Eye Tracking the Visual Search of Click-Down Menus

Michael D. Byrne, John R. Anderson, Scott Douglass, Michael Matessa
Psychology Department
Carnegie Mellon University
Pittsburgh, PA 15213
byrne@acm.org, ja+@cmu.edu, sd3n+@cmu.edu, mm4b+@cmu.edu

ABSTRACT

Click-down (or pull-down) menus have long been a key component of graphical user interfaces, yet we know surprisingly little about how users actually interact with such menus. Nilsen's [8] study on menu selection has led to the development of a number of models of how users perform the task [6, 2]. However, the validity of these models has not been empirically assessed with respect to eye movements (though [1] presents some interesting data that bear on these models). The present study is an attempt to provide data that can help refine our understanding of how users interact with such menus.

Keywords

Menu selection, eye tracking, visual search, cognitive models

INTRODUCTION

Menus of one form or another have been a central feature of the user interface for some time (see [9] for a review). Mouse-based pull-down (requiring that the mouse button be held down) and click-down (which stay open once clicked until another click occurs) menus are more recent advances that have become ubiquitous in the modern graphical user interface. Recently, there has been increased effort devoted to understanding how users interact with click-down or pull-down menus. This research has included computational cognitive modeling [6, 2] and eye tracking research [1, 4]. Detailed approaches have the potential to provide understand interaction with menus at a fine grain.

Nilsen [8, Experiment 2] performed an experiment which provided detailed enough data to constrain computational cognitive modeling. In this experiment, users were presented with a single digit on the screen and a menu button. They clicked the button, and then searched for that digit in a menu of randomly-ordered digits that appeared as a result of the button click. Nilsen's data (shown in Figure 1) shows that users' response time is an approximately linear function of serial position in the menu, with each successive position being approximately 100 ms slower than the last. The exception is serial position 1, or the first menu position. Time for this position is slightly higher than response time for position 2.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
CHI '99 Pittsburgh PA USA
Copyright ACM 1999 0-201-48559-1/99/05...$5.00

The data suggest that Fitts' Law (see [5], chapter 2, Welford form), while an excellent predictor of mouse movement time, is not a good characterization of the menu search process. Users took much longer and had steeper slope as a function of target position than would be predicted by Fitts' law. Thus, it was argued by Nilsen and others that the bulk of the users time spend on this task is time for visual search.

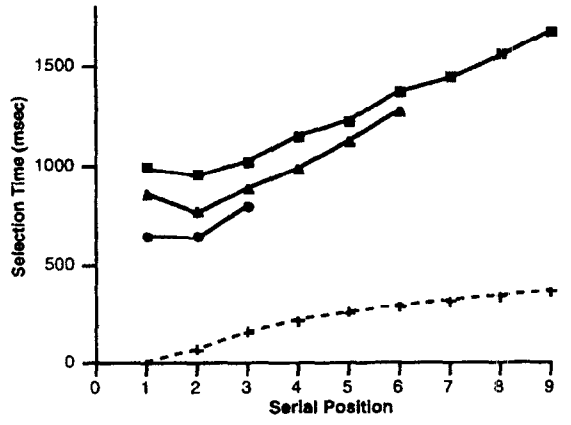


Figure 1. Nilsen's results: Response time as a function of menu size and target position. Lower dotted line is Fitts' Law predicted time[1]

Computational cognitive models that reproduce these results have been produced in both ACT-R [2] and EPIC [6], which are production-system theories that can be used to predict latency, accuracy, and ease of learning for a wide variety of HCI-related tasks. These models are similar in that they both produce response times that approximate the observed data. However, the models differ in the details of eye movement and mouse movement.

[1] Serial position 1 required a one-pixel mouse movement, so the Fitts' Law time is very close to zero.

The EPIC model[2] predicts the following:

1) Eye movement patterns should conform to a pattern that consists of 50% sequential top-to-bottom searching, and 50% randomly-ordered searching;

2) In cases of serial top-to-bottom search, the users' eyes should move down the menu a constant distance in each saccade, which is exhaustive in that every item of the menu from item 1 to the target item is examined;

3) The eye should "overshoot" the target item by one saccade with some regularity, since users are examining multiple items in parallel; and

4) No mouse movement occurs until the target has been located.

In contrast, the ACT-R model predicts:

1) Eye movements should be exclusively top-to-bottom;

2) The distance moved on each saccade should vary from trial to trial and menu to menu—items which do not share features with the target will be skipped over and thus the not every item is searched;

3) The eye never overshoots the target item; and

4) Mouse movements trail the saccades in a regular fashion but occur before the target has been located.

So, while these models cannot be differentiated by response time data, it should be possible to assess the validity of each model through careful eye and mouse tracking. While [1] attempted to shed more light on the validity of the EPIC model through an eye-tracking study, their experiment had several problems with respect to distinguishing among the current models. Their menus consisted of non-random arrangements of words where the exact target was sometimes unknown to participants, thus introducing reading, comprehension, memory for location, and the like. While this study is perhaps in some sense more ecologically valid, it is not a good evaluation of either the EPIC or ACT-R models and does not clearly relate to the Nilsen data.

EXPERIMENT

Procedures

The tasks used was essentially the same as the one used in the Nilsen experiment and a subsequent follow-up experiment [2]. Users were first shown a screen containing a rectangle with the word "Target:" followed by a target character. When the user clicked on this rectangle, a menu of characters appeared (see Figure 2). Users then searched for the target item in the menu and clicked on it. Visual point-of-regard (POR) and mouse position were tracked throughout the entire trial, and response time and accuracy were recorded for each trial..

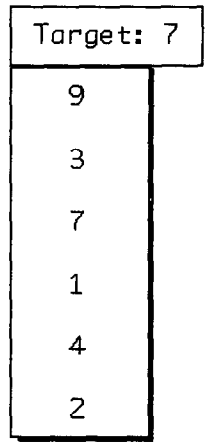


Figure 2. Example menu used

Participants

11 undergraduate participants were paid for their participation in the study. They were paid a base amount for simple participation and additional "bonus" money which was determined by their performance. All participants had normal uncorrected vision and were familiar with the use of computer menus.

Design

There were two primary within-subjects factors in the experimental design: menu length and target location. Three menu lengths were used: 6, 9, and 12 items. We decided to use longer menus than those used in the original Nilsen experiment because pilot data showed a general lack of interesting eye movements for 3-item menus. All target locations were used for each menu length.

The were other within-subjects factors in the design as well: target type and distractor type. Targets could be either letters or digits, as could non-target distractors. Thus, there were a total of 108 trials in the experiment: 6 6-item menu trials (one for each target location) + 9 9-item menu trials + 12 12-item menu trials X 2 target types X 2 distractor types. The 108 trials were randomly ordered by the experimental software. Participants also received 36 practice trials with randomly-chosen values on all factors. There was also a between-subjects manipulation. In one condition, the "Target" field remained on the screen when the menu appeared (as in Figure 2) and in the other, the "Target" button disappeared when it was clicked. The effects of the target type, distractor type, and presence of the target button are beyond the scope of the current presentation and will not be considered further.

[2] [6] presents several models. "The EPIC model" referred to here is the final model presented, the "Parallel Processing Dual Strategy Varying Distance Hybrid Model."

Apparatus/Materials

The eye tracker used was an ISCAN RK726/RK520 HighRes Pupil/CR tracker with a Polhemus FASTRACK head tracker. Head-mounted optics and a sampling rate of 120 Hz were used in this experiment. This system, like most other laboratory eye trackers, works by shining an infrared light on the eye and taking a video image of the eye. From that image, it is possible to determine the pupil center and the point on the cornea closest to the camera (the corneal reflection) and take the vector between them. This vector changes as the eye orients to different positions on the screen and with calibration to known points, it is possible to compute visual POR. The magnetic polhemus is used to compensate for head movements. POR reports by the eye-tracking equipment are typically accurate to within one degree of visual angle.

POR and mouse position were recorded approximately every 8 ms by the experimental software. Stimulus and POR/mouse data for each trial were recorded so that all individual trials could be "replayed" at various speeds. An experimenter monitored each experimental trial and recalibrated the eye tracker if there appeared to be sizable disparity between reasonable expectations about where users would look (in particular, users needed to look at the target on each trial) and the position reported by the tracker.

Users were seated approximately 30 inches from a 72 ppi computer display. Characters were 13 pixels high (approximately 0.34° of visual angle) with 26 pixels (approximately 0.69° of visual angle) separating characters. Thus, simultaneously foveating three characters would require a fovea of approximately 2.4° visual angle in diameter. (EPIC assumes that the fovea covers 2° of visual angle and characters must be foveated to be recognized.)

RESULTS

Analysis Technique

Sampling at 120 Hz, despite short trials, generates a great deal of raw data over 108 trials. However, from this raw data it is possible to compute where and when fixations have occurred. This can be done either by assuming that any eye position within a given region for more than some threshold number of milliseconds is a fixation (dwell-based) or assuming that any period of time showing relatively low velocity is a fixation (velocity-based). For the current data set, both methods were initially used and both methods yield approximately the same result. Because the velocity based method yields slightly less noisy data, the results presented here are based on that method of post processing. For each trial, the location of each fixation (with location 1 being the top item in the menu) was recorded. Mouse data were treated similarly; that is, post-processing analysis was used to identify the number and location of mouse "fixations" for each trial.

Response Time

Results for response time are presented in Figure 3. Clearly, response time is a function of target location, with higher locations generating longer response time. This is consistent with the Nilsen data. However, other aspects of Nilsen's data set were not reproduced as clearly. First, the slope of the function for the two larger menu sizes is somewhat shallower, around 75 ms (as opposed to 103 observed by Nilsen) and is even shallower for 6-item menus. Further, there appears to be very little main effect of menu size (controlling for position), as opposed to what Nilsen found. This may be a function of the larger spacing between items used here, which was necessary to make it possible to discriminate fixations on adjacent items. A second distinct possibility is that this is a practice effect; Nilsen's subjects had many more trials (1440) than our participants. Error rates were negligible in all conditions and will not be discussed.

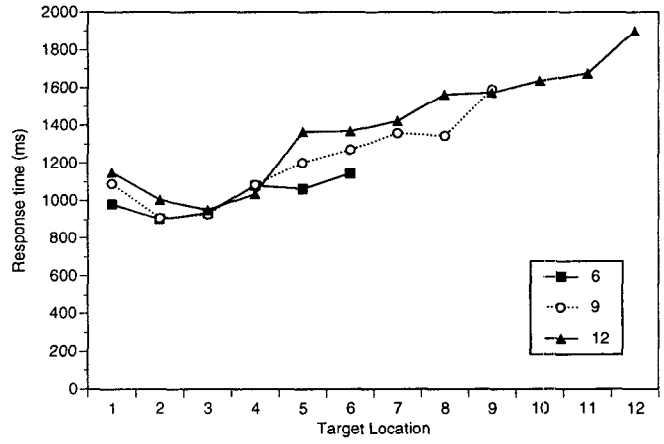


Figure 3. Response time by menu size and target location

Eye fixations

Random search models predict that the number of fixations should not be a function of target position (half of the time for EPIC). Ordered and exhaustive models (the other half of EPIC's trials) predict a strictly-increasing step function in the number of fixations required—for example, if the search takes in three items per fixation, targets locations 1-3 should require 1 fixation, targets at locations 4-6 two fixations, and so on. Ordered selective models (ACT-R) also predict a shallow and graded increase in the number of fixations with target location. Results are presented in Figure 4.

For six-item menus, the number of fixations is relatively insensitive to target location. For both the longer menus, the best-fitting regressions have an intercept of approximately 2.5 fixations and a slope of just under 0.2 fixations per serial position—thus, there is evidence that locations further down the menu do indeed require somewhat more fixations.

This is consistent with both the EPIC mixed top-down/random model and ACT-R's top-down feature search, which is encouraging for both models.

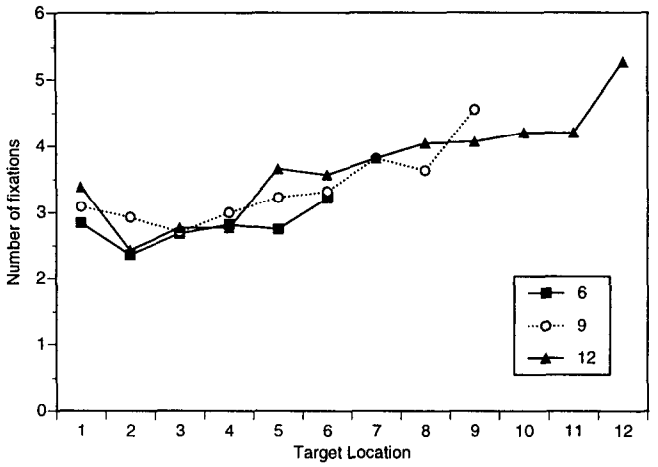


Figure 4. Number of fixations vs. target location by menu size

However, there is considerably more to the story than the raw number of fixations—in particular, the locations of those fixations is quite revealing. If the EPIC model is correct, then on 50% of the trials, every menu item has an equal probability of being foveated in the initial fixation, with the remaining 50% of the trials fixating somewhere in the first two (or perhaps three) items. Thus, the initial fixation would be to one of items 4, 5, or 6 approximately 25% of the time. (Fully random would predict 16.7% for each item; half random predicts 8.3% per item, summed across three items.) It would similarly predict the initial fixation be to items 4 or higher 33% of the time in a 9-item menu and 38% of the time in a 12-item menu. Figure 5 shows that this is clearly not the case.

Users clearly have a non-random preference for the first three menu items with their initial fixation, and a particularly strong preference for the first item. Clearly, the total number of initial fixations on items 4 and higher in the menus is less than what the EPIC model predicts. This is also inconsistent with the ACT-R model, which predicts that users will fixate on the first item in the menu which has a feature in common with the target item. While it is likely that a character in the first three items meets this criterion, the ACT-R model under-predicts the preference shown for the initial item.

Examining only the initial fixation does not, however, provide a complete characterization of the overall search process. Another way of looking at the fixation data is to consider the number of times each location on the menu is visited, on average, as a function of the serial position and the location of the target. These data are presented for 9-item menus only (due to space considerations—the patterns for 6- and 12-item menus are quite similar) in Figure 6.

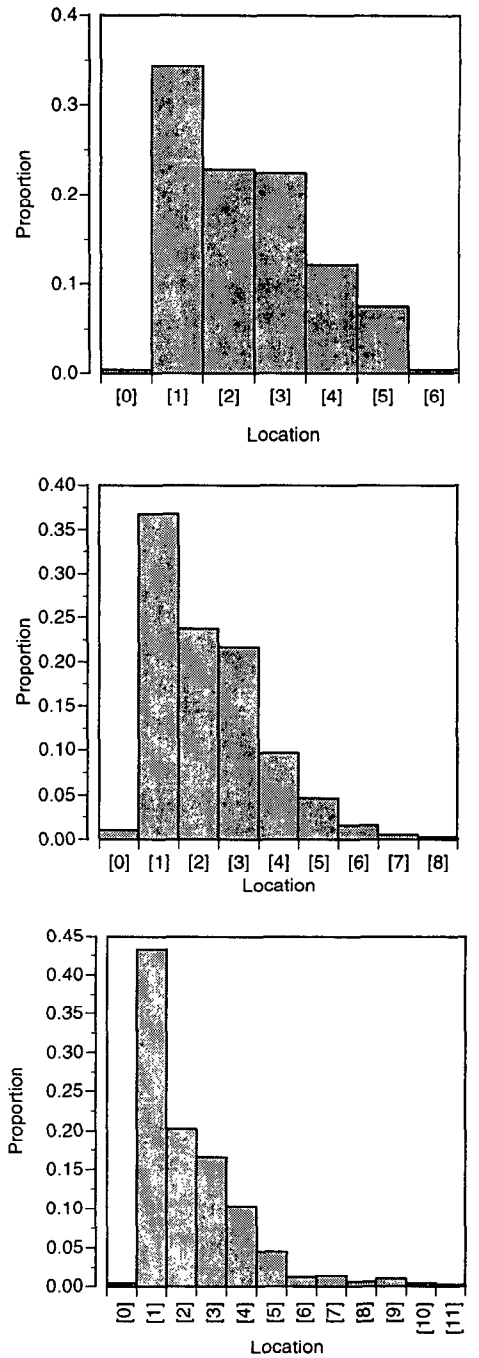
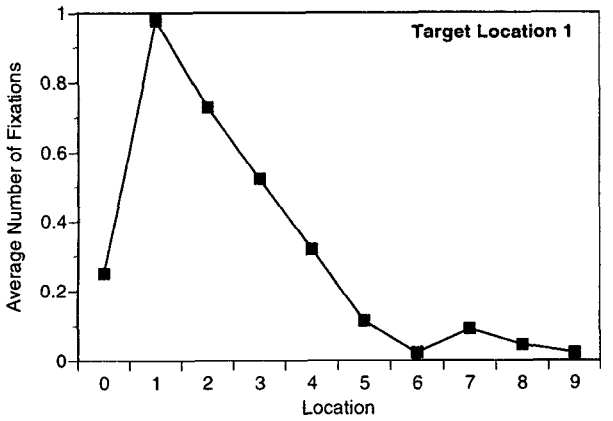
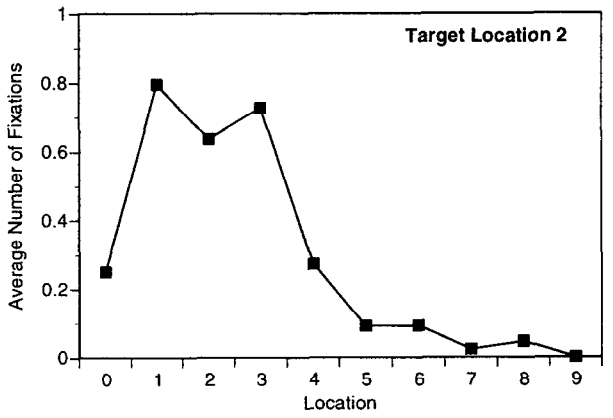
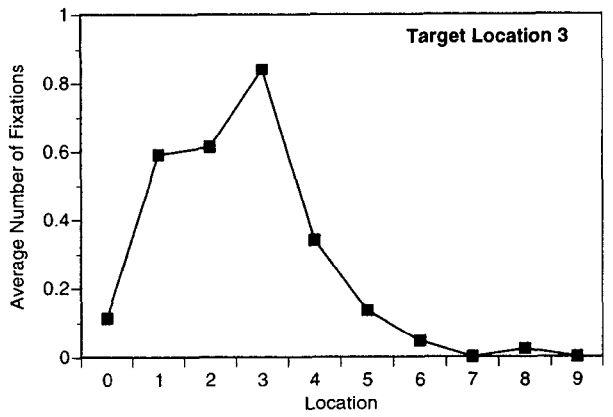


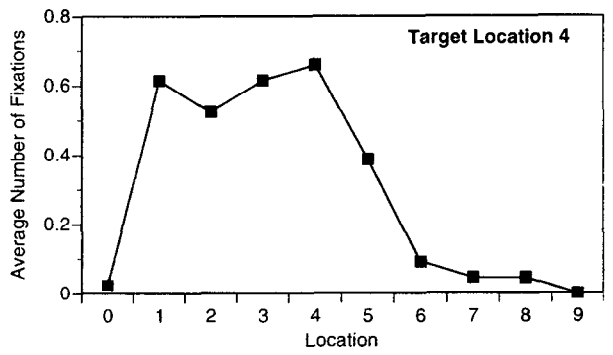
Figure 5. Histogram of initial fixation locations for 6-item, 9 item, and 12-item menus. (Items 9 and 12 are omitted from the 9-item and 12-item graphs because they are both exactly zero. Item 0 is the "Target: X" item.)

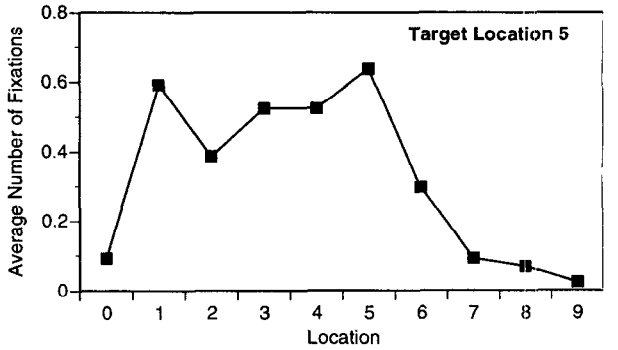
These data are especially revealing with respect to the status of random search. A random search would, for example, in a 9-item menu, still occasionally visit items 7-9 of the menu when the target appeared in items 1-3. However, the

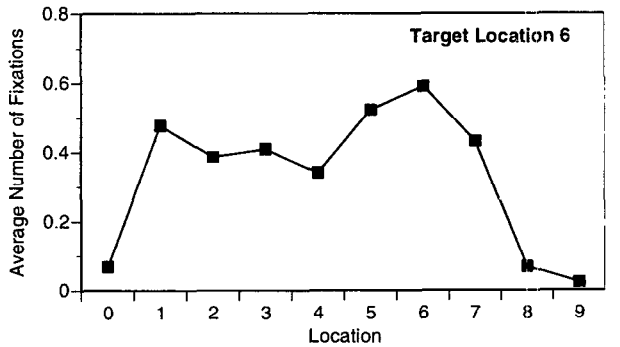


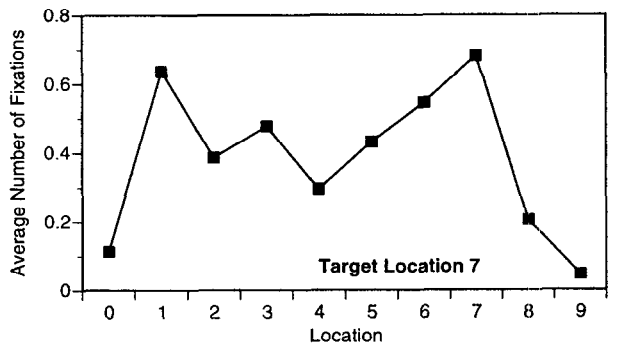












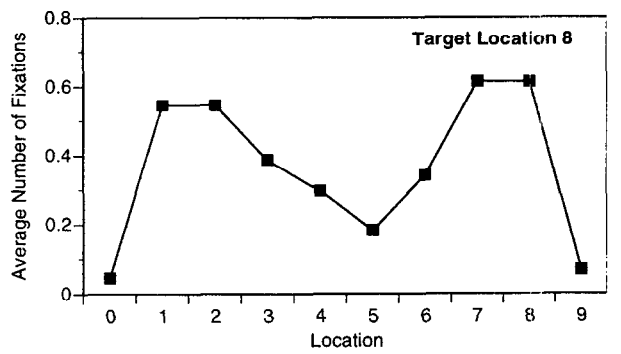
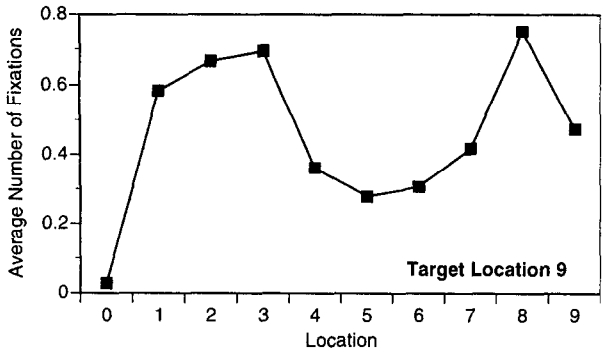


Figure 6 (continued on following page). Fixation frequency as a function of serial position for targets at location 1 through 9. Item 0 is the "Target: X" item.



average number of fixations for positions 7-9 when the target is in the upper third of the menu is effectively zero. Even when the target is in the second third of the menu (items 4-6), the number of fixations on locations 8 and 9 is quite small—much smaller than a random search model would predict. Even if a random search model were modified to systematically begin a search with an item in the 1-3 range (as seems the users' pattern), later fixations should still be seen with some regularity on the lower items in the menu—but this is not what appeared in our data.

Instead, the modal fixation location moves systematically down the menu as the target moves down the menu, suggesting a search that is predominantly top-to-bottom. These data further suggest that the search is non-exhaustive; that is, not all menu items with location less than the target item receive full consideration. Since it is unlikely that users can foveate three items at a time, a systematic and exhaustive search of the menu should average 1/2 of a fixation on each item all the way to the bottom (if necessary). However, for targets at the bottom of the menu (items 8 and 9), the middle menu items average less than half a fixation. That is, there is evidence that users skip intermediate items on their way to the bottom of the menu. This is consistent with the ACT-R model but not with the EPIC model.

On the other hand, the data are not wholly consistent with the ACT-R model, either. The ACT-R model predicts that if the item is in position N, none of the items with position greater than N will be examined—that is, search is entirely top-to-bottom with no extra search below the target item. This, too, is clearly not the case. Figure 5 shows that users frequently begin with their first fixation below the initial menu item, which should not happen if the ACT R model is correct. Further, the first few panels of Figure 6 show that users average well over one fixation in the later part of the menu even when the target is in the first two menu locations, meaning there is at least some search past the target. This is inconsistent with the ACT-R model of menu search.

Interestingly, there is considerable variability both between users and from trial to trial in the each user. There are trials that certainly appear to be top-to-bottom exhaustive searches, trials that appear more or less random (especially for 6-item menus), and trials that appear to be top-to-bottom searches that skip items. However, neither of these models appears to be a good characterization of either the mean or modal behavior of the users.

Mouse Movement

Both the EPIC model and the ACT-R model make predictions about mouse movements as well as eye movements. The EPIC model predicts that there should be a single aimed mouse movement from the initial position to the target item once that target item has been located. Timing of this movement should be governed by Fitts' law. The ACT-R model, on the other hand, predicts that the mouse should "trail" the eyes such that once the target item is located, there should be an approximately constant and short distance between the current mouse location and the target. This predicts multiple mouse movements, directly related to the number of eye movements.

Once again, the data appear inconsistent with both models. Figure 7 depicts the number of mouse "fixations" (as defined by the velocity-based post processing algorithm described earlier) vs. the target location for the three menu sizes used.

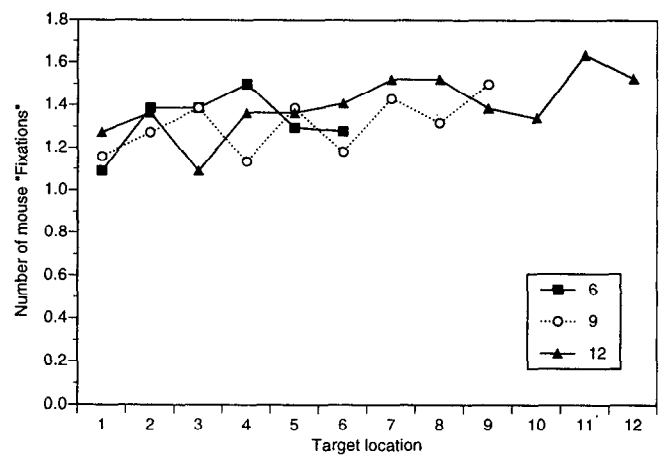


Figure 7. Average number of mouse fixations as a function of target location, by menu length

The EPIC model predicts that this should be a flat function at 1.0. The ACT-R model predicts this should be a monotonically increasing function that should approximate the eye movement data (Figure 4). Instead, the data show that, while users seem to average more than one mouse fixation, they average considerably fewer fixations than they do eye movements. There seems to be a very slight upward trend in the number of mouse movements as the

target appears in later menu positions, but this effect appears quite small (approximately 0.03 additional mouse fixations per serial position).

Further information about users' strategies in terms of moving the mouse can be found in Figure 8, which shows the average location of the initial mouse fixation. The EPIC model predicts that users make one fixation on the target item and thus a slope of 1 throughout the full range of serial positions. The average initial mouse position increases linearly with target position until the last few target positions where it tends to flatten off. The slope for the linear portion is approximately 0.65, implying that the initial mouse fixation is often short of the target.

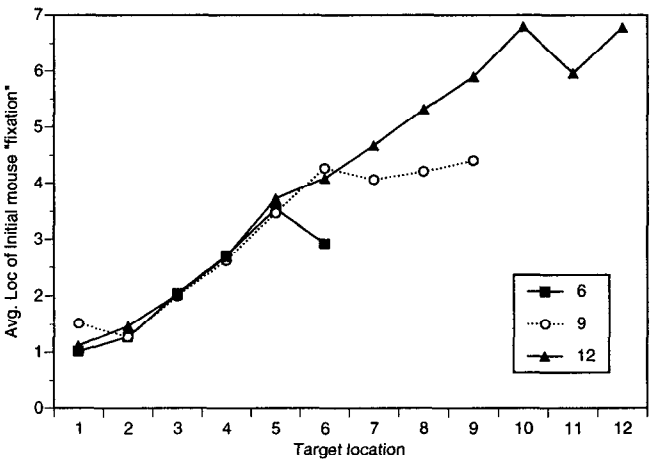


Figure 8. Average location of the initial mouse fixation as a function of target location, my menu size

The ACT-R model predicts that the initial mouse fixation should be relatively insensitive to the target location, particularly for late targets in longer menus. So neither model is consistent with these data. [3]

As with eye movements, both mouse movement strategies (wait until the target is found and trail the eyes with the mouse) quite clearly appear in the data on some trials for some users. The actual data appears to be some mixture of these two (and possibly other) strategies.

DISCUSSION

Despite the fact that both the EPIC and ACT-R models provide good fits to the Nilsen latency data, a finer-grained analysis at the level of eye and mouse movements reveals that neither of the models accurately characterize the visual search and mouse movement strategies employed by real users. What would more accurately characterize the search? These properties seem to be key:

• The initial eye fixation is modally to the initial menu item and almost always to one of the first three items.

• Search is primarily, though not exclusively, top-to bottom. Search rarely appears to be random.

• Some items are not foveated at all, that is, that they are "skipped" by the top-to-bottom search.

• Though the evidence is not conclusive, examining the individual protocols indicates that most violations of top-to-bottom search come from the target item being skipped by the top-to-bottom search and found by backtracking. Users also occasionally move their eyes down the menu without passing the target, backtrack, and then proceed back down.

One plausible model is a "noisy" top-to-bottom search that sometimes skips items and backtracks, which in some cases would give the appearance of a random search—especially for short menus—but would, particularly for longer menus, produce predominantly top-to-bottom searches.

With respect to mouse movement, the truth appears to lie somewhere in between the EPIC model's single move strategy and the ACT-R model's many-move strategy. Further analysis of the data will be necessary to generate a clear model of the mouse movement data and its relationship to eye movements.

It is interesting to note that other researchers have claimed both entirely random [4] and entirely top-to bottom [7] searches.What we observed is something in between, though not a simple 50/50 mixture of the two. With lower resolution eye-tracking equipment or a slightly different task (perhaps one more like reading, which likely biases the users towards top-to-bottom), it is not difficult to imagine coming to one or the other conclusion. However, as is often the case in studies of human behavior, the story is more complex than it initially appears.

The immediate design implications for this work are not obvious, as the menu task itself is somewhat artificial. However, these data suggest that interface designers should make few assumptions about what items users will and will not process when they make selections from unfamiliar menus. While it is very likely that users will see one of the initial items, it cannot be assumed that users will have seen intermediate items, particularly for longer menus.

[3] A secondary issue with respect to mouse movements is the detection of sub-movements, as Fitts' movements have been shown to consist of several sub-movements [e.g. 10]. In order to avoid this problem, the velocity threshold was set conservatively high—high enough that the terminal stop on the target item was occasionally not detected. Thus, we are confident these results are not a function of detection of sub-movements.

The ultimate value of such work lies primarily in its informing of more accurate models of human cognition and performance. Such models are valuable to the field in that they allow the evaluation of interfaces, even ones that have not been constructed yet, without expensive user tests. This work also highlights the utility of advanced data collection methodologies such as eye tracking.

FUTURE WORK

The analyses presented here merely scratch the surface of an incredibly rich data set. We have only begun to examine certain aspects of the data, such as the temporal relationship between the eye and mouse movements. The post-processing algorithms are constantly being improved and may shed more light on the data we already have. We have also not yet examined in great detail the influence of target type, distractor type, and the presence/absence of the initial button during the course of the trial. Clearly, there is a great deal of work yet to be done.

Once we have achieved a clearer understanding of the data, we hope to construct a computational cognitive model of the menu search process using a new extension of ACT-R, ACT-R/PM [3], which incorporates a number of key features of the EPIC system into the original ACT-R architecture. Hopefully, through a clearer understanding of both the data and a more complete model, it will be possible to improve upon current guidelines and tools for evaluating displays used in human-computer interfaces.

ACKNOWLEDGEMENTS

This research was supported in part by the Office of Naval Research (ONR), Award #N00014-96-10491 and by the National Institute for Mental Health (NIMH), fellowship #2732-MH19102. The views and conclusions expressed here are those of the authors and do not reflect the views or opinions of the ONR, the NIMH, or the U. S. government

REFERENCES

1. Aaltonen, A., Hyrskykari, A., & Räihä, K. (1998). 101 Spots, or how do users read menus? In *Human Factors in Computing Systems: Proceedings of CHI 98* (pp. 132–139). New York: ACM Press.

2. Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction, 12*, 439–462.

3. Byrne, M. D., & Anderson, J. R. (1998). Perception and Action. In J. R. Anderson & C. Lebiere (Eds.) *Atomic Components of Thought* (pp. 167–200). Mahwah, NJ: Lawrence Erlbaum.

4. Card, S. K. (1984). Visual search of computer command menus. In H. Bouma & D. G. Bouwhuis (Eds.) *Attention and Performance X: Control of Language Processes* (pp. 97–108). London: Lawrence Erlbaum.

5. Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction.* Hillsdale, NJ: Lawrence Erlbaum

6. Hornof, A. J., & Kieras, D. E. (1997). Cognitive modeling reveals menu search is both random and systematic. In *Human Factors in Computing Systems: Proceedings of CHI 97* (pp. 107–114). New York: ACM Press.

7. Lee, E., & MacGregor, J. (1985). Minimizing user search time in menu retrieval systems. *Human Factors, 27*, 157–162.

8. Nilsen, E. L. (1991). Perceptual-motor control in human-computer interaction (Technical Report Number 37). Ann Arbor, MI: The Cognitive Science and Machine Intelligence Laboratory, the University of Michigan.

9. Norman, K. L. (1991). *The Psychology of Menu Selection: Designing Cognitive Control of the Human/Computer Interface*. Norwood, NJ: Ablex.

10. Rosenbaum, D. A. (1991). Human motor control. New York: Academic Press.