Profile Before Optimizing: A Cognitive Metrics Approach to Workload Analysis

Wayne D. Gray **Michael J. Schoelles** **Christopher W. Myers**
Rensselaer Polytechnic Institute
Troy, New York USA
[grayw, schoem, myersc] @rpi.edu

ABSTRACT

The Intelligence Analyst (IA) community will soon be the designated users of many new software tools. In the multitasking world of the IA, any one tool cannot be permitted to greedily consume cognitive resources. This situation requires a new approach to usability assessment; one that profiles the moment-by-moment demands placed on embodied cognition by a given software tool during task performance. The approach we have taken relies on families of cognitive models that interleave cognition, perception, and action at the 1/3 to 3 sec timescale. This is the level of analysis where embodied cognition forms *interactive routines* that adapt to the cost-benefit structure of the software tool. Our proof-of-concept is a model that performs a task that the IAs find challenging. From the trace of the model, we derive a *cognitive metrics profile* that pinpoints dynamic changes in workload demands on human cognitive, perceptual, or action systems.

Author Keywords

cognitive modeling, intelligence analysts, embodied cognition, interactive behavior, cognitive architectures

ACM Classification Keywords

H.1.2 User/Machine Systems; H.5.2 User Interfaces

INTRODUCTION

We are involved in an effort that seeks to bridge the gap between basic research and application, bringing new technologies for searching massive data, hypothesis generation and test, and more to the desktop of the Intelligence Analyst (IA). Whereas other groups are attempting to assist the IA in her or his role as problem solver, problem finder, hypothesis generator, hypothesis evaluator, and so on; our role is to assist the IA as *user*.

This effort is somewhat unique in that a single user community, the IA, is the designated user of a score or more projects that will put new and innovative information technologies on their desktops. Given that these technologies will be used alone and in various unforeseeable combinations, a new approach to usability assessment is vital; one that provides a *cognitive metrics profile* for the dynamic changes in the demands that various software tools make on the IA's internal cognitive, perceptual, and action resources (i.e., embodied cognition) during the course of task performance. In the multitasking world of the IA, any one tool cannot be permitted to greedily hoard the IA's cognitive resources.

Although our users are Intelligence Analysts, their job demands match the profile of an increasing number of knowledge workers who daily shift through gigabytes of text, image, and sound files to answer or ask ill-defined questions. Hence, we believe that our developments and their discussion will be of interest to a broad swath of the CHI community.

In this short report we provide an overview of the problem as we have defined it, a description of our approach, a summary of our accomplishments to date, and discussion of our next steps. Although our report is full of citations to the literature on cognitive science and is concerned with cognitive modeling, evaluation, and empirical data collection ours is not primarily an empirical or theoretical report; rather, it is a high-level description of a new methodology and tool for usability analysis.

EMBODIED INTERACTIVE ROUTINES

Interactive behavior emerges out of a complex mix of *interactive objects* and *interactive devices* [7]. Interactive objects include text fields, text-based drop-down or pop-up menus, scroll-bars, 2-D and 3-D icons, static graphics with and without text, 2-D and 3-D animations, sound, and so on. The mouse, keyboard, and human eye continue to be the dominant interactive devices. At the same time older technologies such as handwriting and gestures, as well as newer technologies such as the force joystick, data glove, or eye gaze as a selection device are beginning to enjoy modest usage.

For any given set of interactive objects and interactive devices there is a finite set of interactive behaviors that produce visible (or invisible) changes in the state of the software tool. At the level of analysis required to model these interactive behaviors, human cognition is embodied cognition [3, 5] and is composed of combinations of cognitive, perceptual, and action operations that form sets

CHI 2005, April 2–7, 2005, Portland, Oregon, USA.
ACM 1-59593-002-7/05/0004.

of *interactive routines* (similar in concept to Ullman's visual routines [13]).

Modeling interactive routines requires describing the critical interactions between embodied cognition and the task environment at the 1/3 to 3 sec level of analysis. For our models we use the ACT-R architecture of cognition, in part, because the most recent version of ACT-R [v 5.0, see 1] supports the writing of models that interact directly with the same software environments that people use but, in large part, because of the distinction ACT-R affords between central control and various semi-independent functional modules.

In ACT-R central control is instantiated by production rules (condition–action pairs) with a cycle rate of 50 ms. Only one production rule can fire at a time. Productions initiate, interleave, and harvest the results from functional modules responsible for hand movements, eye movements, visual attention, memory retrieval, auditory attention, and more. Although space limits constrain more detailed discussion, multiple production firings and the initiation and execution of multiple functional processes are required for any given interactive routine. In an ACT-R model, interactive routines such as locating and moving visual attention to a given icon, or locating a target and moving the mouse cursor to its screen location, require approx 1/3 to 3 sec to complete; that is, the time to execute an interactive routine in ACT-R matches the timescale postulated for embodied cognition [3].

The same set of interactive routines can be applied to many different task environments and to many places within the same task environment. One interactive routine is selected over another based on its *expected utility*. The expected utility of an interactive routine changes as experience accumulates with its cost and success in the current task environment. Eventually the set of interactive routines brought to bear in a particular task environment represent an adaptation to the cost-benefit structure of that task environment. The central controller makes no functional distinction between knowledge in-the-head versus in-the-world or the means of acquiring that information (such as eye movement, mouse movement and click, or retrieval from memory [8]). (Note that adaptation does not imply *optimization*. Stable, suboptimal performance may result, as the adaptation process tends to be near-sighted in emphasizing immediate costs over delayed benefits [6].)

PROFILE BEFORE OPTIMIZING

It would be wonderful if we had a device that directly measured mental workload. As the user performed our tasks we could see what parts of the task stressed memory, visual attention, auditory attention, or the motor parts of the brain. Knowing, for example, that one subtask placed high demands on both visual attention and memory retrieval, we could work to redesign the task and interface to reduce the mental workload during performance.

Although it is not far-fetched to imagine that the next 10 yrs might bring brain-imaging technology to such a state, it is far-fetched to imagine that this would solve the user-testing problem. Whatever the technology, we would still need to bring users to the usability lab and this process would continue to consume much time and many resources. In addition, to the extent that we test new systems or newly redesigned systems, our human users would continue to be novices, not experts, during usability testing.

The charge of determining how task performance stresses human cognition is somewhat analogous to determining how to improve the performance of software. In that realm, the maxim that has emerged is *profile before optimizing*. Once code is running, "if you start haphazardly trying to optimize before you actually know where things are bogging down, you're guaranteed to be pessimizing your development efficiency" [2]. The accepted wisdom is to "profile the code to see where it's actually spending its time" and then to "focus on the few high-payoff areas and leave the rest alone."

We advocate applying this advice to interface design. Rather than trying to second guess how a design stresses visual attention or memory, interface designers should focus on designing consistent systems that are pleasurable to use, easy to learn, easy to recover from errors, and that meet the general performance requirements for which the system is being designed. Once a prototype is up and running it should be profiled. It is at this stage where modern computational cognitive models allow us to take a giant step forward.

COGNITIVE METRICS PROFILING

Although arguably possible in theory, in the usability lab it is impossible in practice to profile the dynamically changing demands that software places on human cognitive, perceptual, and action resources. However, modern architectures of cognition provide an alternative to profiling human users. To the extent that our models provide a high-fidelity representation of human embodied cognition then we can use our models as surrogates to measure the stresses that using an interface places on embodied cognition.

To be clear, we do not believe that it is practical to model all of human cognition that a user, for example, an IA, requires to perform their job. We do, however, believe that:

- Based on the cost-benefit structure of a given designed environment, we can predict which interactive routines will be adopted by users for task performance
- Based on our understanding of embodied cognition at the 1/3 to 3 sec level of analyses we can predict the stress that interactive routines will place on various cognitive, perceptual, and action systems

To some degree both of these claims are long-term research projects that form the focus of our basic research efforts. However, it is important to emphasize that both of these

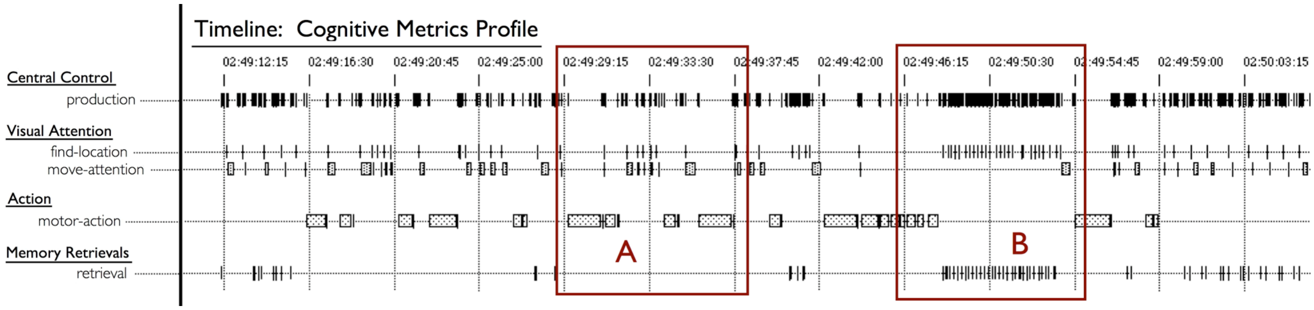

Figure 1: Cognitive Metrics Profile across the first 51 sec of the *Temple of the Sun* task. (Timeline shows time in hr, min, sec, and tics.) Each event is a separate entry whose duration is indicated by its length. All productions are 50-ms in duration. Hence, entries that appear longer than others indicate two or more production firings in close succession. Find-location events are instantaneous. The duration of move-attention events averages at 135-ms. Motor actions vary with Fitts Law for movement times, but times for mouse clicks and key clicks are constant. Finally, in this task, memory retrievals mostly involved the retrieval of the location of frequently used objects. Hence, most retrievals were under 50-ms.

claims are embedded in a tradition of research that extends back at least to Card, Moran, and Newell's classic work on the *Psychology of Human-Computer Interaction* [4]. For those interested in applications, but not in the minutia of theoretical disputes there has always been much practical value in applying current theory to current problems. We argue that the current state-of-the-art is such that although much about the embodied cognition of interactive routines remains to be worked out, that there is much value that can be obtained from applying current knowledge. Furthermore, we argue that advances in software engineering have increased the usability of computational cognitive modeling approaches well beyond the usability of older formalisms such as GOMS [4].

A Practical Example

We walk through one cognitive metrics profile at two levels of analyses. The profile was derived from an ACT-R model performing the Temple of the Sun scenario. This scenario was developed as part of the Sage project [9] to design a non-classified task environment in which to study IA performance.

In concept, the procedure for producing a cognitive metrics profile is simple. First, run one or more idiot savant models (more on this below) called *simBorgs* [10] on simple, but representative tasks using the same software interface as humans use. Unlike human users, the simBorgs produce a log file that provides a detailed trace of the activities of embodied cognition. Second, parse the log file into a form that can be imported into a program such as MacSHAPA [11] that can produce timelines. As in Figures 1 and 2, this literally yields the cognitive metrics profile. The profile is then inspected to determine time periods or subtasks that seem to have a higher than average workload.

In the figures, we have parsed embodied cognition into four major components. Central control is represented by production firings. Each production requires 50-ms to fire. Hence, the denser the space, the more productions fired per unit of time (in Figure 1, compare the top line of region A with region B). ACT-R incorporates a Treisman & Gelade [12] type theory of visual attention. This class of theories divides visual attention into a highly parallel and effortless feature detection phase (*find-location*) and a slower, serial *move-attention* phase. It can be seen that region A requires many fewer find-locations than region B, but entails many more shifts of visual attention (move-attention). In the figure we have collapsed different types of Actions (move mouse, move hand to/from mouse, mouse clicks, and key presses) into one motor-action operator. This allows us to determine that region A requires much more intensive motor activity than region B. Finally, memory retrievals are not much of an issue for region A but are much more important in region B.

Whereas Figure 1 zooms in on the initial 51-s of the task, Figure 2 zooms out to encompass the entire 6 min period. The impression that emerges from Figure 2 is one of a task in which memory demands vary greatly throughout task

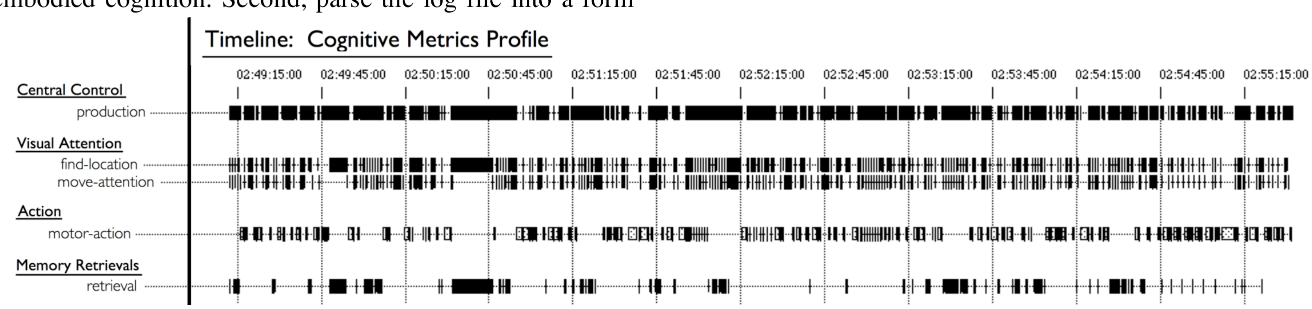

Figure 2: Cognitive Metrics Profile across the 6 min *Temple of the Sun* task.

performance. Demands on action are more constant than the demands made on memory. In contrast, the demands placed on visual attention appear to be both high and constant throughout. As the model is an expert at this task that has learned the location of many interactive objects, the designer might be concerned that the constant demands on visual attention are too high. Certainly, this would not be a good task to perform along with another task that also demanded the use of visual attention.

Cognitive metrics profiling is the missing component in interface design. It is a new component that does not eliminate the need for good software design and does not eliminate the need to attend to interface standards, guidelines, heuristics, walkthroughs, and other tools of the interface designer. With all else equal, it allows us to examine how the interactive routines demanded by the interface contribute to the workload imposed on the elements of embodied cognition.

DISCUSSION AND ISSUES IN COGNITIVE METRICS PROFILING

In this paper we have provided an overview of cognitive metrics profiling and a proof-of-concept. We have shown that architectures of cognition can provide an alternative to profiling human users namely cognitive metrics profiling through the use of cognitive models as high-fidelity representations of human embodied cognition. We raised many issues in passing that would have required many more pages to fully address. Some of these are issues for which we have ready answers; others are issues that demand the validation that can only come from extensive empirical testing.

We stated that we view our models as idiot savants in that they are experts at one or more aspects of embodied cognition, but do not incorporate extensive models of problem solving. This enables reuse of the models across a wide range of software minimizing the time spent rewriting the model. The design of our simBorgs [10] for implementing cognitive metrics profiling raises many issues for the field of cognitive modeling and architectures. Among them are: How should cognitive metric profiling be incorporated into a cognitive architecture such as ACT-R? How can interactive routines be incorporated into the architecture in order to facilitate reuse across models, domains, and even projects?

ACKNOWLEDGMENTS

This research has been funded in part by contract # MDA-904-03-C-0408 to Booz Allen Hamilton from the Advanced Research and Development Activity, Novel Intelligence from Massive Data Program.

REFERENCES

1. Anderson, J.R., Bothell, D., Byrne, M.D., Douglas, S., Lebiere, C. and Quin, Y. An integrated theory of the mind. *Psychological Review*, *111*, 4 (2004). 1036-1060.
2. Anonymous. Profile before optimizing. *Portland Pattern Repository*. 2004-12-11. <http://c2.com/cgi/wiki?ProfileBeforeOptimizing>
3. Ballard, D.H., Hayhoe, M.M., Pook, P.K. and Rao, R.P.N. Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*, 4 (1997). 723–742.
4. Card, S.K., Moran, T.P. and Newell, A. *The psychology of human-computer interaction*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1983.
5. Clark, A. Re-inventing ourselves: The plasticity of embodiment, sensing, and mind. *Journal of Philosophy and Medicine* (in press).
6. Fu, W.-T. and Gray, W.D. Resolving the paradox of the active user: Stable suboptimal performance in interactive tasks. *Cognitive Science*, *28*, 6 (2004). 901-935.
7. Gray, W.D. and Boehm-Davis, D.A. Milliseconds Matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior. *Journal of Experimental Psychology: Applied*, *6*, 4 (2000). 322–335.
8. Gray, W.D. and Fu, W.-T. Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science*, *28*, 3 (2004). 359-382.
9. Gray, W.D., Schoelles, M.J., Bringsjord, S.A., Burrows, K. and Colder, B. Sage: Five powerful ideas for studying and transforming the intelligence analyst's task environment *47th Annual Conference of the Human Factors and Ergonomics Society*, Human Factors and Ergonomics Society, Santa Monica, CA, 2003, 1019–1023.
10. Gray, W.D., Schoelles, M.J. and Veksler, V.D. Simborgs: Towards the building of simulated human users for interactive systems design. in *48th Annual Conference of the Human Factors and Ergonomics Society*, Human Factors and Ergonomics Society, Santa Monica, CA, 2004, 362-366.
11. Sanderson, P.M. MacSHAPA information and download. *University of Queensland*. 2004-12-08. <http://www.itee.uq.edu.au/cerg/macshapa.htm?
12. Treisman, A.M. and Gelade, G. A feature integration theory of attention. *Cognitive Psychology*, *12* (1980). 97–136.
13. Ullman, S. Visual Routines. *Cognition*, *18*, 1-3 (1984). 97-159.