Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 3 (2015) 4266 – 4272

Procedia MANUFACTURING

6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015

Progress towards automated human factors evaluation

Shi Cao*

University of Waterloo, 200 University Avenue West, Waterloo, N2L 3G1, Canada

Abstract

Human factors tests are important components of systems design. Designers need to evaluate users' performance and workload while using a system and compare different design options to determine the optimal design choice. Currently, human factors evaluation and tests mainly rely on empirical user studies, which add a heavy cost to the design process. In addition, it is difficult to conduct comprehensive user tests at early design stages when no physical interfaces have been implemented. To address these issues, I develop computational human performance modeling techniques that can simulate users' interaction with machine systems. This method uses a general cognitive architecture to computationally represent human cognitive capabilities and constraints. Task-specific models can be built with the specifications of user knowledge, user strategies, and user group differences. The simulation results include performance measures such as task completion time and error rate as well as workload measures. Completed studies have modeled multitasking scenarios in a wide range of domains, including transportation, healthcare, and human-computer interaction. The success of these studies demonstrated the modeling capabilities of this method. Cognitive-architecture-based models are useful, but building a cognitive model itself can be difficult to learn and master. It usually requires at least medium-level programming skills to understand and use the language and syntaxes that specify the task. For example, to build a model that simulates a driving task, a modeler needs to build a driving simulation environment so that the model can interact with the simulated vehicle. In order to simply this process, I have conducted preliminary programming work that directly connects the mental model to existing task environment simulation programs. The model will be able to directly obtain perceptual information from the task program and send control commands to the task program. With cognitive model-based tools, designers will be able to see the model performing the tasks in real-time and obtain a report of the evaluation. Automated human factors evaluation methods have tremendous value to support systems design and evaluation.

Keywords: Systems design; Usability tests; Cognitive architecture; Human performance modeling; Mental workload; QN-ACTR

* Corresponding author. Tel.: +1- 519-888-4567 ext. 36377; fax: +1-519-746-4791.
E-mail address: shi.cao@uwaterloo.ca

doi:10.1016/j.promfg.2015.07.414

1. Introduction

Human factors tests are important components of systems design. In complex human-machine systems, operators are often required to perform multiple complex tasks simultaneously, and human performance is critical to the overall performance of the whole system. System designers need to evaluate operators' performance and workload while using a system and compare different design options to determine the optimal design choice. Such evaluation and analysis have become increasingly crucial in the design and quality control of complex systems in a wide range of industrial domains including healthcare, human-computer interaction, transportation, manufacturing, and aviation. As a result, system designers and policy makers have an increasing need for quantitative human factors evaluation methods.

Traditionally, human factors evaluation and tests mainly rely on empirical user studies. For example, to evaluate an in-vehicle navigation system, designers usually recruit drivers to interact with the system and record their performance during the interaction. Empirical studies, however, add a heavy cost to the design process; in addition, it is difficult to conduct comprehensive user tests at early design stages when no physical interfaces have been implemented. To address these issues, I develop computational human performance modeling techniques that can simulate operators' interaction with machine systems. During the simulation, the human model interacts with the machine model and generates performance and workload results without the need for human participants.

Human performance models can be categorized into cognitive models and physical models. My current research focuses on the cognitive aspects such as decision making and system operation control, with a goal to apply the models to human factors tests and evaluation. Cognitive psychologists also develop and use cognitive models to explain psychological phenomena and test theoretical hypotheses. My modeling approach is built upon previous modeling work in the cognitive science literature. In contrast, a different field of research is digital human modeling that focuses on the physical aspects such as anthropometry and biomechanics. These aspects are also important in systems design and evaluation, but they are currently not the focus of my modeling work.

In the development of computational human performance models, there are two major challenges. The first one is how to comprehensively cover the wide range of human capabilities. The human cognition is so versatile and adaptive that humans can master very complicated tasks and learn almost unlimited new skills. To model and explain human cognitive capabilities, researchers usually use a divide-and-conquer strategy, studying each aspect individually. As a result, there are many isolated models each accounting for one aspect of human cognition; however, human performance (e.g., reaction time) is the combining result of all the aspects such as perception, memory, and motor control. Therefore, an integrated model is needed to explain and predict human performance as a whole. This need for unified theories of cognition was advocated by Newell [1] and echoed by many other researchers. Along the same line of research, my approach is an integrated cognitive architecture. The second challenge is how to accurately reflect human limitations, such as limited memory capacities and mental processing speed. This challenge is often addressed by accumulating model validation work in this research field. Earlier work focusing on individual cognitive aspects aims to validate parameter values that reflect human limitations; once validated, later models could use the same parameter values in similar task scenarios. For example, the processing cycle time of the central cognitive processor has been established as 50 ms, and this value is commonly used by many models. Human performance modeling is still a young research subject. As the accumulation of research in this field, an ultimate goal is to achieve automated human factors evaluation in which a modeler can utilize model components and parameter values validated in previous work.

The work reviewed in this paper focuses on the recent development of QN-ACTR [2], an integrated cognitive architecture combining two isolated but complementary architectures, Queueing Network (QN) [3,4] and Adaptive Control of Thought-Rational (ACT-R) [5,6]. A cognitive architecture is a unified theory of cognition implemented as a computerized simulation program. As a theoretical framework, it unifies the underlying cognitive mechanisms of human performance. The modules or servers of a cognitive architecture and their computational mechanisms are biologically inspired; they are based on psychological and neurological evidence of how the human brain works. At the implementation level (using discrete event simulation), the human mind is represented as an information processing system. Mental representations and knowledge are programmed as symbols, and mental processing is programmed as computational functions applied to the symbols. The parameters of the architecture can represent human factors such as visual processing speed and working memory capacity. System inputs include the descriptions of the task and task-specific knowledge; system outputs include simulated performance such as the contents of responses, processing time, and correct rates as well as workload. The term “architecture” refers to the generic cognitive framework representing the general processing capabilities and mechanisms of the human mind. In contrast, the term “model” often refers to a set of task-specific information, including the tasks to be performed, the knowledge required to perform the tasks, and the parameter values. To simulate different tasks, modelers use the same cognitive architecture but need to specify a task-specific model for each task, although similar tasks may share the same model components.

The goal of integrating QN and ACT-R is to take the advantages of both and overcome the limitations of each method alone [7]. QN has advantages in modeling human multitasking performance and workload [8,9], whereas ACT-R has advantages in modeling complex cognitive activities such as learning and decision making [6,10]. QN-ACTR currently has two versions, one implemented in C# based Micro Saint® Sharp (http://www.maad.com) and the other implemented in Java. Both versions have been verified that they have implemented the advantages of both ACT-R and QN. Completed studies using QN-ACTR have modeled multitasking scenarios in a wide range of domains, including transportation, healthcare, and human-computer interaction. To validate a model, the general process includes the following steps. First, human empirical studies are conducted to collect human performance and workload data, often in computerized test environments. Second, models are programmed to interact with the same test environments, producing modeling results as the simulation runs. Finally, the modeling results are compared with the human results. If the results are very similar, the models can be regarded as representations providing a plausible account for the mechanisms of human performance and workload, and the models could be used to predict human performance and workload in similar task scenarios. The next section will present the completed studies in more details. The success of these studies demonstrated the modeling capabilities of QN-ACTR.

2. Review of completed work

2.1. Modeling driving and speech comprehension

Driving is a complex task requiring the coordination of perception, cognition, and motor activities. Modeling human driving performance and workload has values for both the development of modeling theories and the evaluation of in-vehicle interfaces. Cao and Liu conducted an empirical study that examined human performance and workload in a dual-task of lane keeping and speech comprehension [11], and a QN-ACTR model was built to simulate the human performance and workload [12]. In the lane keeping single-task, participants drove a car in a driving simulator and were instructed to maintain lane position at the center of the lane while the speed was fixed at a constant level. In the speech comprehension single-task, participants listened to pairs of sentences and were asked to judged whether they have the same meaning or not. The dual-task condition required participants to perform both tasks simultaneously. This empirical study found that the standard deviation of lane position (*SDLP*) was increased when the driving speed was faster. Concurrent speech comprehension had no significant effect on *SDLP*. However, the correct rate of comprehension was decreased in the dual-task condition compared with the speech comprehension single-task condition. In addition, workload was significantly higher in the dual-task condition compared with the two single-task conditions.

The QN-ACTR model combined the mechanisms used in previous ACT-R and QN models. The integration was necessary because it was tested that using ACT-R or QN alone could not successfully simulate the human performance as observed from the empirical study. Previous ACT-R studies have modeled driving single-tasks [13] and speech comprehension single-tasks [14]. The descriptions of task-specific knowledge and parameter values in the QN-ACTR model followed previous ACT-R models. QN mechanisms were implemented to schedule dual-task processing and produce overall workload estimations. A QN filtering discipline was implemented to coordinate limited mental resources between the two tasks. It was found that this QN mechanism was necessary to produce results similar to the human data. In contrast, a simple first-in-first-out queueing mechanism was inappropriate for the scheduling of the dual-task demands in this case.

The QN-ACTR model produced results similar to the human results. For the lane keeping single-task, the model's *SDLP* at the high speed was also larger than the value at the low speed as the human results. The mean absolute percent error (*MAPE*) was 12% in comparison to the human results. For the sentence comprehension single-task, the model's reaction time result had an absolute percent error (*APE*) of 7%, and the correct rate *APE* was 6%. The model's overall workload results were also similar to the human data, showing the highest workload in the dual-task condition and the lowest workload in the lane keeping single-task condition.

2.2. Modeling diagnostic decision making with concurrent tasks

In the healthcare domain, an important human factors issue is the workload of physicians and nurses (e.g., in emergency departments) who are frequently interrupted and distracted by multiple concurrent tasks. It is important to examine and model the effects of multitasking on diagnostic decision performance and workload. Cao and Liu conducted an empirical study examining these effects using an abstract diagnostic decision task [15]. The task was to diagnose the single true disease among eight candidates. Three diagnostic tests were provided, each of which could reveal a property of the true disease. This abstract diagnostic decision task was designed to allow the measurement of decision making speed, accuracy, and strategy (analytic vs. heuristic). The decision task was displayed visually. Through the auditory channel, there were two kinds of concurrent tasks that can be paired with the decision task to form two dual-task conditions. The simpler auditory task was to monitor a beeping sound and watch for a flat tone; the more complex task was to memorize and keep tracking the dynamic emergency levels of three simulated patients and answer questions when prompted. The human results showed that diagnostic decision performance was impaired by the concurrent memorization task. In contrast, the concurrent sound monitoring task did not affect diagnostic performance. Both types of concurrent tasks significantly increased workload. Diagnostic decision strategies were not significantly changed between the single- and dual-task conditions.

Two kinds of models were developed in QN-ACTR representing the two types of diagnostic strategies (analytic vs. heuristic) [16]. The difference between strategies were implemented as different production rules used by the models. The QN filtering discipline was implemented to resolve dual-task resource conflict at the mental module level. The modeling results were similar to the human results for both types of strategies. For the analytic strategy model, *MAPE* was 2%, and *root mean square error* (*RMSE*) was 0.2 s. For the heuristic strategy model, *MAPE* was 2%, and *RMSE* was 0.3 s. Models without the filtering discipline were also tested (a simple first-in-first-out queueing mechanism was used in this case); however, such models could not complete the decision-memorization dual-task condition, because the controlled processes required to maintain the working memory storage for the memorization task were interrupted by inappropriate task switching. This result demonstrated the value of the QN perspective in modeling multitasking performance.

2.3. Modeling target shooting and arithmetic computation

In a recent work, a QN-ACTR model was built to simulate human performance and workload in a dual-task of target shooting and arithmetic computation [17]. The empirical results were available from a previous study, which required soldiers to shoot targets on a shooting simulation test platform [18]. The target types included enemy (brown) and friendly (green); the duration of target appearance had two levels (high time stress 3 s and low time stress 5 s); the task types included a shooting-only single-task condition and a dual-task condition that had a concurrent auditory-verbal arithmetic computation task. The human results showed that performance measures, including friendly hit percent, target hit percent, and addition correctness, were better in the low time stress condition compared with the high time stress condition. Dual-task significantly increased overall workload.

The QN-ACTR model used overall expected utilization (OEU) as an index of workload. For each QN server, expected utilization is defined as the ratio of the server processing time required by task demands to the total task time available. OEU is the average expected utilization values calculated from all servers that have non-zero service time. When the required task cannot be fully completely within the total task time available, for example, in the case of high time stress, expected utilization can be obtained through the simulation of an ideal condition in which all tasks are given sufficient time.

The modeling results were similar to the human data in terms of both performance and workload measures. OEU was directly calculated while the model perform the tasks, without using any extra free parameter. These findings showed that the OEU concept in QN-ACTR is a sensitive index of workload especially in time-stressed task conditions.

2.4. Modeling transcription typing and reading comprehension

Transcription typing is a fundamental task of human-computer interaction and has been regarded as an important task to test and evaluation human performance models. There have been a considerable amount of empirical data about human typing performance accumulated in the literature. Skilled typists can achieve a typing speed about 60 words per minute. Previous work using QN models has successfully simulated most of the typing phenomena reported in the literature but cannot model the phenomena related to reading comprehension [19], because previous QN models did not have mechanisms to account for declarative memory and semantic processing. Incorporating the declarative memory mechanisms from ACT-R, QN-ACTR gains the advantages in modeling a wider range of transcription typing performance. Completed work using QN-ACTR has modeled 29 transcription typing phenomena, in particular, the phenomena involving the complex cognitive activities of reading comprehension and the phenomena involving concurrent tasks and skilled typing [2]. These results demonstrated the benefits of the integrated cognitive architecture, because it was difficult to model such phenomena in either QN or ACT-R alone.

2.5. Simplifying model building process

Cognitive-architecture-based models are powerful tools for cognitive engineering and systems design. However, learning how to build the models could be a difficult task itself. It usually requires at least medium-level programming skills to understand and use the language and syntaxes that specify the task. For example, to build a model that simulates a driving task, a modeler often needs to build a driving simulation environment so that the model can interact with the simulated vehicle. To achieve automated human factors evaluation, the model building process needs to be simplified.

In a preliminary work, I have established a protocol to directly connect user mental models in QN-ACTR (Java version) to existing task environment simulation programs via User Datagram Protocol (UDP) connections (Fig. 1). In particular, a driving simulator TORCS (torcs.sourceforge.net) that has been used in many previous driver behavior studies [21–23] was used to test the UDP connection method. In each simulation cycle (about 20 ms for TORCS), TORCS sends traffic information such as the road heading and the lead vehicle's headway distance to QN-ACTR. The information serves as perceptual inputs for the mental model. TORCS then waits for the model's response. QN-ACTR updates its perceptual inputs accordingly and then sends updated information of controllers (i.e., the position of the steering wheel and the pedals) to TORCS. A QN-ACTR model processes the perceptual information and issues motor commands according to its own internal clock. QN-ACTR's clock and TORCS's clock are synchronized each time when they exchange information. The preliminary work was successful. QN-ACTR driving models were able to control simulated cars in TORCS. The benefit is that human participants' performance and models' performance can be compared on exactly the same testing platform. Modelers can also see how a model drives the car as the simulation runs in real time. The simulation speed factor can be adjusted to allow faster or slower than real time simulation.

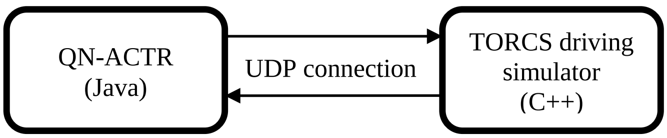

Fig. 1. A connection example between QN-ACTR and a driving simulation program (TORCS) via User Datagram Protocol (UDP).

Efforts have also been made to improve the usability of QN-ACTR as a cognitive engineering tool [20]. A Model Setup Assistant program was developed to facilitate and simplify the model building process. It uses a click-and-select user interface that allows users to setup a model by selecting from menu items and filling in blanks, avoiding the needs to memorize modeling syntaxes and parameter names. Users can define a model following natural language and experiment logic. Parameters are listed in a table with their meanings and default values displayed so that users can easily select a parameter and specify its value. Auto-check functions have also been developed to prevent typos.

3. Discussion and future research

The completed modeling work using QN-ACTR has demonstrated the advantages of the integrated cognitive architecture in modeling multitasking scenarios involving complex cognitive activities. The advantages of ACT-R and QN are combined to allow the simulation of tasks that are difficult to model by either ACT-R or QN alone. Unique QN features such as the filtering discipline and the server utilization concept have been demonstrated to be valuable in modeling complex cognitive multitasking performance and workload. Work has also been done to simplify the model building process. The completed work has moved forward to the long-term goal of automated human factors evaluation.

Future studies are needed to further accumulate task-specific models and determine the proper parameter values in a wider range of task scenarios. For example, previous desktop typing models can be extended and modified to simulate mobile touchscreen typing. Regarding the connection between cognitive models and machine models, an alternative method besides UDP connection is to use image recognition, which allows a mental model to visually perceive a machine/computer interface. In this case, modelers need to provide the images of key objects on the interface. With the help of image recognition techniques, cognitive models can directly perceive the visual objects. This method removes the need for coding the information path from the machine system interface to the cognitive model, and therefore it is suitable for the evaluation of software programs without their source codes available. Future research will test this technique.

In conclusion, cognitive-architecture-based modeling methods have the benefits of unifying isolated theories and explaining a wider range of human performance and workload in a variety of complex task scenarios. Completed work using QN-ACTR has demonstrated these benefits and made initial steps towards automated human factors evaluation. Model-based methods will enable early design tests and evaluation in simulated environments without the need for physical system prototypes or human participants. Such methods will have tremendous value to support systems design and evaluation. There is still a long road to go before reaching this ultimate goal, but it is well worth the effort.

References

[1] A. Newell, Unified Theories of Cognition, Harvard University Press, Cambridge, MA, 1990.

[2] S. Cao, Y. Liu, Queueing network-adaptive control of thought rational (QN-ACTR): An integrated cognitive architecture for modelling complex cognitive and multi-task performance, Int. J. Hum. Factors Model. Simul. 4 (2013) 63–86.

[3] Y. Liu, Queueing network modeling of elementary mental processes, Psychol. Rev. 103 (1996) 116–136.

[4] Y. Liu, R. Feyen, O. Tsimhoni, Queueing Network-Model Human Processor (QN-MHP): A computational architecture for multitask performance in human-machine systems, ACM Trans. Comput.-Hum. Interact. 13 (2006) 37–70.

[5] J.R. Anderson, C. Lebiere, The atomic components of thought, Erlbaum, Mahwah, NJ, 1998.

[6] J.R. Anderson, D. Bothell, M.D. Byrne, S. Douglass, C. Lebiere, Y. Qin, An integrated theory of the mind, Psychol. Rev. 111 (2004) 1036–1060.

[7] Y. Liu, QN-ACES: Integrating queueing network and ACT-R, CAPS, EPIC, and Soar architectures for multitask cognitive modeling, Int. J. Hum.-Comput. Interact. 25 (2009) 554–581.

[8] Y. Liu, Queueing network modeling of human performance of concurrent spatial and verbal tasks, IEEE Trans. Syst. Man Cybern. 27 (1997) 195–207.

[9] C. Wu, Y. Liu, Queuing Network modeling of driver workload and performance, IEEE Trans. Intell. Transp. Syst. 8 (2007) 528–537.

[10] F.J. Lee, J.R. Anderson, Does learning a complex task have to be complex?: A study in learning decomposition, Cognit. Psychol. 42 (2001) 267–316.

[11] S. Cao, Y. Liu, Concurrent processing of vehicle lane keeping and speech comprehension tasks, Accid. Anal. Prev. 59 (2013) 46–54.

[12] S. Cao, Y. Liu, Modeling driving and sentence comprehension dual-task performance in Queueing Network-ACTR, Proc. Hum. Factors Ergon. Soc. Annu. Meet. 58 (2014) 808–811.
[13] D.D. Salvucci, Modeling driver behavior in a cognitive architecture, Hum. Factors. 48 (2006) 362–380.
[14] J.R. Anderson, R. Budiu, L.M. Reder, A theory of sentence memory as part of a general theory of memory, J. Mem. Lang. 45 (2001) 337–367.
[15] S. Cao, Y. Liu, Effects of concurrent tasks on diagnostic decision making: An experimental investigation, IIE Trans. Healthc. Syst. Eng. 3 (2013) 254–262.
[16] S. Cao, Y. Liu, Queueing Network-ACTR modeling of concurrent tasks involving multiple controlled processes, Proc. Hum. Factors Ergon. Soc. Annu. Meet. 57 (2013) 768–772.
[17] S. Cao, Y. Liu, Modelling workload in cognitive and concurrent tasks with time stress using an integrated cognitive architecture, Int. J. Hum. Factors Model. Simul. (In Press).
[18] S.E. Kerick, L.E. Allender, Effects of cognitive workload on decision accuracy, shooting performance, and cortical activity of soldiers, U.S. Army Research Laboratory Human Research and Engineering Directorate Aberdeen Proving Ground, 2004. http://handle.dtic.mil/100.2/ADA433487.
[19] C. Wu, Y. Liu, Queuing Network modeling of transcription typing, ACM Trans. Comput.-Hum. Interact. 15 (2008) 6: 1–45.
[20] S. Cao, Y. Liu, An integrated cognitive architecture for cognitive engineering applications, Proc. Hum. Factors Ergon. Soc. Annu. Meet. 56 (2012) 323–327.
[21] S. Cao, Y. Qin, M. Shen, Modeling the effect of driving experience on lane keeping performance using ACT-R cognitive architecture, Chin. Sci. Bull. Chin. Version. 58 (2013) 2078–2086.
[22] S. Cao, Y. Qin, X. Jin, L. Zhao, M. Shen, Effect of driving experience on collision avoidance braking: An experimental investigation and computational modelling, Behav. Inf. Technol. 33 (2014) 929–940. doi:10.1080/0144929X.2014.902100.
[23] J. He, A. Chaparro, B. Nguyen, R.J. Burge, J. Crandall, B. Chaparro, et al., Texting while driving: Is speech-based text entry less risky than handheld text entry?, Accid. Anal. Prev. 72 (2014) 287–295. doi:10.1016/j.aap.2014.07.014.