

Proceedings of the Human Factors and Ergonomics Society Annual Meeting

<http://pro.sagepub.com/>

Mental Workload Modeling in an Integrated Cognitive Architecture

Shi Cao and Yili Liu

Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2011 55: 2083

DOI: 10.1177/1071181311551434

The online version of this article can be found at:

<http://pro.sagepub.com/content/55/1/2083>

Published by:



<http://www.sagepublications.com>

On behalf of:



[Human Factors and Ergonomics Society](#)

Additional services and information for *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* can be found at:

Email Alerts: <http://pro.sagepub.com/cgi/alerts>

Subscriptions: <http://pro.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://pro.sagepub.com/content/55/1/2083.refs.html>

>> [Version of Record](#) - Sep 1, 2011

[What is This?](#)

Mental Workload Modeling in an Integrated Cognitive Architecture

Shi Cao and Yili Liu
University of Michigan
Ann Arbor, Michigan 48109 USA

Mental workload analysis is an important component in the test and evaluation of human-machine systems. Existing empirical workload measures have limited applicability when human-in-the-loop tests are impractical, which produces the need for theory-based workload modeling and prediction methods. ACTR-QN is a theory-based integrated cognitive architecture combining the advantages of Adaptive Control of Thought-Rational (ACT-R) and Queueing Network (QN) architectures. The research reported in this paper proposes and examines a theory and method for modeling and visualizing mental workload in ACTR-QN. Validation with an empirical study of a semantic judgment task showed that an ACTR-QN model produced both performance and mental workload data similar to the human results. In addition, different components of the multi-dimensional mental workload can be visualized with ACTR-QN. Mental workload modeling in ACTR-QN provides a new tool for human factors evaluation of mental workload.

INTRODUCTION

As system operations grow more complex in cognitive demands, mental workload analysis becomes of increasing importance in human factors tests and evaluations. Although there is still no clear definition of mental workload, researchers have generally agreed that: (1) mental workload is related to the relationship between task demands and the capacity of human information processing; (2) it is multi-dimensional (Wickens, 2008); (3) it cannot be measured by primary task performance alone. The goal of mental workload analysis is to identify and help prevent both mental overload, which may decrease performance potentials and cause fatigue, and underload, which may lead to the loss of vigilance.

Two groups of methods can be used to assess mental workload. Human-in-the-loop empirical testing methods include performance measures such as secondary task methods, physiological measures such as blink rate, and subjective measures such as NASA-Task Load Index (TLX, Hart, 2006). The other group is computational modeling methods, which include task-analysis-based methods (e.g., Mitchell & Samms, 2009; Wang, Cain, & Lu, 2010) and cognitive-architecture-based methods (e.g., Gray, Schoelles, & Sims, 2005; Wu & Liu, 2007).

This paper introduces a cognitive-architecture-based method for mental workload modeling. The architecture is ACTR-QN, which integrates Adaptive Control of Thought-Rational (ACT-R, Anderson et al., 2004) and Queueing Network (QN, Liu, Feyen, & Tsimhoni, 2006) cognitive architectures. The server structure of ACTR-QN is illustrated in Figure 1.

ACT-R represents the human mind as a production rule system. It assumes two types of knowledge

representations: declarative chunks and production rules (rules, for short). A chunk's retrieval time and error rate are determined by its activation level, which is jointly determined by the chunk's learning history and association with other chunks. Rules represent procedural knowledge in the form of condition-action (IF-THEN) pairs, and its action will be fired when its condition matches the current "mental state". A mental state consists of the state of each module, and each module is a cognitive component, such as the vision module and the declarative module. ACT-R "thinks" and "acts" by firing rules until a goal state is reached.

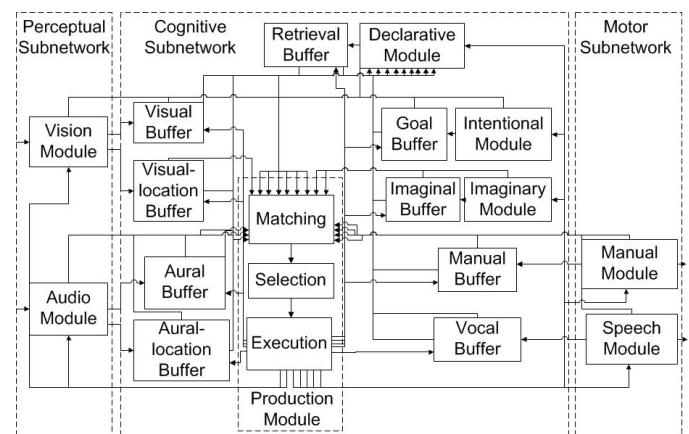


Figure 1. The server structure of ACTR-QN.

The QN cognitive architecture represents the human mind as a queueing network, where information processing is the process of servers holding and processing entities. Each server represents millions of neurons from a certain functional field of the brain, and each entity represents certain information to be processed and passed within the brain. Compared with ACT-R, QN has the advantages of visualizing mental

information processing, defining and visualizing mental workload (Wu & Liu, 2007), and modeling multi-task performance (Liu, 1997; Wu & Liu, 2008). However, QN currently does not have the capability to model a wide range of complex cognitive activities like those that can be modeled with ACT-R. To benefit from both QN and ACT-R architectures, ACTR-QN integrates the two cognitive architectures by representing ACT-R as a QN. A completed study (paper submitted for review) has implemented ACTR-QN in a C# based discrete event simulation platform Micro Saint Sharp (www.maad.com). ACT-R modules and their buffers are represented as servers in QN. ACT-R buffer requests, chunks, production rules, and the notice of completion are represented as entities that flow between these servers. ACTR-QN can read the same ACT-R codes to define a model's task-specific knowledge and parameters. It has been verified that ACTR-QN is able to generate identical mental processes and results as ACT-R for the ACT-R 6.0 tutorial models (act-r.psy.cmu.edu), and it also replicated three models of threaded cognition (Salvucci & Taatgen, 2008), a multitask performance theory implemented in ACT-R.

The work reported in the paper further develops the function of ACTR-QN by adding a theory and method for modeling and visualizing mental workload. Mental workload was quantified by server utilization, which has been used in a previous QN modeling study (Wu & Liu, 2007) to model driver workload and its NASA-TLX measurements. In queueing theory, utilization is the proportion of time during which the system's resources are used by the entities in the system. Considering server capacities in QN as mental resources in the brain, one can intuitively link utilization with workload. Since different servers/modules represent different mental resources, the utilization values from different servers processing different entities can be used to estimate different components of the multi-dimensional workload. For example, the utilization value computed from the visual module processing text entities will represent the visual verbal perception component of mental workload.

Although the current released version of ACT-R does not include any mental workload modeling theory and method, several efforts have been made to propose and test possible workload modeling methods with ACT-R (Gray et al., 2005; Jo, Myung, & Yoon, 2010; Lebiere, 2001). In comparison, the work in this paper has its unique values. First, it modeled both the error rate performance and the mental workload NASA-TLX data from an empirical study (Colle & Reid, 1998) with the same model. Second, it can visualize the multi-dimensional components of mental workload.

METHOD

ACTR-QN uses server utilization to estimate mental workload. We assume that, in general, mental workload has a linear relationship with server utilization. At the implementation level, the instant utilization of a server/module is 1.0 when it is busy (processing any entity) and 0.0 otherwise. The averaged utilization over a second, for example, equals to the busy time of the module within the second. ACTR-QN computes utilization for all servers except the intentional module and all the buffers, because their processes do not take any time, which is the same case in ACT-R.

In the current implementation, the utilization of a network is computed as the average of all server utilizations in the network. The perceptual sub-network utilization is averaged from the vision and the audio modules. The cognitive sub-network utilization is averaged from the production, the declarative, and the imaginary modules. The motor sub-network utilization is averaged from the motor and the speech modules. These three utilization values correspond to the workload components of perception, cognition, and response, which are the three levels of the processing stage dimension in the multiple resource theory view of mental workload (Wickens, 2008). Finally, the overall utilization of a model is computed as the average of the three sub-network utilization values. We assume that the overall utilization is linear to aggregated subjective workload measurements such as NASA-TLX's overall workload rating.

For model validation we used the empirical data from Colle and Reid (1998) Experiment 2, which is a semantic category judgment task. This study was selected because it can demonstrate both ACT-R's advantage in modeling complex cognitive performance and QN's advantage in modeling mental workload embedded in ACTR-QN. In each trial, participants judged whether or not two words shown on the screen belong to the same category by pressing one of two buttons. Words were randomly selected from 10 categories, each of which had 10 words. All words and categories were frequently used ones such as "dog – animal." Each trial had a fixed display duration, and there was no time interval between successive trials. A response was counted as correct if and only if it was the correct classification and was made within the display duration of a trial. Trials came in groups that each lasted 2 min 20 s. The presentation rate in a group was manipulated (6, 14, 22, 30, 38, 46, 54, and 62 word pairs/min) as an independent variable. The two dependent variables used in this validation were percent

correct and NASA-TLX overall mental workload. The experiment had another independent variable, context effect, but its effect was not statistically significant. Therefore, the results used in this validation were averaged over different levels of context effect.

A model was built in ACTR-QN to simulate the empirical data. The task-specific knowledge was defined using the same syntaxes as ACT-R. Each word-category relationship was defined as a declarative chunk. For example,

(pA0 ISA property object "A0" attribute category value A)

defined that an abstract word "A0" belongs to an abstract category "A". The initial goal chunk for each trial had its judgment, 1st word, 2nd word, 1st category, and 2nd category all unknown. Eight production rules were defined to represent the procedural knowledge. Table 1 shows their English descriptions.

Table 1. English descriptions of the model's production rules.

Rule	English description
Attend	IF the judgment is unknown, visual-location buffer has a visual location, and visual module is free, THEN encode the content of the visual item.
Retrieve-first	IF 1 st word is unknown, visual buffer has a text chunk, and declarative module is free, THEN 1 st word is the text, retrieve its category, and maintain the visual buffer chunk.
Retrieve-failed	IF the judgment is unknown, visual buffer has a text chunk, and retrieval failed, THEN retrieve the text's category, and maintain the visual buffer chunk.
First-retrieved	IF 1 st category is unknown, and a category is retrieved, THEN 1 st category is the category, find visual-location of the second word.
Retrieve-second	IF 1 st word is known, 2 nd word is unknown, visual buffer has a text chunk, and declarative module is free, THEN 2 nd word is the text, retrieve its category, and maintain the visual buffer chunk.
Second-retrieved	IF 1 st category is known, 2 nd category is unknown, and a category is retrieved, THEN 2 nd category is the category.
Yes	IF the two categories are the same, and manual module is free, THEN press the yes key, judgment is yes, and the goal is done.
No	IF the two categories are different, and manual module is free, THEN press the no key, judgment is no, and the goal is done.

The default values of the ACTR-QN's parameters are identical to those in ACT-R. To model the particular task situation of Colle and Reid (1998), only three parameters were changed from their default values, two of which were set according to the related ACT-R methods and only one was adjusted to fit the data. Specifically, parameter :esc (enable subsymbolic computations) was set to true to allow random noise in chunk activation. Parameter :ans (activation noise s), which controls the scale of chunk activation noise, was set at 0.5 following the *paired* word-number memorization model in the ACT-R 6.0 tutorial.

Parameter :blc (base-level constant) was set at 1.8 to fit the empirical results. Since the experiment used frequent words and categories and controlled the learning effect, we assumed that the learning effect was negligible and chunk activation did not change during the experiment. As a result, we did not turn on the declarative chunk learning mechanism.

The task environment (i.e., displays and controls) was set using the *single-discrete-two-stage* template provided in ACTR-QN to simulate the empirical experiment. Each presentation rate level was repeated three times, and the results were averaged before analyses. Detailed module activity traces, response results, and server and network utilization were recorded.

RESULTS

The model fitness for percent correct results is illustrated in Figure 2. The mean absolute percentage error (MAPE) was 2.6%, the root mean square error (RMSE) was 2.450, and the coefficient of determination (R^2) was 0.986.

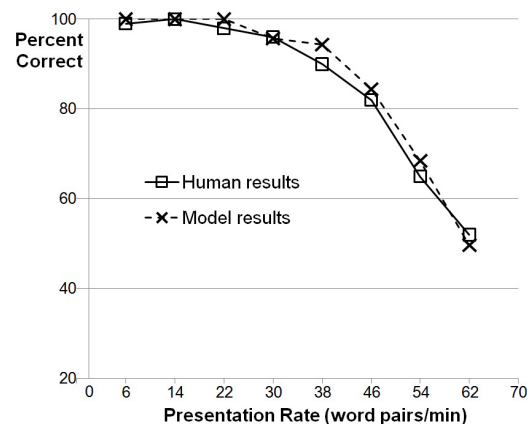


Figure 2. Comparing human and model percent correct results.

Examining the module activity traces revealed two types of errors. The first type is no response. Due to the random noise in chunk activation, the retrieval of a word-category relationship from the declarative memory may take a long time or even fail. Although the model will keep retrieving if it fails, it may not have enough time to complete the judgment and issue the manual key-press response within the duration of a trial. As the duration decreases with increasing presentation rates, no-response errors become more frequent. The second type is incorrect response. It happens when the model has issued a key-press motor command, but the next trial starts before the "finger" actually hits the "key". If the second trial's correct response is different from the first

trial's correct response, the key press will become an incorrect response to the second trial.

Linear regression was used to fit the relationship between model overall utilization (*OU*) and NASA-TLX overall workload (*OW*) ratings ($R^2 = 0.988$). The resulted regression function is

$$OW_{predicted} = -5.8 + 570.4 * OU. \quad (1)$$

Figure 3 shows the subject-reported NASA-TLX overall workload and corresponding model prediction under different task demand conditions.

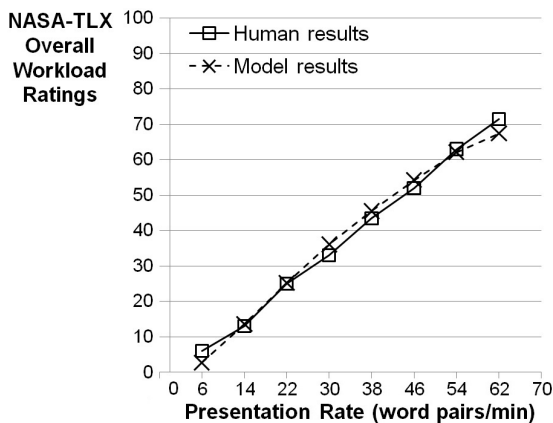


Figure 3. Comparing human and model overall workload results.

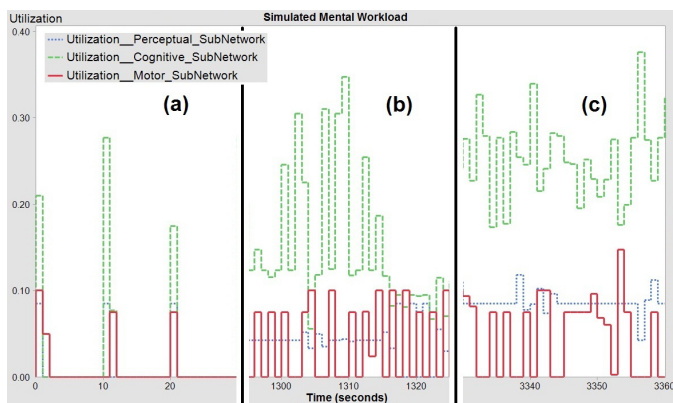


Figure 4. Visualization of mental workload components in ACTR-QN under different task demands. Word presentation rates: (a) 6 pairs/min, (b) 30 pairs/min, and (c) 62 pairs/min.

Although the authors (Colle & Reid, 1998) did not report subjective ratings for each mental workload component, ACTR-QN computed and visualized sub-network utilization values, which we assume to have linear relationship with corresponding workload components. Figure 4 shows the utilization of perceptual, cognitive, and motor sub-networks under three different presentation rates. The visualization clearly demonstrates workload increasing with faster presentation rates and provides more detailed estimation about each workload components.

DISCUSSION

As argued by Byrne and Pew (2009), a merit of modeling techniques is to quantitatively define concepts like workload, rather than qualitatively define them in a verbal form. This study proposed and validated a quantitative mental workload modeling theory and method in ACTR-QN, which is an integrated cognitive architecture combining the advantages of ACT-R and QN human performance modeling. Server utilization in QN was used to model mental workload. Modeling results were validated against human data from a semantic category judgment task. The same ACTR-QN model generated both percent correct performance results and NASA-TLX overall mental workload results that were very similar to the corresponding human data ($MAPE < 5.0\%$ and $R^2 > 0.95$). In addition, different components of mental workload were visualized in ACTR-QN. The visualization clearly showed the increase of mental workload with greater task demands.

ACT-R is perhaps the most influential cognitive architecture in the research field, but its application in human factors analysis has mainly focused on modeling performance. This study introduces a new theory and method to measure and visualize mental workload in ACTR-QN. The theory and method is in accord with three important features of mental workload: sensitive to task demands, multi-dimensional, and different from primary task performance. Since ACTR-QN uses exactly the same modeling syntaxes as ACT-R to define task-specific knowledge and parameters, as demonstrated by this study, researchers who are familiar with ACT-R can enjoy a seamless modeling of mental workload in ACTR-QN.

Compared with human-in-the-loop empirical methods, mental workload evaluation using predictive models such as ACTR-QN does not require any real time measurement of operator in-the-loop testing. This feature makes ACTR-QN especially suitable for early-stage evaluation without any full-scale system, quick evaluation of a large amount of design alternatives, and evaluation in high-risk domains. In this regard, task-based (e.g., Cassenti, Kelley, & Carlson, 2010) and task-time-based (e.g., Wang et al., 2010) methods usually need expert judgments to estimate workload scales and time constraints for unit cognitive tasks. In contrast, ACTR-QN's prediction of mental workload can complement or substitute expert subjective judgments, as it is based on a synthesis of a large amount of related findings and governed by the psychological and neuroscience theories underlying the cognitive architecture. These advantages make ACTR-QN a valuable one in the arsenal of system test and evaluation tools. Future research can further develop and examine

ACTR-QN's workload modeling function in more complex task scenarios.

Wu, C., & Liu, Y. (2008). Queuing network modeling of the psychological refractory period (PRP). *Psychological Review*, 115(4), 913-954.

REFERENCES

- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, 111(4), 1036-1060.
- Byrne, M. D., & Pew, R. W. (2009). A history and primer of human performance modeling. *Reviews of Human Factors and Ergonomics*, 5, 225-263.
- Cassenti, D. N., Kelley, T. D., & Carlson, R. A. (2010). Modeling the workload-performance relationship *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 54, 1684-1688.
- Colle, H. A., & Reid, G. B. (1998). Context effects in subjective mental workload ratings. *Human Factors*, 40(4), 591-600.
- Gray, W. D., Schoelles, M. J., & Sims, C. (2005). Cognitive metrics profiling. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 49, 1144-1148.
- Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 50, 904-908.
- Jo, S., Myung, R., & Yoon, D. (2010). A proposed methodology for the theory-based prediction of mental workload with a cognitive architecture. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 54, 1941-1945.
- Lebiere, C. (2001). A theory-based model of cognitive workload and its applications. *Proceedings of the 2001 Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)*, Arlington, VA.
- Liu, Y. (1997). Queueing network modeling of human performance of concurrent spatial and verbal tasks. *IEEE Transactions on Systems, Man, Cybernetics*, 27, 195-207.
- Liu, Y., Feyen, R., & Tsimhoni, O. (2006). Queueing Network-Model Human Processor (QN-MHP): A computational architecture for multitask performance in human-machine systems. *ACM Transactions on Computer-Human Interaction*, 13(1), 37-70.
- Mitchell, D. K., & Samms, C. (2009). Workload warriors: Lessons learned from a decade of mental workload prediction using human performance modeling. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 53, 819-823.
- Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review*, 115(1), 101-130.
- Wang, W., Cain, B., & Lu, X. L. (2010). Predicting operator mental workload using a time-based algorithm. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 54, 987-991.
- Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, 50(3), 449-455.
- Wu, C., & Liu, Y. (2007). Queueing Network modeling of driver workload and performance. *IEEE Transactions on Intelligent Transportation Systems*, 8(3), 528-537.