Expanding GOMS-HRA from Analog to Digital Human-Machine Interfaces

Thomas A. Ulrich and Ronald L. Boring
Idaho National Laboratory, Idaho Falls, USA

Abstract: The majority of human reliability analysis (HRA) methods were developed for analog based human-machine interfaces and their applicability to digital HMIs has come in question. Digital HMIs change the nature of the task by presenting information in a windowed interface that requires the user to navigate between displays to access information or controls. GOMS-HRA is a new method developed to analyze human reliability based on the times required to perform tasks in conjunction with the typical human error probability. GOMS-HRA quantifies tasks by decomposing and mapping tasks onto basic task level primitives. It is based off of the goals, objects, methods, and selection rules (GOMS) cognitive model. This paper uses GOMS-HRA to examine operator-in-the-loop study data on nuclear operators interacting with analog and digital HMIs. The results, though inconclusive, suggest digital HMIs may require longer times to complete the same task level primitives compared to analog HMIs. However, this result may be attributed to a deficiency in the GOMS-HRA methodology, since it does not easily account for navigation tasks required for digital HMIs. The GOMS-HRA model was revised to include an additional task level primitive for navigation to allow the methodology to better capture the nuances of interactions for digital HMIs.

Keywords: GOMS-HRA, Human Reliability Analysis, Analog and Digital HMIs, Nuclear Process Control, Nuclear Main Control Room

1. INTRODUCTION

In the U.S., nuclear power plant operators use largely analog based human-machine interfaces (HMIs) when monitoring and controlling the plant from the main control room. Many of the currently operating reactors are seeking license amendments to operate beyond their original 40-year lifespan. As these plants seek license extensions, they are contending with obsolete, difficult to maintain, and expensive analog hardware comprised of discrete instrumentation and controls (I&C). Plants are adopting digital control systems to modernize and eliminate their reliance on the antiquated analog technology and take advantage of the new digital HMI capabilities. Digital HMI affords a number of advantages including real time sharing of data, data aggregation and visualizations, continuous monitoring, and fault diagnostics [1]. These new digital capabilities change the cognitive nature of the operators' task since operators have more information available to ascertain the state of equipment. Furthermore, the digital HMIs typically use a windowed interface, which requires operators to access information and controls. This is in stark contrast to the always visible discrete controls and displays in existing analog control rooms. To understand and demonstrate human error is within as low as reasonably achievable (ALARA) levels [2], human reliability analysis (HRA) methods are employed. These methods were developed for analog I&C, and the applicability of these methods to digital HMIs is questionable [3]. This paper examines empirical timing data from an analysis of a group of reactor operators working with analog I&C and a group of reactor operators working with a digital HMI to assess any differences in timing between the two formats.

A literature review was conducted by [3] to compare human error potential (HEP) estimates from canonical HRA methods, including THERP, ASEP, SPAR-H, and HEART, to empirically observed HEP estimates with digital interfaces. In their analysis, reports were classified as object level and holistic level error data. Object level errors involve a particular interface object such as a text entry field or menu structure, while the holistic level errors involve a failure at the system level such as failing to diagnose an alarm or perform a plant function such as turbine startup. The authors analysis revealed the estimates for both object and holistic level errors differed by a significant magnitude in the empirical studies compared to the HRA method estimates. Specifically, the authors noted THERP provides overly optimistic estimates of error in comparison to the observed values from the included studies in the literature review. The authors conclude a meaningful and significant difference between the measures and observed HEP estimates using digital HMIs and therefore the HRA methodologies must be revised to accommodate the shift towards more digitally based control rooms.

HRA methods appropriate to capture the new interactions between operators and digital HMIs are needed to provide accurate analyses of human error in the upcoming modernized and more digital based main control rooms. Goals-Operators-Methods-Selection rules-Human Reliability Analysis (GOMS-HRA) [4], is a promising new HRA method with the potential to adequately address digital interactions when analyzing human error potential.

GOMS, an acronym for its four comprised components, i.e. goals, operators, methods, and selection rules, is a cognitive model of human-computer interaction first proposed by [5]. Goals represent tasks users strive to achieve, operators are the available actions users can take, methods are the steps or combinations of actions the user performs to achieve the goal, and selection rules are decisions about which methods to perform the achieve the goal. GOMS is used to analyze interactions within the context of specific goals set by the user of a system. GOMS can produce both qualitative and quantitative diagnostics of an individual performing a task. To perform a GOMS analysis, the task must be sufficiently decomposed into basic units of analysis. Central to the GOMS methodology is observing timing data for each of the basic units of analysis so that an accurate model of the human-computer interaction can be derived. This model then becomes a powerful tool since it can now be used to predict and identify more efficient paths towards specific goal. Furthermore, the model can be used to understand how the time to complete a task changes when adding additional subtasks, such as inserting a safety check.

GOMS-HRA, proposed by [4], relies on the Keystroke-Level Model (KLM) variant of GOMS, which classifies tasks based on the series of operators required to achieve the goal [6]. The KLM offers a suite of basic units or subtasks to represent the individual operators at the lowest level of the model. KLM quantifies the time to complete a goal by summing the duration for each subtask across the operators to yield an overall timespan for the task. Boring and Rasmussen leveraged the concept of the basic suite of operators from KLM along with an expansion of the Systematic Human Error Reduction and Prediction Approach (SHERPA) error taxonomy [7] to identify process control specific operators to support GOMS analysis within an HRA framework. The resulting operators, referred to as task level primitives (TLPs), can be seen below in Table 1 (from Boring and Rasmussen, 2016).

Table 1: GOMS Task Level Primitives (TLPs)

TLP	Description
A_C	Performing required physical actions on the control boards
A_F	Performing required physical actions in the field
C_C	Looking for required information on the control boards
C_F	Looking for required information in the field
R_C	Obtaining required information on the control boards
R_F	Obtaining required information in the field
I_P	Producing verbal or written instructions
I_R	Receiving verbal or written instructions
S_C	Selecting or setting a value on the control boards
S_F	Selecting or setting a value in the field
D_P	Making a decision with procedure guidance
D_W	Making a decision without procedure guidance
W	Waiting

The GOMS-HRA approach was first demonstrated in an example probabilistic risk assessment (PRA) analysis of a station blackout scenario analogous to what occurred during the Fukushima Daiichi accident. GOMS-HRA was developed as part of the Human Unimodel for Nuclear Technology to Enhance Reliability (HUNTER) method [8]. HUNTER is a computation-based human reliability analysis framework developed to support dynamic PRA. Existing dynamic PRA methods, such as RAVEN which is based on RELAP-5, are capable of accurately simulating the reactor coolant system and core physics for operational transients and accident scenarios to characterize the risks from various equipment failures associated with postulated accident scenarios. The HUNTER framework allows for dynamic modeling of human actions in the form of a virtual operator comprised of a simplified model of human cognition and decision and action capacity. HUNTER was designed to couple with the RAVEN framework to dynamically model key human actions, HEPs, and action timings associated with those actions by using a GOMS-HRA modelling approach.

The GOMS-HRA primitive timing data [9] was gathered from an operator-in-the-loop study conducted as part of a larger control room modernization effort under the light water reactor sustainability program. The timing data consisted of procedure start and completion timestamps along with operator dialogue. This data was coded with the task level primitives, and timing distributions were assigned. The HEPs for the simulation were taken from THERP. The timing data was sampled from a control room study and therefore, only the primitives pertaining to control board activities were examined, i.e., all field activities denoted by the subscript f were discarded. Together the timing and HEP values were used to simulate the human actions occurring during the station blackout in conjunction with the plant physics. A Monte Carlo simulation was performed and the various combinations of the human actions and plant states generated the distribution of outcomes in which core damage was either prevented or occurred.

The timing data used in this simulation stemmed from an operator-in-the-loop study using an analog platform and thus the timing data reflects operator activities in the context of analog instrumentation and controls. Only primitives pertaining to control room activities were included, and therefore the subscript denoting field or control room activities will be dropped henceforth in this paper. The goal of this research is to address the question as to whether the previously collected GOMS-HRA timing data can be used in a digital context. There are several reasons to suspect that there may be some differences in the time required to perform the task level primitives in a digital setting. In particular, the navigation requirement for accessing displays in digital interfaces is not readily captured by any of the task level primitives. Additionally, the mouse input for selection versus physical manipulation is fundamentally different and could result in a difference in the time required to perform physical actions or selecting or setting value on the different formats. Obtaining information may or may not differ between the two formats. Digital interface technology affords greater flexibility in value presentation, and therefore it could prove easier to perceive and interpret values. To address this question of how this timing data may differ in an analog versus digital format for the GOMS-HRA primitives, the timing data from another operator-in-the-loop study using a digital system was collected and analyzed. The authors predicted there would be meaningful differences between the timing data for the analog and digital HMI formats.

2. METHOD

2.1. Operator-in-the-Loop Studies

The timing data was collected from two different operator-in-the-loop studies completed as part of a larger control room modernization effort. The first study was performed in the Human Systems Simulation Laboratory (HSSL) at the Idaho National Laboratory. The HSSL is a full-scope, fully reconfigurable glasstop simulator capable or digitally representing a nuclear power plant control room. Timing data was collected only for activities involving mimics of analog I&C. The second study was performed on a full-scope glasstop simulator at a collaborating utility simulator facility. Both simulators were configured to mimic the front panels of the control room. The two studies did use different plants; however, both plants were quite similar in that they were of similar vintage pressurized water reactor design. Additionally, both studies focused on the turbine control system operation with scenarios covering the activities performed during a turbine start up. These scenarios included latching the turbine, ramping the turbine, testing the overspeed protection systems, and valve testing. Indeed, since these two plants are of such similar design, the standard Westinghouse start up procedure, GP-005, was used in both studies. Of course, the procedures are customized for each plant variant, but the general flow of the procedures and the terminology between the two procedures is quite similar.

2.2. Data Recording and Analysis

Both studies collected the timing date using an observer working with a digital pre-populated task and procedure log. As the operators completed the procedures, the observer followed along and checked off each task, which automatically logged a timestamp for the corresponding task. This pre-populated task and procedure log reduced the recording error made by the observer since he was not required to log times, but rather indicated with the press of a key when each observed task was performed. Two researchers then independently went through and classified each task as one of the GOMS-HRA primitives. Procedure level primitives were mapped to task level primitives [10]. This process resulted in a data set comprised of a total of 253 analog and 179 digital HMI format observations of GOMS-HRA primitives and times. As this was a study performed in a main control room, there were no observations pertaining to field activities. Additionally, there were insufficient observations for the W (waiting) task level primitive and therefore this primitive could not be examined.

3. RESULTS

A 2 (HMI format: analog vs. digital) by 6 (task level primitive type) analysis of variance (ANOVA) was performed to analyze the time observation data. Note that the W (waiting) task level primitive, commonly used for monitoring periods, is implicitly variable and is not included in this analysis. The results of the ANOVA can be seen in Table 2, while a graphical representation of the main effects and interactions can be seen in Figure 1. There was a significant effect for HMI format, with the analog format consistently demonstrating shorter times for each task level primitive. A post hoc analysis using Tukey's honest significance difference (HSD) showed significant differences in time for the A, C, I, and R primitives.

Table 2: Results from the ANOVA analysis for the analog and digital format and primitive type

Source	SS	df	MS	F	p
HMI Format	18089.13	1	18089.13	53.88	< 0.001
Primitive Type	19884.20	5	3976.84	11.85	< 0.001
Format * Primitive Type	10500.35	5	2100.07	6.26	< 0.001

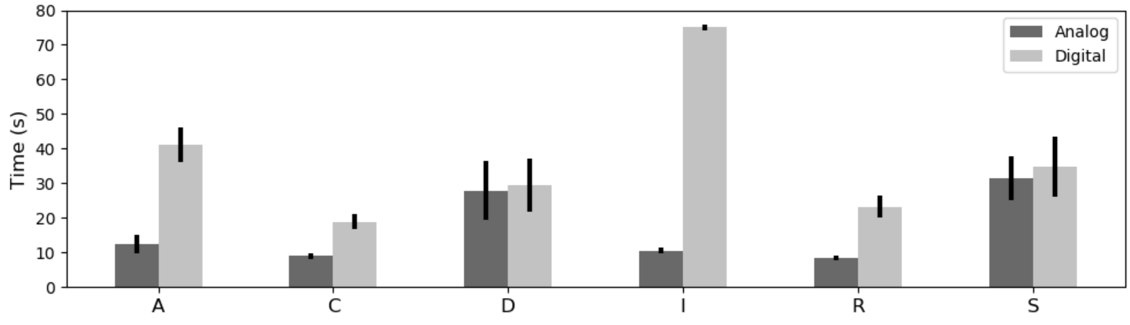

Figure 1: Main effects and interactions for HMI format and primitive type analysis, error bars represent standard error.

4. DISCUSSION

The results suggest there is a meaningful difference in the time required to perform the various primitives when using an analog versus a digital interface. Overall the digital interface demonstrated longer primitive times. It should not be assumed that these numbers necessarily reflect a generalizable finding of slower performance for digital interactions. The study may suggest that familiarity may have confounded the results. The operators are trained extensively on the analog interface, while the operators using the digital interface had only a short training session prior to the study. As a result, the longer times observed consistently across the primitives for the digital interface may suggest that a lack of familiarity may be a causal factor for taking more time to perform the primitive. This is also reflected in the timing data for the decide (D) primitive, which did not demonstrate a significant difference between interface formats. The decide primitive is a cognitive action that should not be affected by the interface, since it is occurring internally within the operators and should not have been influenced by particular technological implementation. Familiarity with the interface is irrelevant since a functional understanding of the system, which did not change between interfaces, should impose the same cognitive demands and result in similar times to make a decision. The action (A) and select (S) primitives also provide evidence for familiarity being a confounding factor. The action primitive requires the operator to identify the control and then manipulate it. The manipulation itself is negligible, and the difference in time between the interfaces can likely be attributed to the additional time required for the operators to find the control in the digital interface with which they were less familiar. As a counterargument, the select (S) primitive requires an extra and time demanding activity of entering the desired value in addition to locating the correct control within nested windows on the screen. The time to find the control is likely still greater for the digital interface, but since both the analog and digital interface require a time consuming entry activity, this offsets any search times and resulted in similar times for the primitive between both interface formats.

Despite the confounding familiarity issue, the magnitude of the increased times observed for the digital interfaces provides evidence that there may be a fundamental difference in the time required to perform primitives using a digital interface. If only a main effect was found for interface format, then familiarity could account for the increased time required to perform primitives in the digital interface. However, since there was a significant interaction, there is evidence that familiarity alone cannot account for the increased time required to perform primitives using the digital interface. There are some fundamental differences between the analog and digital interfaces worth noting that may account for the increased time. Since the digital interface does not display all necessary controls, navigation is required. The current GOMS-HRA primitives can only implicitly capture the navigational time. For example, the time required to perform a control action (A) also includes the time required to navigate to the appropriate display to manipulate the desired control.

This issue highlights a potential shortcoming of the existing GOMS-HRA primitive suite for digital applications. An additional primitive may be required to capture this variance and prevent it from being lumped into the other primitive classifications. Therefore, it is proposed to include a new navigation (N) task level primitive for GOMS-HRA applications with a digital interface. The inclusion of this navigation primitive would eliminate some of the discrepancy between the digital and analog time required to perform each primitive. An examination of the timing data from the digital interface study did yield some specific instances in which the procedure step explicitly refers to a navigation task. These steps were excluded from the primary analysis since they were specifically navigational steps that could not be appropriately classified by the other primitives nor linked to comparable activities in the analog data. The mean time for these navigational steps was 29.67 s (SD = 15.50 s). A correction factor of this magnitude would eliminate any discrepancy between the times for the digital and analog primitives. It is important to note that this mean time for navigation represents instances in which navigation was explicitly called for by the procedure; however, other tasks implicitly require navigation and a thorough analysis of the tasks would be required to determine how many and which also contain a navigation component. Many procedure level primitives would require navigation while others may not depending upon whether the indicator or control of interest happens to be located on the same display or window as a prior task. Therefore, applying this correction uniformly would be inappropriate, and it can merely be speculated at this stage that it could eliminate the significant difference between the digital and analog formats.

The increased time to perform primitives using a digital interface observed in this study should not be interpreted as conclusive or as evidence for not adopting digital interfaces in nuclear process control.

First and foremost, the data are inconclusive and merely reflect the GOMS-HRA methodology as a means to assess the time to perform these primitives. Furthermore, the digital interfaces here, though adequate for operators in a nuclear process control task, do not represent the advanced graphical capabilities the digital interfaces can afford. Automated information presentation that synthesizes disparate data sources into an aggregated view can provide an enormous advantage for providing the operating with state of the system in an intuitive manner [11]. Leveraging these advanced principles reduces the need for a complicated series of windowed displays and shortens the time required for the operator to perceive and interpret the information. Further, the timing data do not imply any decreases in operator reliability for digital HMIs. The authors emphasize the importance of embracing digital interfaces.

4. CONCLUSION

As this research moves forward, a number of avenues are being explored. A follow-up study is of great interest to record timing data for operators with experience working with the new digital system to eliminate the lack of familiarity confound. Advanced digital interfaces that leverage advanced visualizations merit examination to expand upon this basic digital interface evaluated in this current study. Lastly, automating the process of mapping the tasks and procedures to task level primitives would greatly increase the usability and reduce the time required to perform the analysis, since manually coding the tasks and procedures is a tedious and laborious process. A deep learning neural network based approach is proposed for this process and has already experienced some success in the authors' limited application, but it is premature and the results of this process were not included in this paper. A follow-on paper will describe this technique, which may also be applicable to other HRA areas.

5. DISCLAIMER

The opinions expressed in this paper are entirely those of the author and do not represent official position. This work of authorship was prepared as an account of work sponsored by Idaho National Laboratory, an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately-owned rights. Idaho National Laboratory is a multi-program laboratory operated by Battelle Energy Alliance LLC, for the United States Department of Energy under Contract DE-AC07-05ID14517. This research was funded through the Laboratory Directed Research and Development program at Idaho National Laboratory.

6. REFERENCES

[1] G. F. Franklin, J. D. Powell, & M. L. Workman, "Digital control of dynamic systems", Addison-Wesley, 1998, Menlo Park, CA.

[2] R. L. Boring. "As low as reasonable assessment (ALARA): Applying discount usability to control room verification and validation", Risk, Reliability and Safety: Innovating Theory and Practice, Proceedings of the European Safety and Reliability Conference, pp. 950-955, (2016).

[3] E. M. Hickling and J.E. Bowie. "Applicability of human reliability assessment methods to human–computer interfaces", Cognition, technology & work, 15(1), pp. 19-27, (2013).

[4] R. L. Boring and M. Rasmussen. "GOMS-HRA: A method for treating subtasks in dynamic human reliability analysis", Risk, Reliability and Safety: Innovating Theory and Practice, Proceedings of the European Safety and Reliability Conference, pp. 956-963, (2016).

[5] S. Card, T. Moran, and A. Newell, "The Psychology of Human-Computer Interaction," Lawrence Erbaum Associates, 1983, Hillsdale, NJ.

[6] S. K. Card, T. P. Moran, and A. Newell. "The keystroke level model for user performance time with interactive systems", Communications of the ACM, 23, pp. 396-410, (1980).

[7] N. A. Stanton, P. M. Salmon, L. A. Rafferty, G. H. Walker, and C. Baber, "Human Factors Methods: A Practical Guide for Engineering and Design, Second Edition", Ashgate Publishing, 2013, Aldershot, UK.

[8] R. Boring, D. Mandelli, M. Rasmussen, S. Herberger, T. Ulrich, K. Groth and C. Smith, "Human Unimodel for Nuclear Technology to Enhance Reliability (HUNTER): A framework for computational-based human reliability analysis", 13th International Conference on Probabilistic Safety Assessment and Management (PSAM 13), pp. 1-7 (2016).

[9] T. Ulrich, R. L. Boring, S. Ewing, and M. Rasmussen, "Operator timing of task level primitives for use in computation-based human reliability analysis", Advances in Intelligent Systems and Computing, 589, pp. 41-49 (2017).

[10] R. L. Boring, M. Rasmussen, T. Ulrich, S. Ewing, and D. Mandelli, "Task and procedure level primitives for modeling human error", Advances in Intelligent Systems and Computing, 589, pp. 30-40, (2017).

[11] R. Boring, T. Ulrich, R. Lew, C. Kovesdi, B. Rice, C. Poresky, Z. Spielman, and K. Savchenko, "Analog, Digital, or Enhanced Human-System Interfaces? Results of an Operator-in-the-Loop Study on Main Control Room Modernization for a Nuclear Power Plant", INL/EXT-17-43188, Idaho National Laboratory, Idaho Falls: ID, (2017).