Comparing Cognitive Effort in Spatial Learning of Text Entry Keyboards and ShapeWriters

Arindam Das, Wolfgang Stuerzlinger
Computer Science and Engineering, York University, Toronto, Canada
{arindam, wolfgang} @cse.yorku.ca

ABSTRACT
Stable structured layouts of buttons are a primary means of control for input in current graphical user interfaces. Such layouts are ubiquitous—from tiny iPhone screens to large kiosk screens in the malls—they are found everywhere. Yet, there has been relatively little theoretical account that compares the impact of cognitive effort on learning such stable layouts. In this paper, we demonstrate that prior empirical results on cognitive effort in learning stable layouts are theoretically predictable through the memory activation model of a cognitive architecture, ACT-R. We go beyond previous work by quantitatively comparing the level of cognitive effort in terms of a newly introduced parameter in the declarative memory model of ACT-R. We theoretically compare the cognitive effort of two different layouts of graphical buttons with respect to their label representativeness in the domains of traditional keyboard and ShapeWriter.

Author Keywords
Modeling; ACT-R; Text Entry; Keyboard; ShapeWriter; Cognitive Effort.

ACM Classification Keywords
H.5.2 [Information Interfaces and Presentation]: User Interfaces—Theory and methods.

INTRODUCTION
Graphical Buttons are one of the primary controls for inputting information in graphical user interfaces. Commensurate with their importance, there have been two research works which empirically studied the effect of cognitive effort on learning a stable structured layout, the earliest work being that of Ehret [1] and the other one by Cockburn, Kristensson, Alexander & Zhai [2]. Each of them compared the learning of alternative layouts that varied in terms of label representativeness of buttons—from visible labels requiring low cognitive effort to learn, to invisible labels requiring high cognitive effort to learn. Both these works observed that the difference in cognitive effort necessary to learn different layouts is manifested in terms of a difference in the rate of learning. Their observation also suggests that such a difference in the rate of learning across layouts is typically evident within the first two to three sessions of rehearsals.

AVI'12, May 21-25, 2012, Capri Island, Italy.

By the term "cognitive effort" we refer to memory operations carried out while executing a visuo-spatial task. Modeling endeavors to compare cognitive effort in interface learning are important. This is because empirical studies involving greater level of cognitive effort are very laborious, time consuming, and at times frustrating for the participants as suggested by Cockburn, Kristensson et al [2]. Modeling, in such situations, would support a quick comparison of the necessary effort to learn interfaces before a full scale user study is undertaken.

While the empirical studies have yielded strong results, only Ehret [1] tried to *model* the extent to which different levels of cognitive effort promote the learning of a layout. Ehret's approach used a rule-based simulation program that was crafted specific to the graphical layout used in his study. Our work provides an alternative look at Ehret's modeling effort [1]. Unlike Ehret's computer-based simulation, we take a mathematical approach to create a predictive model. Our model is relatively simpler in that it consists of only two equations, the Base-Level Activation equation and the Reaction Time equation, which are also used by the ACT-R declarative memory simulation. Keeping the model simple this way also helps us to reduce the number of free parameters to three (3) in our work.

To a large extent, our modeling endeavor is guided by the *soft constraints hypothesis* of Gray and associates [3, 4, 5]. The soft constraints hypothesis is a rational analysis approach which proposes that the mixture of perceptual-motor and cognitive resources allocated for interactive behavior is adjusted based on temporal cost-benefit tradeoffs, such that the least-effort path of executing the visuo-spatial task at hand, gets implicitly chosen. As perceptual-motor effort increases, users will normally choose the least-effort option of fewer perceptual-motor operations and more memory operations, even if the memory retrieval is imperfect. Conversely, as perceptual-motor effort decreases, users will normally choose the least-effort option of more perceptual-motor operations and fewer memory operations.

In stable structured layouts, information stays put; that is, items in the layout have a fixed position. Gray and Fu [3] points out that in such cases, the total effort expended in locating an item on the layout can be considered as:

$$\text{TOTAL EFFORT} = \text{PM}_{(\text{access})} + \text{MEMORY}_{(\text{storage + retrieval})}$$

Effort Equation

$\text{PM}_{(\text{access})}$ signifies the effort to access information at a location on the layout (Examples of “access” include an eye movement to an icon in a menu ribbon of Microsoft Word or moving the mouse and clicking on a key of a graphical keyboard) and $\text{MEMORY}_{(\text{storage + retrieval})}$ includes the effort of storing an item in memory as well as the effort of subsequently retrieving that item from memory.

The soft constraints hypothesis concludes that the tradeoff between selecting the perceptual-motor versus cognitive behavior minimizes the total effort (performance cost) measured in the currency of time. Motivated by this, we introduce a parameter in the Base-Level Activation equation of ACT-R as a coefficient of practice time, as shown later in the paper. We hypothesize that this new parameter acts as a reasonable surrogate for measuring the proportion of cognitive effort, i.e. memory operations. By virtue of being a proportionality constant for practice time it is also able to describe the rate of learning. We call this parameter *effort factor*.

The other two free parameters in this work are the latency factor parameter and latency exponent parameter [6] that belongs to the Reaction Time equation of ACT-R described later. The latency exponent parameter is left at a fixed value across all experimental conditions, while the latency factor parameter is varied.

We validate our model against the previous human data from two different domains; a traditional text entry keyboard and a shapewriter. The human data was collected for each domain in two different label conditions; keys with labels on them (*visible* label condition) and keys with no labels on them (*invisible* label condition). A measured human datum from a trial that we validate against, refers to the response time required in finding a pre-cued item on a layout and either clicking it with a mouse or gesturing it using pen strokes, depending on the domain. *The same amount of practice time was given to all conditions* per domain [2] (p. 1578).

In order to validate our model, we first set the values of the new free parameter, *effort factor*, for the two different label conditions, *visible* and *invisible*, of the keyboard. We adjust the value of the effort factor once for each condition such that the model best fits the human data of that condition. The same two values of the effort factor (one for visible and the other for invisible condition) are then used to fit the human data from the shapewriter in each of the label conditions. The newly introduced parameter, *effort factor*, helps us differentiate the rate of learning across visible versus invisible label conditions of a domain, keyboard or shapewriter, thereby indicating the differences in the level of cognitive effort required to learn interfaces that vary in the obscurity of labels.

Being guided by the theory of soft constraints hypothesis of Gray et al [3, 4, 5], our model concurs with the prior results that the item locations on a layout are learned faster when the least-effortful strategy available in the layout explicitly requires retrieval of location information, given *constant practice time* across all conditions of label representativeness.

ACT-R THEORY

The ACT-R cognitive theory [7, 10] describes a modular system that aims to replicate the human mind. The theory can be viewed as a framework of mathematical equations that models the neural computations in order to realize human dynamic behavior. We focus here on the two equations behind the ACT-R declarative memory system that are relevant to our objective.

ACT-R Base-Level Activation Equation

In ACT-R declarative memory, chunks, i.e. memory traces of items, have different levels of activation to reflect their past use: chunks that have been used recently or chunks that are used very often receive a high activation. This activation decays over time if the chunk is not used. The activation of a chunk controls both its probability of being retrieved and its speed of retrieval. In the case where there are multiple candidates for retrieval, the chunk with the highest activation has the highest probability of being retrieved. A retrieval threshold sets the minimum activation a chunk can have and still be retrieved successfully.

The equation describing the activation of a chunk (representing an item) is given by

$$A = \ln\left(\sum_{j=1}^{n}[kt_j]^{-0.5}\right)$$ **Base-Level Activation Equation**

where n is the number of practices of the item completed so far, t_j is the age of the j-th practice of the item, and the negative exponent -0.5 controls how quickly the activation decays. The factor k in the equation is the new parameter, *effort factor*, described earlier. Following the soft constraint hypothesis of Gray et al [3, 4, 5], k is introduced as a temporal cost factor that would help comparing different degrees of cognitive effort required to accomplish different levels of information access conditions, *when the total practice time is held constant across all conditions*. It is to be noted that Anderson [8] (p. 277) as well as Stewart and West [9] (p. 235) both had previously suggested the usage of a cost factor similar to k, albeit in a different form and context.

Overall, A is the strength of the memory trace of an item after n practices of that item. A *practice* of an item is said to occur whenever a trace of that item is presented to the declarative memory. Presentation may happen because of either recognition or recall of that item.

ACT-R Reaction Time Equation

The time required for the declarative memory to respond to a request (recognition or recall) for a chunk (representing an item) is given by the following equation:

$RT = I + F * exp(-f * A)$ **Reaction Time Equation**

where I is an intercept time reflecting the time cost of perceptual (visual) encoding and motor response [10] (p. 1043). F is the latency factor, and maps activation to time. f is the latency exponent that remains fixed across all experimental conditions. The purpose of parameters F and f is only to scale the time to retrieve a chunk from memory. The fixed time cost of a visual encoding is set at 185 ms and the time cost of a motor response is set according to the task specific behavior discussed later. This equation can be interpreted as the temporal analog of the Effort Equation, the terms I and $F*exp(-f*A)$ being the temporal surrogates of $PM_{(access)}$ and $MEMORY_{(storage + retrieval)}$ respectively.

We model two tasks. One task involves finding a pre-cued Webdings symbol on a traditional keyboard and then entering it by clicking on the appropriate key using a mouse. The other task involves entering a pre-cued English word in a ATOMIK shapewriter by sliding through all the letters in the word sequentially using a stylus. Guided by the work of Gray et al [5], we estimate the time cost of motor response to be 300 ms for the keyboard. Based on the work of Cao and Zhai [11], we estimate the time cost of motor response to be 903 ms for the ATOMIK shapewriter [12].

MODEL BASED COMPARISON

Two data sets of mean reaction time (trial time) over five sessions of practice, one for the *visible label* condition and the other for the *invisible label* condition, corresponding to each domain (keyboard or shapewriter), are used. A trial involves entering a pre-cued target in a text entry task. We leave the latency exponent parameter, f, fixed at 0.77 across all the conditions for every domain. We first fit our model to the human data of the keyboard by adjusting the value of the *effort factor* parameter, k, once for the visible label condition and then for the invisible condition. We keep the values of these two parameters fixed and thereafter utilize them to fit the human data of the shapewriter across the two conditions.

Comparison in Keyboard

Cockburn, Kristensson et al [2] (fig. 2, p. 1574) carried out an experiment that tests how well users learn the location of keys on a graphical keyboard. The aim was to identify how the rate of learning varied in the early phase of practice between the condition requiring more cognitive effort and the one requiring less.

The labels on the keyboard were iconic symbols chosen from the Microsoft Webdings font. Users were trained on two different layouts of the keyboard; one with keys having labels (*visible* layout) and the other with keys having no labels (*invisible* layout). There were 18 symbols displayed in a separate target-cueing region, with the next target item highlighted in green. Tasks involved selecting the pre-cued target on either layout using mouse. Figure 1 shows the observed data in solid lines. The observation strongly suggests that the difference in the rate of learning between the invisible versus the visible layout is substantial in the first two sessions itself, the rate of decrease in reaction time being much higher for the invisible compared to the visible.

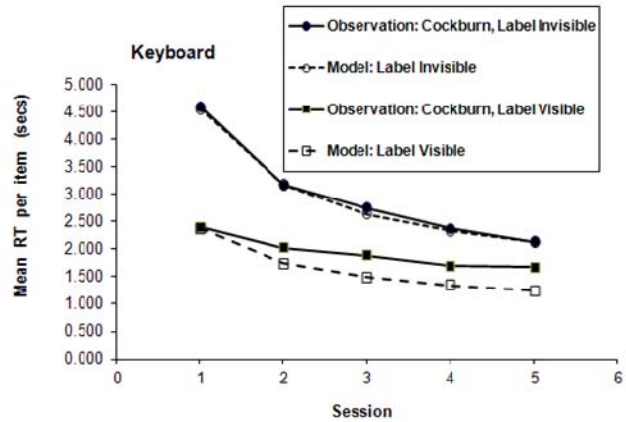

Figure 1. Mean Reaction Time, RT (in secs) per item (symbol) selected on the graphical keyboard, as observed in Cockburn, Kristensson *et al* [2] (fig. 2, p. 1574). The solid lines are the observed data and the dotted lines are the model data. The effort factor, *k* is adjusted to 1 for the *invisible* label condition and to 400 for the *visible* label condition.

In order to facilitate our data fitting process, we had to make a few assumptions, as the corresponding information was not given explicitly in their paper. From the observed data, we estimate that each session got completed in roughly 30 seconds on average. We further assume the inter-session periods to be constant. Also, except for the target-precue, we assume that environmental context cuing is minimal and can be ignored for our purposes. Figure 1 shows our model data against the observed data. With $r^2 = 0.998$, RMSE = 0.058 for invisible layout and $r^2 = 0.984$, RMSE = 0.332 for visible layout, our model closely fits to the observed data. In order to fit the data for the keyboard, the effort factor k was adjusted to 1 for the *invisible* condition and, to 400 for the *visible* condition. Next, we use these same values of the effort factor to estimate the model data for the invisible versus visible condition of shapewriter.

Comparison in ShapeWriter

Cockburn, Kristensson et al [2] (fig. 4, p. 1577) carried out an experiment that tests how well users learn the location of keys on a ShapeWriter hosting the virtual ATOMIK keyboard [12]. Shape writing is a text entry technique where the user strokes through all the letters in the word on a graphical keyboard without lifting his or her finger. The graphical keyboard in this case is called ShapeWriter. Shape writing is thus a task that involves both location and trajectory learning. The aim of this study was again to find out how the rate of learning varied in the early phase of practice between the condition requiring more cognitive effort versus the condition requiring less cognitive effort.

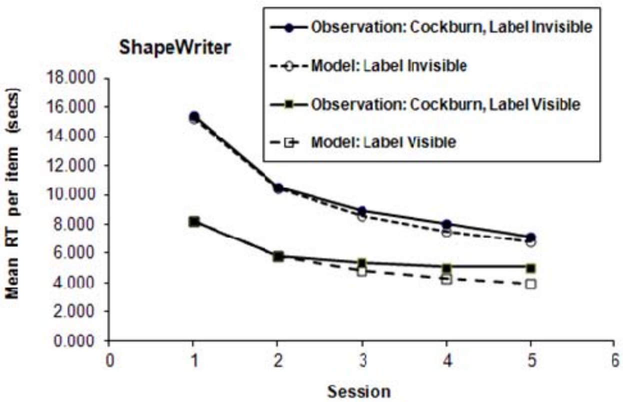

Figure 2. Mean Reaction Time, RT (in secs) per item (word) entered on the ShapeWriter, as observed in Cockburn, Kristensson et al [2] (fig. 4, p. 1577). The solid lines are the observed data and the dotted lines are the model data. Previously fixed values of the effort factor *k*=1 for *invisible* label condition and *k*=400 for *visible* label condition are used.

The labels on the keys to be stroked were English alphabet. Users were trained on two different layouts of the ShapeWriter; one with keys having labels (*visible* layout) and the other with keys having no labels (*invisible* layout). Tasks involved repeatedly writing 16 English words, each pre-cued randomly, on either layout, using a stylus.

The values of the effort factor *k* that were already fixed earlier for the traditional keyboard are utilized here. In order to fit human data, we had to make a few assumptions similar to the keyboard study, due to lack of information. From the empirical measures, we estimate that each session got completed in roughly 81.8 seconds on average. Figure 2 demonstrates the curvilinear trend of the model data. It correlates reasonably well with the observed data (r^2 = 0.998, RMSE = 0.226 for *invisible* layout and r^2 = 0.964, RMSE = 0.319 for *visible* layout).

CONCLUSIONS

With total practice time held constant across all conditions of label representativeness in a domain (traditional keyboard or shapewriter), our model data suggest that the newly introduced *effort factor* parameter, *k*, of the ACT-R Base-level activation equation can assist us in comparing the different levels of cognitive effort spent in learning layouts of different levels of label representativeness in the given domain. A lower value of *k*=1 refers to a certain degree of label representativeness of a layout whose locations are learned more quickly when the least-effortful strategy available requires more memory operations and less perceptual-motor operations, whereas a higher value of *k*=400 implies less memory operations but more perceptual-motor operations, instead.

It is worth repeating that our extended ACT–R model was fit to the empirical measures of just one stable layout (the traditional keyboard), while it closely matched the human performance of a second stable layout (the shapewriter) using the same parameter settings of *k* and *f*. The values for the latency factor parameter *F* were different though: *F* was higher for the invisible condition, corresponding to the higher latencies associated with memory retrievals (failures as well as successes) in the early phase of practice. Such difference in *F* values, therefore, does not seem to violate the soft constraints hypothesis perspective of Gray et al [3, 4,5], and hence we consider this an acceptable compromise.

ACKNOWLEDGEMENTS

We thank Professor Wayne D. Gray of RPI, New York for his invaluable suggestions on the final version of this paper. We thank York University, Toronto for funding this project.

REFERENCES

1. Ehret, B. (2002). Learning Where to Look: Location Learning in Graphical User Interfaces, CHI 2002, 211-218.
2. Cockburn, A., Kristensson, P. O., Alexander, J., and Zhai, S. (2007). Hard Lessons: Effort-Inducing Interfaces Benefit Spatial Learning, CHI 2007, 1571-1580.
3. Gray, W. D., & Fu, W.-T. (2001). Ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head: Implications of rational analysis for interface design. CHI Letters, 3(1), 112–119.
4. Gray, W. D., & Fu, W.-T. (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. Cognitive Science, 28(3), 359–382.
5. Gray, W. D., Sims, C. R., Fu, W.-T., & Schoelles, M. J. (2006). The soft-constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. Psychological Review, 113, 461–482.
6. Bothell, D. (2005). ACT-R 6.0 Reference Manual. March 2012, from act-r.psy.cmu.edu/actr6/reference-manual.pdf
7. Anderson, J. R. & Lebiere, C. (1998). The atomic components of thought. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
8. Anderson, J. R. (1983). A Spreading Activation Theory of Memory , Journal of Verbal Learning and Verbal Behavior, 22, pp. 261-295.
9. Stewart, T. C. and West, R. L. (2007). Deconstructing and reconstructing ACT-R: Exploring the architectural space. Cognitive Systems Research, 8, 227–236.
10. Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., and Quin, Y. (2004). An integrated theory of the mind. Psychological Review, 111(4), 1036–1060.
11. Cao, X., & Zhai, S. (2007). Modeling human performance of pen stroke gestures. CHI 2007, 1495–1504.
12. Kristensson, P.-O. and Zhai, S. (2004). SHARK2: a large vocabulary shorthand writing system for pen-based computers. UIST 2004, 43-52.