The Limits Of Expert Performance Using Hierarchic Marking Menus

Gordon Kurtenbach and William Buxton

Dept. of Computer Science
University of Toronto
Toronto, Ontario Canada, M5S 1A1
Phone: (416) 978-6619, Email: gordo@dgp.toronto.edu, willy@dgp.toronto.edu

ABSTRACT

A marking menu allows a user to perform a menu selection by either popping-up a radial (or pie) menu, or by making a straight mark in the direction of the desired menu item without popping-up the menu. A hierarchic marking menu uses hierarchic radial menus and "zig-zag" marks to select from the hierarchy. This paper experimentally investigates the bounds on how many items can be in each level, and how deep the hierarchy can be, before using a marking to select an item becomes too slow or prone to errors.

KEYWORDS: Marking menus, pie menus, gestures, pen based input, accelerators, input devices

INTRODUCTION

The first couple of times I went into the restaurant I got a menu and surveyed my choices. I generally ordered vermicelli and barbecue pork by saying "dish number 30, please". On my fifth or sixth visit I knew what I wanted and was in a hurry. I didn't wait to see a menu. I looked my waiter in the eye and said, "Lloyd, bring me a number 30". Things happen faster when you know what you want.

This story reveals the philosophy behind an interaction technique we call "marking menus". Marking menus are a type of pop-up menu where, if the user can recall the location of the item in the menu, the item can be selected without having to pop-up the menu. Just like in the story, it's nice not having to wait for Lloyd to bring a menu when you know what you want and how to order it.

Menus are used extensively in human computer interfaces. They provide critical information on what commands are available and a way to invoke commands. Unfortunately, many computer menus do not provide the kind of service that the restaurant in our example does. One cannot order without having to look at the menu and this can be a problem. Some menus require substantial computing before display and this delays the user. Also, menus appearing and disappearing on the screen can be visually disruptive. Finally, a menu may obscure objects on the screen that are the focus of attention.

Some systems do provide methods to by-pass menus but the by-pass mechanism requires an action that is radically different than selecting using the menu. For example, in some systems, a user selects from a menu using the mouse but by-passes the menu using an "accelerator key" on the keyboard. Using our restaurant example again, this would be like changing from ordering verbally, when one has the menu, to ordering with hand signals when one doesn't have the menu. The problem is that one has to learn two different protocols.

Marking menus are designed to overcome this problem. Using a marking menu with a pen based computer works as follows. A novice user presses down on the screen with the pen and waits for a short interval of time (approximately 1/3 second). A radial menu [13] [4] then appears directly under the tip of the pen. A user then highlights an item by keeping the pen pressed and making a stroke towards the desired item. If the item has no sub-menu, the item can be selected by lifting the pen. If the item does have a sub-menu, it is displayed. The user then continues, selecting from the newly displayed sub-menu. Figure 1 (a) shows an example. Lifting the pen will cause the current series of highlighted items to be selected. The menus are then removed from the screen. At any time a user can indicate "no selection" by moving the pen back to the center of the menu before lifting, or change the selection by moving the pen to highlight another item before lifting. Finally a user can "back-up" to a previous menu by pointing to its center.

The other, faster, way to make a selection without popping up the menu is by drawing a mark. A mark can be drawn by pressing the pen down and immediately moving. The shape of the mark dictates the particular series of items selected from the menu hierarchy. Figure 1 (b) shows an example.

The first important point to note is that *the physical movement involved in selecting an item from the menu is identical to the physical movement required to make the mark corresponding to that item*. With marking menus, a user actually *rehearses* the physical movement involved in making the mark every time a selection from the menu is made. We believe that this helps users learn the markings. The second point to note is that *supporting radial menus with markings in this way helps users make an efficient transition from novice to expert*. Novices perform menu selection because they are not familiar with the menu and its layout. As they become experts, they begin to use the
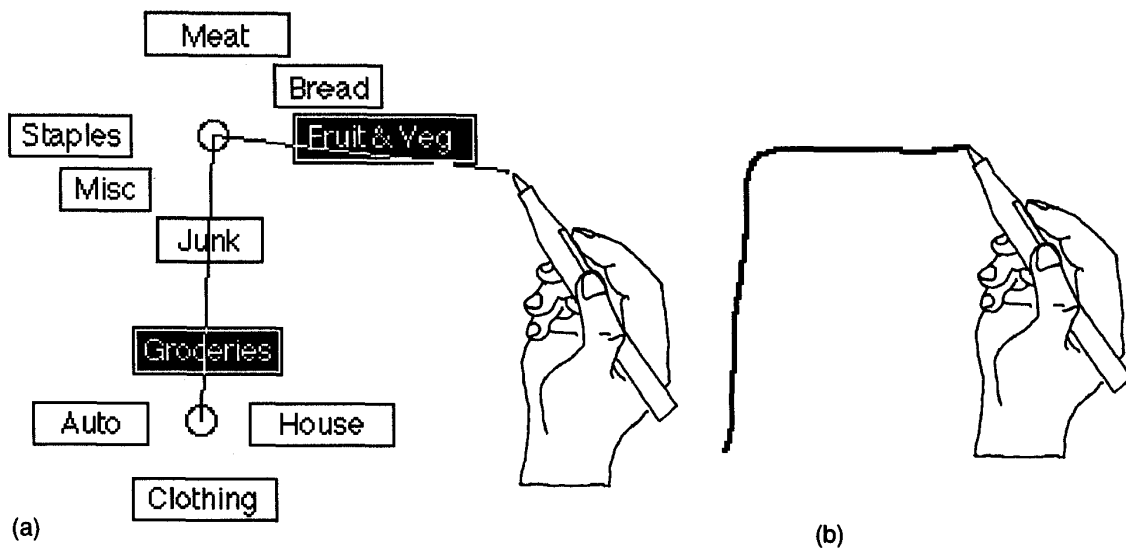
Figure 1. Hierarchic marking menus can be selected from using two different methods. Using method (a), radial menus can be sequentially displayed and selections made. Method (b) uses a marking to make the same selection. Method (a) is good when the user is unfamiliar with the menu. Method (b) is good when the user is familiar with the menu and wants to avoid waiting for the display of each menu.

markings instead. Novices, in effect, "learn on the job" because these two activities are so similar.

Is marking much faster than using the menu? In a study of user behavior with non-hierarchic marking menus in a real application, we found that using a mark was approximately 3.5 times faster than using the menu, even if the 1/3 of second delay to pop-up the menu was subtracted from menu selection time (for example, one user required on average 0.2 seconds to select using a mark and 0.7 seconds to select using the menu) [8]. Displaying the menu actually takes a small amount of time (0.15 seconds in our system). The larger amount of time is consumed waiting for the user to react to the display of the menu, even if the location of the desired menu item is known. While these time savings may seem trivial, one user performed approximately 15,000 selections over 36 hours of work. Using marks helped her complete the task 1.25 hours sooner.

We performed an experiment to address questions one asks when designing an interface that will use hierarchic marking menus.

Q1: Are hierarchic marking menus a feasible idea? Non-hierarchic marking menus have proven to be feasible [9]. Other research has shown that radial menus are faster than linear menus [1]. Thus we can expect marking menus, even without using the faster marking ahead technique, to be faster that traditional linear menus. Nevertheless, the question remains as to whether it is possible to use a marking to select hierarchic menu items.

Q2: How deep can one go using a marking? Just how "expert" could users become? Could an experienced user

use a mark to select from a menu which had 3 levels of hierarchy and twelve items at each level? By discovering the limitations of the technique we would be able to predict what menu configurations, with enough practice, will lead to reliable selection using a marking, and which menu configurations, regardless of the amount of practice, will never permit reliable selection using a marking. Also, will some items be easier to select regardless of depth? For example, it seems easier to select items that are on the up, down, left and right axes even if the menus are cluttered and deep.

Q3: Is breadth better than depth? Will wide and shallow menu structures be easier to access with markings than thin and deep ones? Traditional menu designs have breadth/depth trade-offs [5]. What sort of trade off exists for marking menus?

Q4: Will mixing menu breadths result in poorer performance? A previous experiment on non-hierarchic marking menus has shown that the number of items in a menu and the layout of those items in the menu affects selection performance when markings are used [9]. Specifically, menus with 2, 4, 6, 8 and 12 items work quite well for markings. What will be the effect of selecting from menu configurations where the number of items in a menu varies from sub-menu to sub-menu?

Q5: Will the pen be more suitable than the mouse for making marks? The experiment mentioned above also compared making selections from non-hierarchic marking menus using a tablet with a stylus, a trackball and a mouse. The trackball was the worst performer, while the tablet with stylus and mouse performed equally. However, hierarchic

483

marking menus require more complex marks. Will the mouse prove inadequate?

THE EXPERIMENT

Basic Design

In order to determine the limits of performance, we needed to simulate expert behavior. We defined expert behavior as the situation where the user is completely familiar with the contents and layout of the menu and can easily recall the marking needed to select a menu item. To make subjects "completely familiar" with the menu layouts we chose menu items whose layout could be easily memorized. We tested menus with 4, 8 and 12 items. For a menu of 4 items, the labels were laid out like the four points of a compass: "N", "E", "S" and "W". We referred to this type of menu as a compass4. Similarly a "compass8" menu had these four directions plus "NE", "SE", "SW" and "NW". Menus with twelve items, referred to as a "clock" menus, were labeled like the hours on a clock.

Will users of real applications ever be as familiar with menus as they are with a clock or compass? We believe the answer is yes and base this on three pieces of evidence. First, our own behavioral study of users using a marking menu in a real application shows, with practice, markings are used over ninety percent of the time [8]. Other researchers have reported this type of familiarity with pie menus [4]. Second, research has shown that the effects of menu organization disappear with practice [2][10]. In other words, with practice, users memorize menu layouts and navigate directly to the desired menu item. Finally, it must be remembered that a user does not have to memorize the layout of an entire menu. For example, a hierarchic marking

menu could contain 64 items but the user might only memorize the markings needed to select the two most frequently used menu items.

The general design of a trial in our experiment was as follows. The system would ask the subject to select a certain item using a marking (the menu could not be popped up by the subject). The subject would draw the marking and the system would then record the time taken, and whether or not a successful selection was made. We would then vary the menu configuration and input device and see what effect these variables had on selection performance.

Method

Subjects: Twelve right handed subjects were recruited from University of Toronto. All subjects were skilled in using a mouse but had little or no experience using a pen on a pen based computer.

Equipment: A Momenta pen based computer was used. The input devices consisted of a Microsoft mouse for IBM personal computers, and a Momenta pen and digitizer.

Task: A trial occurred as follows. The type of menu configuration currently being tested would appear in the top left corner of the screen. A small circle would appear in the center of the screen. A subject would then press and hold the pen or mouse button over the circle. The system would then display instructions describing the target at the top center of the screen. A subject would then respond by drawing a mark that was hoped to be the correct response. The system would respond by displaying the selection produced by the marking. If the selection did not match the target, the system would beep to indicate an error. The
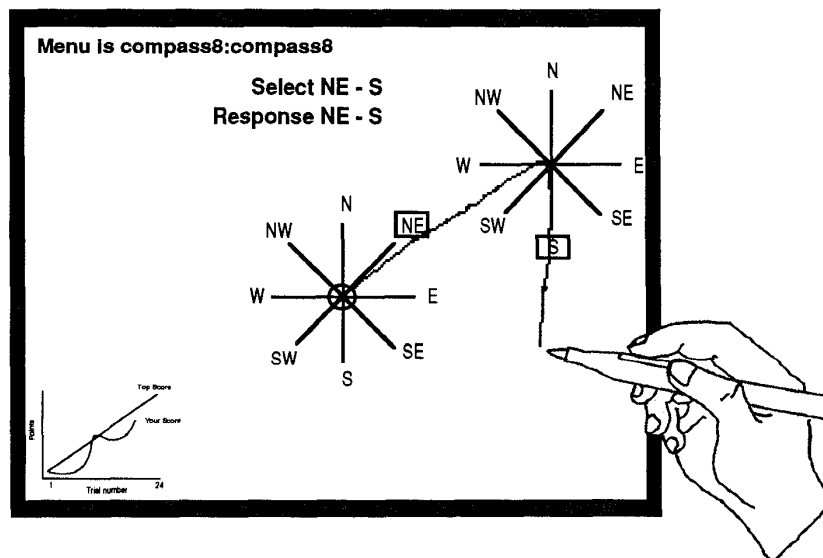

Figure 2: The experiment screen at the end of a trial where the target was "NE-S". After the marking was completed, the system displayed the menus along the marking to indicate to the subject the accuracy of their marking.
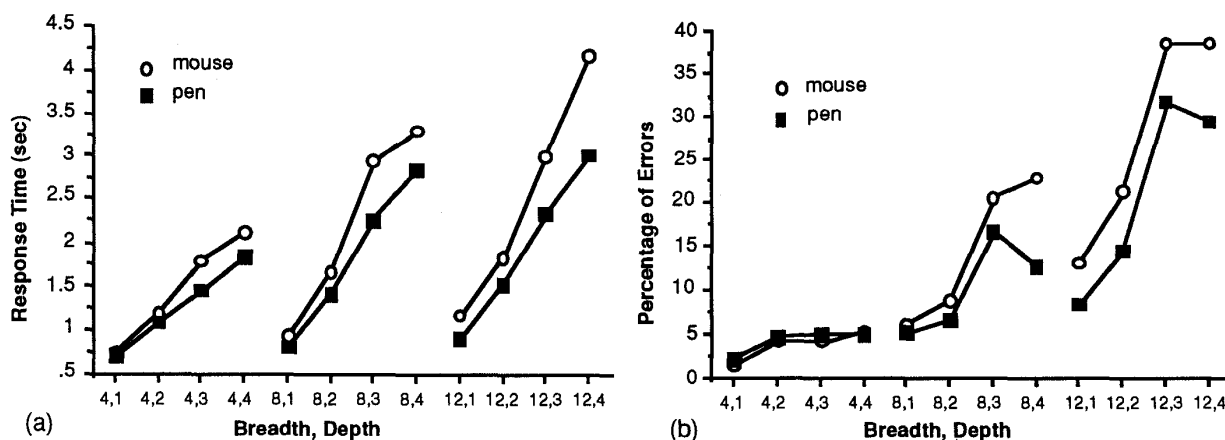
484

Figure 3: Response time and percentage of errors as a function of menu breadth, depth and input device. Each data point is the average of 288 trials.

system would then display each menu in the current menu configuration at its appropriate location along the marking and indicate the selection from each menu. A subject's score would then be shown in the lower left of the screen. Figure 2 shows the experimental screen at this point. If a selection was incorrect, a subject would lose 100 points and the trial would be recorded as an error. If a selection was correct, the subject would earn points based on how quickly the response was executed. Response time was defined as the time that elapsed between the display of the target and the completion of the marking.

Design: All three factors, device, breadth and depth were within-subject. Trials were blocked by input device with every subject using both the pen and the mouse. One half of the subjects began with the pen first while the other half began with the mouse. For each device, a subject was tested on the 13 menu configurations (breadths 4, 8 and 12 crossed with depths 1 to 4, plus the mixed menu configuration of clock:compass8:clock). Menu configurations were presented in random order. For each menu configuration, a subject performed 24 trials. During the 24 trials, subjects were repeatedly asked to select 1 of 3 different targets. Each target appeared eight times in the 24 trials but the order of appearance was random.

Before starting a block of trials for a particular menu configuration, subjects were allowed 8 seconds to study the menu configuration. Before starting trials with a particular input device, a subject was given ten practice trials using the device on a compass4:compass4:compass4 menu.

Results and Discussion

All three factors, input device, breadth and depth affected response time. Analysis of variance revealed a three way interaction between input device, breadth and depth $(F(6,66)=3.32, p < .05)$ affecting response time. Figure 3 (a) shows these relationships. As one would expect, increasing breadth and depth increases response time, however,

subjects' performance degraded more quickly with the mouse than with the pen.

Subjects responded significantly faster with the pen than with the mouse $(F(1,11)=19.7, p < .001)$. The response time averaged across all subjects, breadths and depths for the pen was 1.69 seconds while the mouse averaged 2.07 seconds. As menu breadth and depth increased subjects' performance with the two devices began to differ. This is shown in figure 3.

Subjects produced significantly more errors with the mouse than with the pen $(F(1,11)=6.41, p < .05)$. Both depth and breadth interacted to affect error rate $(F(6,66)=12.28, p < .001)$. Figure 3 (b) shows that mouse and pen error percentages began to differ once menu breadth reached eight items. For either device, error rates were below 10% for up to menus of breadth 8 and depth 2.

We tested for effects of mixing menu breadths in menu configurations by comparing the performance of a clock:clock:clock menu with a clock:compass8:clock menu. We found no significant performance differences between the two menu configurations.

In order to test the hypothesis that markings which consist of "on axis" items (items on the vertical and horizontal axes) out-perform "off axis" markings, we picked targets for menus of breadth twelve, depths two, three and four such that the experimental data could be divided into 3 groups. With each group we associated an "off axis-level": a1, a2 and a3. Experimental data was placed in group a1 if the target consisted strictly of menu items that were on-axis, such as "12-3-9-3". Group a3 consisted of data on targets that consisted of entirely off-axis targets such as "1-2-1-2". Group a2 consisted of data on targets that were a mixture of on-axis and off-axis menu items, such as "12-7-3-9". Figure 4 shows that axis level had a significant effect on response time $(F(2,22)= 104.84, p < .001)$ and on percentage of errors $(F(2,22)=36.2, p < .001)$. Figure 4 (a) shows how the type of device interacted with off-axis level $(F(2,22)=6.93, p <$

485

.05). This indicates that subjects' response time on the worse off-axis targets did not degrade as badly with the pen as it did with the mouse.

CONCLUSIONS

We can now revisit the questions posed at the start of this paper.

Q1: Are hierarchic marking menus a feasible idea? Even if using a marking to access an item is too hard to draw or cannot be remembered, a user can perform a selection by displaying the menus. Nevertheless, since the subjects *could* perform the experiment, it is feasible that markings could be used to select hierarchic menu items.

Q2: How deep can one go using a marking? Our data indicates that increasing depth increases response time linearly. The limiting factor appears to be error rate. For menus of four items, even up to four levels deep, the error rate was less than ten percent. This is also true for menus of eight items, up to a depth of two. However, when using markings for menus with eight items or more, at depths greater than two, selection becomes error-prone, even for the expert. However, our "off-axis" analysis indicates that the source of poor performance at higher breadths and depths is due to selecting "off-axis" items. Thus, when designing a wide and deep menu, the frequently used items could be placed at "on-axis" locations. This would allow some items to be accessed quickly and reliably with markings, despite the breadth and depth of the menu.

What is an acceptable error rate? The answer to this question depends on the consequences of an error, the cost of undoing an error or redoing the command, and the attitude of the user. For example, we have data that indicates, in certain situations, experts produce more errors than novices [11]. The experts were skilled at error recovery and thus elected to trade accuracy for fast task performance. Our experiences with marking menus with six items being used in a real application indicates that experts perceived selection to be error-free. Other research reports that menus with up to eight items produce acceptable performance [4]. Marking menus present a classic time versus accuracy trade-off. If the marking error rate is too high, a user can always use the slower but more accurate method of popping up the menus to make a selection.

Q3: Is breadth better than depth? For menu configurations that resulted in acceptable performance, breadth and depth seems to be an even trade-off in terms of response time and errors. For example, accessing 64 items using menus of four items, three deep, is approximately as fast as using menus of eight items, two deep, and both have approximately equivalent error rates. Thus, within this range of menu configurations, a designer can let the semantics of menu items dictate whether menus should be narrow and deep, or wide and shallow.

Q4: Will mixing menu breadths result in poorer performance? The experiment did not show this to be true. This may be due to the fact that our menu labels strongly suggested the correct angle to draw at, and therefore confusion was avoided. A stronger test might be to compare mixing menu breadths using less suggestive labels. However, our results do indicate that, with enough familiarity with the menus, mixing breadths is not a significant problem.

Q5: Will the pen be better than the mouse for making marks? Overall, subjects performed better with the pen than with the mouse. However, for small menu breadths and depths, subjects' performance, with either the mouse or pen, was approximately equivalent. We found this extremely encouraging because it implies that a marking menu is an interaction technique that not only takes advantage of the pen but also remains compatible with the mouse.

FUTURE DIRECTIONS

We are currently experimenting with designing interfaces that use marking menus. When displays become small or very large, marking menus are effective. A mark or selection can be made at a user's current location without a trip to a menu bar or tool pallet. We are currently trying to exploit this advantage on a electronic whiteboard system
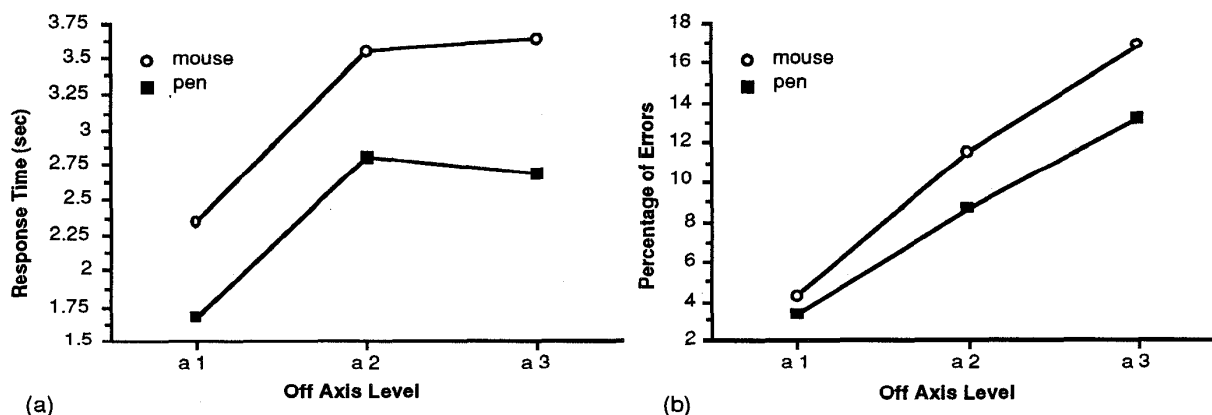

Figure 4: Average response time and percentage of errors for targets with an increasing number of "off-axis" items.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.