an Appearance occurs, the “A” key is pressed; when a Movement occurs, the “S” key is pressed, and when a Disappearance occurs, the “D” key is pressed. The threat detection task is performed the same way as in the baseline model, using the mouse. This position of the left hand on the keyboard and the right hand on the mouse is a common configuration in many PC games. This configuration allows the operator to respond to a change and select a threat simultaneously during the
dual-task scenarios. Using the keyboard is expected to increase operator efficiency because this design allows for solely a key press, where as responding to a change in the baseline requires moving and tracking the mouse to the correct button on the screen and then pressing the left mouse button.

**Alternative Design 2: Voice Recognition.** The second re-design alternative is the incorporation of voice recognition software for performing the change detection task. With this configuration, the operator states the change type that has occurred out loud. The threat detection task is performed the same way as in the baseline model, i.e., using the mouse. This configuration allows the operator to respond to a change and select a threat simultaneously during the dual-task scenarios. The voice recognition software is assumed to be trained for the operator, 100% reliable, and detects all responses spoken (i.e., the operator does not have to repeat responses).

**Alternative Design 3: Touch Screen.** The third re-design alternative replaces the standard computer monitor screen with a touch screen that is the same size as the monitor. In this configuration, the mouse is eliminated. The operator identifies threats by using his or her right index finger to select the threat (eliminates the Threat Detect button). The operator identifies changes by touching the corresponding button on the screen using his or her right index finger. This configuration assumes that the operator only uses his or her right hand, and thus can only respond to one event at a time. The right hand is assumed to rest at the far edge of the right side of the screen, when not responding to threats or changes.

### 2.6 Workload and Performance Outcomes

The workload values predicted by the simulation models of the tasks use the Visual, Auditory, Cognitive, and Psychomotor (VACP) model of Bierbaum, Szabo, and Aldrich’s [10], which is an adaptation of McCracken and Aldrich’s [11] VACP model. The VACP model builds upon multiple resource theory by capturing cognitive workload demands across seven separate resource changes: Visual, Auditory, Cognitive, Fine Psychomotor, Gross Psychomotor, Speech, and Tactile. Each sub-task performed is rated by level of demand on a scale from 0-7 for each resource channel, where zero represents no demand, and 7 represents the highest level of demand. At each point in time, workload values are calculated for each channel, and summed within and across channels. These cognitive workload values are then augmented with an interference value that represents the potential conflict of using more than one channel simultaneously. The interference values are task- and channel-specific and are based on the work of Wickens [12]. The operator’s total workload value is calculated continuously as tasks are performed. These values are, then, transformed into time-weighted averages for each variant.

Performance for the change detection task is calculated as the total number of changes identified by the simulated operator divided by the total number of changes in that variant. Performance for the threat detection task is calculated as the total number of threats identified by the simulated operator divided by the total number of threats in that variant. Dual-task performance is calculated as the average performance of the two single tasks.

### 3. Results

#### 3.1 Scenario 1: Change Detection Single Task

Figure 5 shows the workload-performance curves for the baseline system and the three alternative re-designs for Scenario 1, the Change Detection single task. Each point on the curve represents the respective workload and performance values for a particular taskload (low, medium, high) of the Scenario 1 variant for each of the 10 replications. Thus, each design’s curve consists of 30 points. The curves in Figure 5 reveal considerable overlap between the designs. However, all three alternative designs achieve higher performance than the baseline system. Furthermore, Alternative 1 is able to achieve this higher performance while also reducing the average and peak workload value experienced. Re-designs 2 and 3, on the other hand, produce slightly higher average, and peak workload values than the baseline system. The curves shown in Figure 5 also reveal that for this Scenario, workload has a quadratic shape, with the highest performance for all three systems occurring in the middle workload values. For example, the performance gains from Alternative 1 are greatest at a workload value of 7, but these gains practically disappear at the lowest (6.5) and highest (7.5+) workload values.
Table 2 summarizes the average workload and performance values for the baseline system and the three re-design alternatives for Scenario 1. These values reveal the difference in performance between the baseline and the redesigns, while also highlighting the similarity in performance between the three alternatives. Since Alternative 1, the Left-Handed Keyboard, achieves both decreased workload and increased performance, it would be the preferred re-design option for the Change Detection single task.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>7.227</td>
<td>0.402</td>
<td>69.0%</td>
<td>0.062</td>
</tr>
<tr>
<td>Alt 1</td>
<td>7.068</td>
<td>0.356</td>
<td>75.1%</td>
<td>0.072</td>
</tr>
<tr>
<td>Alt 2</td>
<td>7.317</td>
<td>0.449</td>
<td>75.7%</td>
<td>0.008</td>
</tr>
<tr>
<td>Alt 3</td>
<td>7.339</td>
<td>0.505</td>
<td>75.6%</td>
<td>0.068</td>
</tr>
</tbody>
</table>

3.2 Scenario 2: Dual Task, Variable Change Detection
Figure 6 presents the workload-performance curves for the baseline system and the three alternative designs for Scenario 2, the Dual-Task with variable Change Detection. The curves reveal staggered differences in workload between the design options. Alternative 2 results in higher average and peak workload than the baseline, while Alternatives 1 and 3 each result in lower workload than the baseline. In fact, the peak workload values for Alternatives 1 and 3 correspond with the minimum workload values for the Baseline and Alternative 2.
Furthermore, all three alternative re-designs are able to achieve higher average and peak performance than the baseline system. The curves shown in Figure 6 reveal that for this Scenario, the relationship between workload and performance vary greatly from system to system. Due to the quadratic shape of Alternative 1’s workload-performance curve, it achieves some of the highest and lowest performance scores. Alternatives 2 and 3 have relatively flat workload-performance curves, suggesting that increases in workload produce negligible changes in performance. Both Alternative 1 and 3 are able to avoid excessively high workload values (10+).

Table 3 summarizes the average workload and performance values for the baseline system and the three re-design alternatives for Scenario 2. These values reveal the difference in performance between the baseline and the alternative re-designs, while also highlighting the similarity in performance between the three alternatives. Alternative 1, the Left-Handed Keyboard and Alternative 3, the Touchscreen, result in decreased workload and increased performance. Thus, either of these would be the preferred re-design options over the Baseline.

### Table 3: Average Workload and Performance Values for Scenario 2

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>10.785</td>
<td>0.326</td>
<td>74.3%</td>
<td>0.045</td>
</tr>
<tr>
<td>Alt 1</td>
<td>9.375</td>
<td>0.475</td>
<td>77.7%</td>
<td>0.042</td>
</tr>
<tr>
<td>Alt 2</td>
<td>11.119</td>
<td>0.639</td>
<td>76.1%</td>
<td>0.053</td>
</tr>
<tr>
<td>Alt 3</td>
<td>9.173</td>
<td>0.578</td>
<td>76.7%</td>
<td>0.049</td>
</tr>
</tbody>
</table>

3.3 Scenario 3: Threat Detection Single Task

Figure 7 presents the workload-performance curves for the Baseline system and the three alternative designs for Scenario 3, the Threat Detection single task scenario. Since Alternative 1 (the Left-Handed Keyboard) and Alternative 2 (the Voice Recognition System) only affect the operations for the Change Detection task, these alternatives are expected to have no statistical difference from the Baseline system. Alternative 3, the Touch Screen, produces lower average and peak workload values, with a slight decline in performance. The curves shown in Figure 7 reveal that for this Scenario, the relationship between workload and performance is relatively flat, suggesting that fluctuations in workload produce negligible changes in performance.
Table 4 provides the average workload and performance values for the Baseline system and the three design alternatives for Scenario 3. These values show negligible differences in average performance and a decline in workload for Alternative 3. Based on the standard deviations, there are no statistical differences between any of these systems.

<table>
<thead>
<tr>
<th></th>
<th>Average Workload</th>
<th>Workload st dev.</th>
<th>Average Performance</th>
<th>Performance st dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>2.990</td>
<td>2.310</td>
<td>93.9%</td>
<td>0.057</td>
</tr>
<tr>
<td>Alt 1</td>
<td>2.990</td>
<td>2.310</td>
<td>94.2%</td>
<td>0.054</td>
</tr>
<tr>
<td>Alt 2</td>
<td>2.969</td>
<td>2.308</td>
<td>94.0%</td>
<td>0.057</td>
</tr>
<tr>
<td>Alt 3</td>
<td>2.031</td>
<td>1.717</td>
<td>92.4%</td>
<td>0.065</td>
</tr>
</tbody>
</table>

3.4 Scenario 4: Dual Task, Variable Threat Detection

Figure 8 presents the workload-performance curves for the Baseline system and the three alternative designs for Scenario 4, the Dual Task with variable Threat Detection. Alternative 2, the Voice Recognition System, attains higher performance than the other three design options; however it also attains higher average workload. Alternative 3, the Touch Screen, is the only option that produces lower average and peak workload than the Baseline system. The curves shown in Figure 8 reveal that for this Scenario, all four systems experience declining performance with increased workload. With Alternative 3, the rate of decline is the most dramatic, but its lowest levels of performance are equivalent to the other systems’ lowest levels, and it is able to avoid excessively high workload values (11+).
Table 5 provides the average workload and performance values for the Baseline system and the three design alternatives for Scenario 4. These values show that Alternative 3 achieves a large decrease in workload, compared with the other systems, without impacting performance.

Table 5: Average Workload and Performance Values for Scenario 4

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>11.481</td>
<td>2.319</td>
<td>74.8%</td>
<td>0.045</td>
</tr>
<tr>
<td>Alt 1</td>
<td>12.057</td>
<td>2.712</td>
<td>76.9%</td>
<td>0.050</td>
</tr>
<tr>
<td>Alt 2</td>
<td>11.999</td>
<td>2.722</td>
<td>78.8%</td>
<td>0.055</td>
</tr>
<tr>
<td>Alt 3</td>
<td>9.486</td>
<td>1.141</td>
<td>77.3%</td>
<td>0.054</td>
</tr>
</tbody>
</table>

4. Discussion
The results of this study demonstrate that relationship between workload and performance is complex, and a reduction in workload does not always lead to an increase in performance. When considering design options for a new system, or redesigning an existing system, it is important to consider what effects these changes might have on both workload and performance.

This study examines four different designs using four different tasks. By doing so, the study was able to provide examples of four sets of decision circumstances:

4.1 Result 1: A Single Preferred Alternative
For Scenario 2, workload is significantly lower for Alternatives 1 and Alternatives 3, but average and peak performance is significantly higher only for Alternative 1. Thus Alternative 1 is clearly the best option, maximizing both workload and performance benefits.

4.2 Result 2: Multiple Preferred Alternatives
For Scenario 1, all three alternative designs produce significant increases in performance, with no statistically significant impact to workload. Since a performance increase can be achieve with any of the three alternatives, the decision maker can look to other considerations, such as cost or technical complexity.
4.3 Result 3: No Preferred Alternatives

For Scenario 3, there was no statistically significant difference between any of the options and the Baseline system. This suggests that either a redesign may be unnecessary or other design options should be considered.

4.4 Result 4: Conflicting Goals – Tradeoff Required

For Scenario 4, only Alternative 3 produces statistically lower workload than the Baseline, but only Alternative 2 produces statistically higher performance than the Baseline. Thus, there is a tradeoff between achieving workload and performance goals. The decision maker must decide which takes precedence: reducing workload or increasing performance.

5. Conclusion

This study uses discrete event simulation to establish workload-performance curves for potential design alternative for a system redesign. By examining the relationship between workload and performance, the study demonstrates how to identify the impacts that particular design alternatives can have on workload and performance. The study finds that four types of decision circumstances can occur: 1) a single preferred alternative is identified, 2) multiple preferred alternatives are identified, 3) no preferred alternative is identified, and 4) alternatives achieve conflicting goals. Each of these circumstances has unique implications for the design selection decision. While some instances will enable reduction of workload and increases in performance, other instances may result in improving only one of these facets, or may require the decision maker to tradeoff workload and performance.

Acknowledgements

This work was supported in part by the U.S. Army RDECOM (W91CRB08D00150068) and the contract research support of Macklin Everly, Wyle Aerospace Group (HC1047-05-D-4005 TAT RI-12-0487/DO 0232). 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 RDECOM or the U.S. Government.

References