COGNITIVE METRICS PROFILING

Wayne D. Gray, Michael J. Schoelles, & Chris Sims
Cognitive Science Department
Rensselaer Polytechnic Institute
[grayw; schoem, simsc] @rpi.edu

Cognitive Metrics Profiling promises a new approach to minimizing the cognitive workload of interactive systems. By metering high-fidelity computational cognitive models of embodied cognition, Cognitive Metrics Profiles provide a theory-based prediction of the transient changes in workload demanded by dynamic task environments. Although establishing the reliability and validity of this new approach will not be trivial, our profiles stand on the shoulders of the ACT-R architecture of cognition. More than 30-yrs of research have gone into the ACT line of theories. Over the last decade, hundreds of researchers have used ACT-R to build and test models of human cognition. Hence, although many of the details of the architecture are certainly incomplete, much of ACT-R is approximately correct. We expect that the predictions of a Cognitive Metrics Profile based on ACT-R will provide a better estimate of cognitive workload than the estimates used in current human factors practice.

Cognitive metrics profiling promises to be a theory-based method that serves five goals:

- Produces a profile of the transient changes in workload on the Operator’s cognitive, perceptual, and motor resources demanded by dynamic task environments
- Relates transient changes in workload to specific tasks or subtasks
- Pinpoints tasks or subtasks with excessive resource demands for interactive behavior
- Indicates places where task interleaving might occur without interfering with performance
- Suggests subtasks where tradeoffs might be made between functionally equivalent alternatives; for example, the use of knowledge in-the-world versus knowledge in-the-head

This short discussion is organized into four major sections. First, we first discuss the relevance of Cognitive Metrics Profiling for usability testing. Second, we provide an extremely short discussion of measures of workload that focuses on differences between Cognitive Metrics Profiling, the NASA TL/X (Hart & Staveland, 1988), and Multiple-Resource theory (MRT, Wickens, 2002). Third, we introduce the current form and format of Cognitive Metrics Profiling, and walk through a simple example. Fourth, provides a brief discussion of the distance that Cognitive Metrics Profiling has to go to meet the five goals introduced above.

BEYOND USABILITY

It is time to rethink interface design. As more and more new technologies come online, merely ensuring that each one has an easy to use, pleasurable, congruent, and fun interface is not enough. Given that these technologies will be used alone as well as 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 user’s internal cognitive, perceptual, and action resources (i.e., embodied cognition) during the course of task performance. In a multitasking world, any one tool cannot be permitted to greedily hoard the user’s cognitive resources.

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” (Anonymous, 2004). 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.

MEASURING COGNITIVE WORKLOAD

There are four major approaches to measuring cognitive workload: physiological, subjective, secondary task, and analytic. Cognitive metrics profiling represents a fifth approach; namely, theory-based.

Physiological measures (e.g., Wilson & Russell, 2003) attempt to meter physical processes/activities that are
presumed to reflect mental activity. Of course, obtaining physiological measurements can be awkward. This awkwardness has led to a search for non-intrusive physical measures of workload (e.g., Guhe et al., 2005). Of course, establishing the reliability and validity of physiological measures to general or specific mental processes is both difficult and demanding.

Subjective measures such as the NASA TLX (Hart & Staveland, 1988) are especially popular in the human factors community perhaps because they replace the awkwardness of physiological measurements with the simplicity of a paper-and-pencil (or computer) administered questionnaire. This allure is unfortunate, as it has long been known that subjective estimates of workload vary greatly with a person’s prior experience. For example, subjective assessments of task A vary depending on whether the person’s immediately prior experience was with task B or task C (Hancock, Williams, Manning, & Miyake, 1995). A second important problem is that, in contrast to physiological measures, subjective measures are administered after the task is performed. Retrospective measures must always be treated with caution. For example, retrospective estimates of pain (arguably a type of workload) during a medical procedure are not influenced by the duration of the procedure, but tend to reflect the intensity of pain during the worst part and the final part of the procedure (Redelmeier & Kahneeman, 1996).

Cognitive workload may be measured indirectly by the use of a secondary task. The amount of cognitive resources consumed by the primary task is assumed to be inversely correlated with secondary task performance. Although this procedure yields objective measures of secondary task performance, the interpretation of these measures can be obscure. For example, it is difficult to determine if secondary task performance reflects a cognitive capacity limitation or, more simply, a response bottleneck (Pashler, 1995). Furthermore, performance on the secondary task is complexly related to the nature of the primary task so that for different secondary tasks it would not be uncommon to find different orderings of performance (e.g., task A > task B, task B > task A, task A = task B) across different primary tasks. As far as we know, there are no established batteries of secondary tasks for assessing workload. A more fundamental criticism is that the secondary task approach never generalizes. The workload imposed by any new task cannot be predicted, but always requires a new empirical test.

Multiple-Resource Theory (MRT, Wickens, 2002) represents an analytic measure of workload. MRT captures the intuition that there are many pools of cognitive resources each with its own limits. Tasks that drain the resources of a single pool may be better performed if the task is reengineered to use multiple pools of resources.

MRT packages much cognitive science into a compact package that can be used in an applied setting by people who are not trained in cognitive science. Likewise, as it is not subjective, it is not prone to any of the problems that plague subjective techniques. Furthermore, although this has not been done, we believe that combining MRT with the Cognitive Walkthrough (Wharton, Rieman, Lewis, & Polson, 1994) would produce a tool that would force designers to think critically about the cognitive workload issues of their designs.

Although it has strengths, MRT also has limits. First, it provides a static, not dynamic, view of workload. Interleaving of cognitive resources occurs at the 1/3 of a second level of analysis. MRT does not capture behavior at that level of analysis. Rather, it seems to focus on behaviors with a duration of 3-30 sec (the "unit task" level of analysis, see Card, Moran, & Newell, 1983). Second, it seems better at identifying potential problem areas (hence our recommendation for combining it with the Cognitive Walkthrough) than at predicting specific problem instances (Wickens, Goh, Helleberg, Horrey, & Talleur, 2003). Third, the dimensions of the MRT are described fairly generally and the dimensions lack the unifying framework provided by architectures of cognition such as EPIC (Kiersas, Meyer, Ballas, & Lauber, 2000), Soar (Newell, 1990), and ACT-R (Anderson et al., 2004).

In contrast to older approaches to cognitive workload, Cognitive Metrics Profiling requires a computational cognitive model that can perform the task that is being profiled. The profiles we have performed have all used the ACT-R (Anderson et al., 2004) architecture of cognition.

To be clear, we do not believe that it is practical to model all of human cognition that a user, for example, an expert architect or an Air Force Commander, requires to perform his or her 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 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* (Card et al., 1983). 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 (Card et al., 1983).

**STEEPING THROUGH A COGNITIVE METRICS PROFILE**

We walk through one Cognitive Metrics Profile at two levels of analyses. The profile was derived from an ACT-R model performing the Blocks World task. For Blocks World, subjects are shown a random configuration of 8 colored blocks in a *target window* and are asked to reproduce this configuration in a *workspace window* using colored blocks available in a *resource window*. Blocks World is a simple task; one in which the cognitive, perceptual, and motor requirements can be easily understood. We use it here to avoid the long discussion that would be required to explain a more complex task to the reader.

In concept, the procedure for producing a Cognitive Metrics Profile is simple. First, run one or more idiot savant cognitive models called simBorgs (Gray, Schoelles, & Veksler, 2004) 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 (Sanderson, 2004) that can produce timelines. As in Figure 1 and Figure 2, this literally yields the Cognitive Metrics Profile. The profile is then inspected to determine periods or subtasks that seem to have a higher or lower than average workload.

In the figures, we have parsed embodied cognition into five major components. The control structure of the task – or task analysis – is represented at the top by two levels of goals and subgoals. Central control is represented by production firings; that is, the atomic operations of cognition. Each production requires 50-ms to fire. Hence, the denser the space, the more productions fired per unit of time. In the figures we have collapsed different types of Actions (move mouse, move hand to/from mouse, mouse clicks, and key presses) into one motor-action operator. ACT-R incorporates a Treisman & Gelade (1980) 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. ACT-R incorporates a well-developed theory of memory decay and activation. The last line in the profile shows the individual retrievals as well their latency. Note that at the scale shown in Figure 1, variations in the latency of memory retrievals are not apparent. These

![Figure 1: Cognitive Metrics Profile of trials 44 to 48 of the Blocks World task. (Timeline shows time in hr, min, sec, and tics.) Each event is a separate entry whose duration is indicated by its length. The four vertical gaps on the timeline (i.e., no activity in any of the timeline dimensions) represent the time between the end of one trial and the beginning of the next. (See text for a more detailed explanation.)](image-url)
variations are more apparent in Figure 2.

Whereas Figure 1 zooms out to encompass four trials of Blocks World over a 5-min period, Figure 2 zooms in to focus on one 60-sec trial. The impression that emerges from Figure 2 is one of a task in which memory demands vary as a function of the user’s current goal. The Encode Block goal seems to demand not only fewer memory retrievals, but retrievals of shorter latency than those for the Retrieve & Place goal.

We ran the output of the Cognitive Metrics Profile through several MacSHAPA queries to derive the latency for each memory retrieved for Encode Blocks and Retrieve & Place. As illustrated in Figure 3, this analysis shows that the Retrieve & Place goal (right) is much more memory intensive than the Encode Blocks goal (left).

**STATUS OF THE FIVE GOALS FOR COGNITIVE METRICS PROFILING**

The work presented in this paper demonstrates that we have achieved our first two goals. Figure 1 and Figure 2 profile dynamic changes in workload on cognition, perception, and action, while relating these changes to specific tasks or subtasks.

The third goal is to pinpoint tasks or subtasks with excessive resource demands and the fourth goal is to indicate places for interleaving. The figures suggest that, within anyone cognitive resource, our profiles met these two goals. For example, within motor-action there are clearly blank areas where more motor-actions might occur without interfering with existing motor-actions. Likewise, there are places where it is just as clear that additional actions cannot be made. However, a more stringent test for these goals would be to suggest places where different types of resources could be interleaved. As shown by the zoomed in Figure 2, it is already the case that visual attention is interleaved with memory retrieval. What we do not know is whether this interleaving represents an upper or lower limit on the amount of interleaving that the cognitive system can tolerate.

Finally, our simple example does not touch our fifth goal; namely, the example does not suggest tradeoffs that might be made between alternative ways of implementing the same task or subtask. What is required to make this point (and what would be a better test of the 3rd and 4th points) are Cognitive Metrics Profiles of two
tasks that might be performed concurrently.

SUMMARY & CONCLUSIONS
We believe that Cognitive Metrics Profiling is an idea whose time has come. As basic research has produced more powerful formalisms for modeling embodied cognition, it has reached the stage where engineering applications can now follow. We do not argue that Cognitive Metrics Profiling is complete or that we have established its reliability and validity. We do believe that the years of basic research underlying the various components of the ACT-R architecture means that the results of Cognitive Metrics Profiling should be taken seriously if with a healthy grain of skepticism. If a Cognitive Metrics Profile suggests that a subsystem of embodied cognition may be overloaded during task performance that would be a claim whose validity should be investigated before the system is fielded.

ACKNOWLEDGEMENTS
The research for the development of Cognitive Metrics Profiling has been supported by a grants from the Air Force Office of Scientific Research AFOSR #F49620-03-1-0143, as well as the Office of Naval Research ONR #N000140310046. An exploratory application of this approach has been supported 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


