Navigation Time Variability: Measuring Menu Navigation Errors

Krystian Samp[1], Stefan Decker[1]

[1] Digital Enterprise Research Institute, National University of Ireland, Galway
{krystian.samp, stefan.decker}@deri.org

Abstract. The subject of errors in menu studies is typically limited to reporting error rates (i.e., the number of clicks missing target items) or even completely neglected. This paper investigates menu navigation errors in more depth. We propose the *Navigation Time Variability* (NTV) measure to capture the total severity of navigation errors. The severity is understood as time needed to recover from the errors committed. We present a menu study demonstrating use and value of the new measure.

Keywords: Navigation Time Variability, errors, navigation, menus.

1 Introduction

There has been a great deal of research into alternative menu designs, including numerous cascading menu improvements [1, 4, 7, 12], radial menus [11], marking menus [8], and much more.

Research studies of menu designs typically focus on measuring performance in the form of selection times and error rates in the form of the number of clicks missing the target menu item. Sometimes the subjective perception is also measured with a post-experiment questionnaire.

When it comes to errors, it turns out that the error rates are low, typically below 5% (e.g., [1, 4, 5, 11, 12]). The error rates are often not significantly different for tested designs. Some studies abandon the analysis of errors, concluding that the errors are too sparse to provide any interesting insights (e.g., [5]). These results would suggest that errors do not play an important role in menu selection.

It is important to note, however, that the traditional way of measuring errors focuses only on one particular type of navigation error—i.e., clicks missing a target item. However, navigation does not consist only of mouse clicks but of all motor actions that the user needs to perform when selecting from a menu (e.g., dwelling or moving the mouse pointer). These other actions can result in errors which are not captured by counting the number of incorrect mouse clicks. For example, in the cascading menu small steering errors causing incorrect selection changes or unexpected sub-menu disappearance do not increase number of incorrect mouse clicks. Navigation is an important part of menu selection, concerning both novices and experts [3]. Therefore, it is important to understand the problem of menu navigation errors more thoroughly.

In this paper we propose a simple measure of navigation errors in menus. The data required by the measure is typically collected in menu studies. Our goal is to describe the measure and demonstrate its use and value in a traditional menu study.

The rest of the paper is structured as follows. First, we present the related research. Next, we describe the measure. The menu study follows where we demonstrate the use and value of the measure. The paper finishes with the conclusions.

2 Related research

There are three components of menu selection: visual search, decision, and navigation [3]. Novices search for an item and then navigate to it. Experts, on the other hand, do not have to search for menu items as they know their locations. They decide which item to select and then navigate to it. If the menu is hierarchical, search and navigation (novices) or decision and navigation (experts) are performed multiple times for each menu level.

Navigation concerns novices and experts. It refers to all motor actions that the user needs to perform when selecting from a menu (e.g., moving the mouse pointer, dwelling, mouse clicks). Navigation is an important part of menu selection. Novices and experts can spend between 20-50% of the total selection time on navigation [11].

There are many possible differences in how menus are navigated. Some menus employ a point-and-click interaction style [11], others a dragging interaction style [8]. Some menus restrict navigation trajectory when moving between the sub-menus, others do not [4, 12]. Some menus make the navigation to frequently selected items easier by dynamically increasing their sizes [3]. Some menus facilitate navigation between the sub-menus; for example, by attracting the cursor towards an open sub-menu [1] or by opening a sub-menu faster if the cursor moves towards it [7]. These are only few examples. The plethora of menu navigation techniques makes it important to understand how error prone they are, beyond the number of clicks missing target items.

The importance of navigation errors is informally established in the HCI community. Numerous cascading menu improvements are motivated by various navigation problems [1, 4, 7, 12]. Pastel [10] shows that steering through sharp corners, like in the case of the cascading menu, can induce errors. Kobayashi [7] provides empirical evidence that the number of unexpected submenu appearances, again for the cascading menu, can be substantial. These works hint at some navigation problems beyond traditional error rate but are quite specific in what they focus on (i.e., one type of error particular for the cascading menu). None of the work has focused on a general approach to measuring the total impact of all navigation errors.

To address this problem, an approach focusing on measuring errors of pointing devices could be considered. MacKenzie et al. [9] proposed seven accuracy measures to elicit differences among pointing devices in precision pointing tasks :1) target re-entry 2) axis crossing, 3) movement direction change, 4) orthogonal direction change, 5) movement variability, 6) movement error, and 7) movement offset. The measures, however, do not necessarily represent menu navigation errors. Axis-crossing or movement changes do not have to result in errors such as sub-menu disappearance or

selection changes – i.e., errors that the user needs to correct. This will depend on particular menu design and size of the committed error. The above measures are well suited to assess deviations from an optimal pointing solution. It is arguable, however, if such a general optimal pointing solution exists for hierarchical menus (e.g., should the steering finish at the border of a sub-menu, in its center, or somewhere else?).

The above approach and other approaches based on counting specific types of errors also pose demands on the experimental software. All the possible errors have to be tracked individually. This might be difficult in some environments such as those employing third-party applications or toolkits.

We aim at creating a simpler and more abstract approach. We want to refrain from listing a priori all possible types of navigation errors and tracking these in the experimental software. Our goal is to assess the total impact of all navigation errors, not contribution of the individual predetermined types of errors. Such an approach would enable more immediate view of the problem of navigation errors in menus.

3   Navigation Time Variability measure

There are many possible sources of navigation errors stemming from the many different ways of navigating menus (some examples mentioned in section 2).

We propose the following view of the problem of navigation errors: A user trying to be as fast and as accurate as possible should be able to navigate to the same target item multiple times within a similar time frame. An increased variability in navigation time for the same target item indicates navigation errors. This is because recovering from navigation errors (e.g., re-pointing to a target item) requires additional time which is not strictly related to navigating towards a desired item.

Note that the increased variability is not necessarily connected with an event such as sub-menu disappearance. The corrections done by the users to prevent the errors—for example, a temporary change of speed-accuracy strategy preventing sub-menu disappearance—will also increase variability. This is a desired behavior because such corrections also indicate navigation difficulties which are the object of our interest.

According to the above view, what we focus on is not the occurrence of a particular event indicating an error but rather the occurrence of variability in navigation time indicating extra time spent on recovering from errors. Consequently, it is not we who decide where the navigation error occurred, but rather the user by making necessary, time-consuming corrections.

The above view of navigation errors takes into account severity of the errors. Severity is an important aspect of errors having a strong effect on user perception [6]. In our case, the more severe navigation errors require more time to recover and thus lead to larger variability. This is important because we can expect different types of navigation errors to cause different degree of difficulties.

To formalize the described variability we propose the *Navigation Time Variability* measure (NTV). It is calculated as follows. (1) For each participant we establish the min-max range of navigation times obtained for **the same menu item** with the same menu design. We did not use standard deviation instead of min-max as we expect small number of measurements per item per participant (i.e., two or three—we discuss

this further in section 3.1). Min-max range also assures easy interpretation of the results as it represents the difference between the fastest and the slowest navigation to the same menu item. (2) For each participant, we calculate the average of the min-max ranges across the different menu items. The result is one value per participant representing a single NTV score.

Since the NTV is computed on per participant basis, the inferential statistics can be used to seek significant differences between tested menus.

The NTV measure has four important characteristics. First, the sources of navigation errors do not have to be known in advance. Second, it focuses on all the errors which truly require user time and effort to be corrected. Third, the measure is based on severity of the errors—i.e., more severe navigation errors result in higher scores. Fourth, the measure assures meaningful interpretation—i.e., it expresses, in a statistical sense, how much additional time is required to correct the committed errors. In light of these characteristics, it becomes clear that the measure is not suited to identify sources of the errors or the contribution of the predetermined types of errors. Rather, its goal is to assess and compare the total impact of the navigation errors.

3.1 Measuring navigation times – practical considerations

The navigation times have to be collected in an experiment. However, the experimental task cannot consist only of navigation. Consequently, some variability in collected times (i.e., menu selection times) might be also attributed to other components present in the experimental task. For example, if the task requires a simple decision apart from navigation, the variability of the decision time will contribute to the total measured variability. Therefore, if the goal is to assess navigation errors, it is important to use a task that emphasizes navigation and minimizes contribution of the other components (i.e., the other components should be relatively small compared to navigation and have small variability ranges). In particular, the menu selection task should not require visual search or problem solving because both can take long and introduce extensive variability.

If the task emphasizes navigation, then the NTV computed on selection times will allow one to assess qualitatively which menu causes fewer navigation errors. However, if the goal is to assess the exact quantitative extent of the navigation errors, the contribution of the other components have to be factored out. To this end, the components have to be known and have known variability ranges.

A considerable amount of menu studies have focused on measuring navigation performance. They employ a common task which emphasizes navigation and adheres to the above characteristics. The task is to select an item from a single or hierarchical menu. All items in a selection path (i.e., parent items and the target item) are highlighted. Consequently, the participants do not have to search for each item nor decide which item to select. The participants only need to: 1) respond to visual stimuli and prepare the movement which takes 234 ± 41ms [2] and then 2) navigate to the item. Ahlstrom [1] demonstrated empirically that the above task emphasizes navigation. He accurately modeled total selection times using only navigation component based on the Fitts' and the steering pointing laws.

The above task is commonly used in menu studies (e.g., [1, 4, 12]). It is also common to administer two or more blocks of menu selections for each tested menu design (e.g., [1, 3]). Therefore, there is potential for similar menu studies to use the proposed measure without modifying the experimental design but merely by extending the error analysis part.

4 Experiment

The objective of this section is to demonstrate use and value of the NTV measure in a traditional menu study employing the navigation task presented in section 3.1.

4.1 Menus

Two menus were used in the experiment. The first menu was the Cascading Menu (CS), which we chose because it is known to cause navigation errors [1, 4, 7, 12]. The implemented CS menu was equivalent to those found in contemporary applications such as MS Word. In short, a mouse click was used to open the menu and select the final item and dwelling over a parent item for 1/3 seconds opened a sub-menu. Moving between the sub-menus required steering, the sub-menus did not disappear immediately upon steering errors but after short delay, and the item under the cursor was highlighted blue.
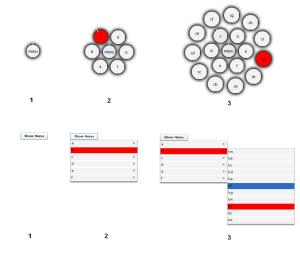
The second menu was the Compact Radial Layout menu (CRL) [11] which we chose because it was designed specifically to decrease navigation errors by the means of eliminating dwelling, eliminating steering, decreasing navigation distances, increasing item sizes, and not restricting navigation paths. The implemented CRL menu was equivalent to that presented in [11]. In short, a mouse click was used to open a menu, select a parent item, and the final item. The items were represented by circles, the levels by concentric rings. The innermost ring corresponded to the first level and again the item under the cursor was highlighted blue. Figure 1 shows examples of selection sequences for both menus.

The size of the CS menu items was set to 215×19 pixels because this size could accommodate three word labels. The diameters for the CRL menu items were 44, 48, and 52 pixels for the first, second and third level respectively. The resulting circles provided enough space to also accommodate three word labels (eventually broken to two or three lines) while avoiding overlaps.

The menu content consisted of 6 items on the first level, 11 items on the second level, and 15 items on the third level. We used one to three character labels to help the user visually separate the menu items; however, they did not have any meaning.

4.2 Measures

The study included quantitative and qualitative measures. Our main dependent variables were selection time and the number of mouse clicks missing target items. In the post-experiment questionnaire, the participants were asked to rank each of the

menus on a 1-7 Likert scale according to the following criteria: *Error prone*, *Frustration*, and *Ease of use*.

Fig. 1. Selection sequence for a target item on the second level for the CRL menu (at the top of the figure) and the CS menu (at the bottom). All items in a selection path are highlighted red.

4.3 Procedure

A total of 28 participants took part in the experiment (8 female and 20 male). All were between ages of 20 and 28. The navigation task presented in section 3.1 was used.

The procedure was as follows. (1) Participants were told the procedure of the experiment. (2) For both menus the system provided a two sentence description of their behavior and allowed the participants to practice for two minutes. (3) The experiment started. The participants were asked to complete the selections as fast and as accurately as possible. For each menu, each participant completed two identical blocks of 45 menu selections (separated by a two minute break): 10 selections on the first level, 15 on the second level and 20 on the third (with a one minute break between the levels). The item sequences were generated randomly. The menu ordering was balanced. The time was measured between the first click opening the collapsed menu and the final click on the target item which collapsed the menu. (4) The participants completed the post-experiment questionnaire.

4.4 Results and Discussion

The traditional measure of error rate indicated that 1.6% of the trials were erroneous (i.e., a mouse click missing a target item) for the CRL menu and 3.7% for the CS menu. Because the scores of error rates were sparse and not normally distributed we used the Wilcoxon Signed-Rank test for statistical analysis. The test indicated that the error rate is not significantly different for both menus ($p > 0.05$).

The results of the questionnaire were as follows. All the participants claimed that they were more error prone with the CS menu. 18 participants were more frustrated with the CS menu, 4 with the CRL menu, and 6 did not see any difference. 6 participants considered the CRL menu easier to use, 5 the CS menu, and 17 considered both menus equal. Because the subjective scores are non-parametric, we again used the Wilcoxon Signed-Rank test. The CRL menu is perceived as significantly less erroneous (p<0.01) and less frustrating (p<0.01) than the CS menu. No difference was found with respect to ease of use.

The results indicate that the CS menu is strongly perceived as more erroneous and more frustrating. However, this perception cannot be attributed solely to the error rate (i.e., the number of clicks missing target items) as it was not significantly different for both menus. The subjective perception is not supported quantitatively. The results hint that there is more to navigation errors than clicks missing their target items.

Using data from two blocks, we calculated the NTV for each participant. Table 1 shows the summary of the results.

Table 1. The NTV on the three menu levels averaged across the participants. The NTV scores followed normal distribution within each menu x level cell.

	Level 1	Level 2	Level 3	Marginal mean
CRL MENU	85 ms	164 ms	232 ms	160 ms
CS MENU	176 ms	862 ms	1005 ms	681 ms

To determine if the NTV differences are significant, we performed pair-wise comparisons using dependent measures t-test. The family-wise significance level was adjusted according to the number of tests performed. The comparisons revealed that the menus were not different on the first level but the CRL menu generated lower NTV than the CS menu on the second and the third level. For the CS menu, each level produced higher NTV compared to the previous levels. For the CRL menu, the first level had lower NTV than the second and third levels which did not differ between themselves. All the reported differences are at level p<0.001.

The CRL menu. For the CRL menu, the NTV on the first level is the lowest (85ms). It can be attributed to the variability of reaction time of aimed movements (82ms [2]). The NTV on the second and the third level are doubled and tripled respectively as these levels require one additional reaction time compared to the previous level. There is no sign of any additional navigation variability which could indicate navigation errors. The participants were able to maintain roughly constant navigation performance on each level when selecting the same items.

The CS menu. For the CS menu, the NTV rapidly decreases between the levels. This indicates that participants, who performed well with the CRL menu, had more problems with the CS menu. The size of the NTV for the CS menu and the size of the differences in the NTV between both menus certainly cannot be attributed to the variability of reaction time of aimed movements. Furthermore, as the NTV for both menus is similar for the first level, we conclude that it is the navigation between the levels that causes rapid increase of the NTV for the CS menu. After factoring out the variability of reaction time of aimed movements[1] from the NTV scores in Table 1, we

---

[1] Reaction time being 82ms [2] for the first level selections, 2×82ms for the second level selections, and 3×82ms for the third level selections.

estimate that recovering from the navigation errors takes approximately 94ms, 698ms, and 759ms for selections on the first, the second and the third level respectively. These amounts are substantial compared to the typical total selection times—e.g., according to [12], approximately 1750ms and 2500ms for the second and the third level selections. This hints at the importance of navigation errors.

The above results provide quantitative support for the subjective findings regarding errors. To remind the reader, the CS menu was perceived as significantly more erroneous (stated unanimously) and significantly more frustrating that the CRL menu. In contrast to the traditional error rate, the NTV captured this difference.

Finally, the results also demonstrate that the NTV can be small, even for selections on the third level. This finding is important because it supports the underlying assumption of the NTV measure stating that a user can navigate to the same menu item within a similar time frame.

5   Conclusions

Our goal in this paper was to describe a new measure of navigation errors in menus and show its use in a traditional menu study. We demonstrated that the proposed NTV measure gives quantitative information on navigation errors beyond traditional measure of error rate. It also supported the qualitative findings.

The measure is not intended to replace the traditional measures (e.g., error rates). Rather, we consider it supplementary measure, with the potential to assess the total impact of navigation errors. The new measure increases the theoretical knowledge base on differences between menu designs.

References

1. Ahlström, D.: Modeling and improving selection in cascading pull-down menus using Fitts' law, the steering law and force fields. In: CHI, 61-70, (2005).
2. Bekkering, H., Adam, J.J., Kingma, H., Huson, A., Whiting, H.T.A.: Reaction time latencies of eye and hand movements in single- and dual-task conditions. Exp. Brain-Res., 97, (1994).
3. Cockburn, A., Gutwin, C., Greenberg, S.: A Predictive Model of Menu Performance. In: International Conference on Human Factors in Computing Systems (CHI), 627-636, (2007).
4. Cockburn, A., Gin, A.: Faster cascading menu selections with enlarged activation areas. In: Graphics Interface Conference (GI), 65-71, (2006).
5. Ellis, J., Tran, C., Ryoo, J., Shneiderman, B.: Buttons vs. menus: An exploratory study of pull-down menu selection as compared to button bars. Technical Report, Vol. 10, (1995).
6. Feng, J., Sears, A.: Beyond errors: measuring reliability for error-prone interaction devices. Journal Behaviour & Information Technology, 29(2), 149-163, (2010).
7. Kobayashi, M., Igarashi, T.: Considering the direction of cursor movement for efficient traversal of cascading menus. In: UIST, 91-94, (2003).
8. Kurtenbach, G.: The Design and Evaluation of Marking Menus. PhD Thesis, (1993).
9. MacKenzie, I.S., Kauppinen, T., Silfverberg, M.: Accuracy measures for evaluating computer pointing devices. In: CHI, 9-16, 2001.
10. Pastel, R.: Measuring the difficulty of steering through corners. In: CHI, 1087-1096, (2006).
11. Samp, K., Decker, S.: Supporting menu design with radial layouts. In: International Conference on Advanced Visual Interfaces (AVI), 155-162, (2010).
12. Tanvir, E., Cullen, J., Irani, P., Cockburn, A.: AAMU: adaptive activation area menus for improving selection in cascading pull-down menus. In: CHI, 1381-1384, (2008).