EYE MOVEMENT AND REACTION TIME ARE BOTH IMPORTANT IN ASSESSMENT OF DIALOG BOX USABILITY

Bruce N. Walker and Raymond M. Stanley
Sonification Lab, School of Psychology
Georgia Institute of Psychology,
Atlanta, GA 30332-0170
bruce.walker@psych.gatech.edu

ABSTRACT

Traditional usability metrics (accuracy and reaction time) were combined with eye movement patterns to study button placement and highlighting in dialog boxes. Participants made button-click responses based on the contents of the dialog box text. Traditional measures and eye movement patterns yielded different results: Reaction time analyses suggested placing the correct button to the left; eye movement patterns suggested placing the correct button on the right. This study demonstrated that eye movements are a rich source of information for usability research, provided theoretical guidelines for future research, and showed the strengths and weaknesses of eye tracking in comparison to more traditional usability metrics. In addition, it provided empirical support for eye movement heuristics that are often implemented in visual interface design, showing that the search patterns for dialog boxes follow a reading pattern.

INTRODUCTION

The dialog box is an essential element in the modern graphical user interface (GUI). Operating systems automate the process of dialog box creation, with the software developer defining only the message text, and sometimes button options. What little research there has been on dialog box design has focused on the algorithms used to generate them more than on the best design of the elements within the box. Perhaps not surprising, then, different operating systems employ different approaches to the placement of the buttons in a dialog box. Style guides (e.g., Mullett & Sano, 1995) suggest that low level design decisions should follow the standards of the operating environment. However, some design approaches, despite being "standardized", remain to be empirically validated. This can lead to interfaces that are consistent, but perhaps consistently sub-optimal.

As an example, Bodart and colleagues (1994) suggest that elements in an interface should have a "logical ordering," emphasize visual cues, and reduce ocular movements. Although these suggestions are empirical in nature, their evaluation was not. Rather, mathematical relationships between elements of the interface were used to assess the quality of the new design, based on visual principles such as balance, symmetry, and proportion. Unfortunately, mathematical relationships do not necessarily directly relate to usability. Kim and Foley (1993) write that the "Confirm button group" (the OK and Cancel buttons on a dialog button) is the last element to be interacted with. Although guidebooks for designing visual interfaces explicitly state this assumption (e.g., Galitz, 1989), these assumptions remain to be empirically validated.

Many heuristics used to guide design of visual interfaces depend heavily on assumptions about eye movements, such as reading a computer screen from left-to-right and reducing ocular movements for maximum efficiency of design. Eye movements, however, are rarely directly assessed. Patterns of fixation and transitions can be a measure of efficiency of the arrangement of display elements (Jacob & Karn, 2000). It is also important to note that there are no usability standards for eye tracking methods, and little has been done to correlate standard usability metrics with eye tracking (Goldberg & Wichansky, 2000). This is most likely due to pragmatics of eye tracking research: There are frequent technical problems, data extraction is labor-intensive, and the data are difficult to interpret (Jacob & Karn).

Reaction time (RT) and accuracy are generally considered to capture efficiency (work per unit time) and effectiveness (how well the user can perform their task) in an interface (Goldberg & Wichansky, 1993). Subjective metrics such as user satisfaction, gathered through structured interviews and cognitive walkthroughs, are also commonly used to assess an interface. Unfortunately these metrics do not allow us to make many inferences about the internal cognitive processes that are occurring in the course of a decision or action. Eye tracking can allow us to infer about attention, distraction, the areas of a display that have been processed, and how a screen's layout affects usability (Goldberg & Wichansky, 2000).

As a concrete example, in this study we combine eye tracking and traditional usability metrics in an attempt to evaluate whether placing the correct button on the left of a dialog box (typical of Windows) or on the right (typical of Macintosh) is "better" in some way, or if such a distinction even makes sense. This approach should combine more sources of information about not only what the user is doing, but how and perhaps why.

METHOD

Participants

Forty Georgia Tech undergraduates participated for course credit; all were native English speakers. Twelve participants were excluded from analysis because they yielded unsuccessful calibrations. In addition, three participants were discarded from analysis because their response accuracy (percent correct) was more than two standard deviations below the mean (less than 69.5% correct) or the number of trials with data missing was more than two standard deviations above the mean (more than 24 trials missing data). Thus, the data from 25 participants (15 males, 10 females, mean age = 19.7 years) were analyzed. Twenty-three participants reported that they used Microsoft Windows 85% or more of the time. The remaining two reported that they used Windows more than 50% of the time, using Macintosh the rest of the time. All but one of the participants were right-handed.

Procedure

On each trial participants bought or sold a fictitious stock, based on the text in a dialog box that stated the current price and asked if he or she would like to take an action (buy or sell) on the stock. For example, the text of on dialog box was "The price of Meba stocks is now $27. Do you want to buy stocks?" Participants could respond, "Yes" or "No", by clicking on the appropriate button; they had been trained to buy if the price was below $25 and to sell if above $35. The dialog boxes were very generic, similar to those generated by the Windows 98 operating system. The correct action (Buy/Sell), the correct response (Yes/No), the location of the correct button (Left or Right in the pair of response choices), and the button that was highlighted (Left/Right) were all counterbalanced evenly. With the different stock price possibilities, there were a total of 32 trials per block, and participants completed two blocks. On each trial, participants fixated on a crosshair target until a dialog box stimulus appeared. After the response, there was a reminder about the criteria for buying and selling the stock and a "Next Trial" button. Order of trials was randomized from the group of possible stimuli.

RESULTS

Traditional Metrics: Reaction Time and Accuracy

The dependent variables in this study were the traditional reaction time (RT) and accuracy measures, as well as the frequency and type of eye movements (discussed below). RT was the duration between the onset of the stimulus and the correct response from the participant. Incorrect responses were excluded from RT and eye movement analyses. Trials with RTs greater than three standard deviations from the grand mean were also discarded (approx 2% of trials overall) from RT and eye movement analyses. A 2 x 2 ANOVA with correct button location (left, right) and type of highlighting (consistent, inconsistent) was completed for both accuracy and RT.

In terms of accuracy, the grand mean was 85.5% ($SE = .009$). ANOVA results showed that accuracy (as measured by percent correct) did not depend on correct box placement, $F(1,399) = .020$, $p > .05$, or highlighting, $F(1,399) = .023$, $p > .05$, and there was no interaction between the two, $F(1,399) = .275$, $p > .05$. This indicates that accuracy did not vary systematically in any of the conditions and that there was no speed-accuracy tradeoff in participants' responses.

For RT, ANOVA results indicated that there was a main effect of correct button location on RT, $F(1,370) = 9.035$, $p < .05$, such that placing the correct button first (i.e., on the left) produced a significantly faster reaction time ($M= 2522$, $SE = 36.26$) than when placing the correct button last (i.e., on the right; $M = 2628$, $SE = 39.26$). There was no effect of highlighting on RT (i.e., whether the correct or incorrect button was highlighted), $F(1,370) = .662$, $p > .05$, nor was there any interaction between placement and highlighting for RT, $F(1,370) = .111$, $p > .05$.

Eye Movement Data

To analyze the eye tracker data, areas of interest were defined. Then for each trial the pattern of fixations as a function of time was categorized into a pattern of movement. There were several common eye movement patterns. On *straight-through* patterns participants fixated on each button only once. On *regressions* participants fixated on one button more times than on the other (e.g., left, right, then left again). A *double-check* meant fixating on the buttons an equal amount of times (e.g., left, right, left, right). A *skip* meant looking at only one button. On a given trial, exactly one eye movement pattern was determined. An arcsine transformation ($A = 2 \arcsin \sqrt{p}$) was applied to the proportions to stretch out the tails of the distribution, so that the unit of measurement was more linearly related to the other variables (Cohen and Cohen, 1983). A 2 (highlighting) x 2 (placement) repeated measures ANOVA was subsequently conducted with the transformed data, for both proportion of regressions and proportion of trials where the first button looked at was the left button.

A one-sample t-test with a test value of 0.5 indicated that the proportion of responses where the left button was looked at first (M=.612, SE=.044) was significantly greater than 0.5, $t(24) = 11.136$, $p < .05$. That is, participants looked at the left button first, more often than looking at the right button first. ANOVA results indicated a statistically significant effect of correct response placement on which response button the participant looked at first, $F(1,24) = 8.338$, $p < .05$, such that the left response button was looked at first more often when the left box was correct ($M = .673$, $SE = .045$) than when the right box was correct ($M = .551$, $SE = .051$). There was also a significant highlighting by placement interaction, $F(1,24) = 5.017$, $p < .05$. Tests of simple effects revealed that when highlighting was consistent, the left response button was looked at more often when it was correct than when the right response button was correct, $t(24) = 3.999$, $p <.05$; when the highlighting was inconsistent, however, there was not an effect of correct response button location, $t(24) = 1.030$, $p >.05$. There was no main effect of highlighting, nor was there an interaction between highlighting and correct response placement.

ANOVA results also indicated that there were significantly more regressions when the correct response button was placed on the left ($M = .106$, $SE = .017$) than when the correct response button was placed on the right ($M = .060$, $SE = .012$), $F(1,24) = 5.346$, $p < .05$. There was no main effect of highlighting and no interaction between highlighting and correct response placement.

DISCUSSION

Interpretation of Results

The analysis and interpretation of the RT and accuracy data is quite familiar and straightforward, if considered alone. The analysis of the eye movement data is somewhat more challenging, involving considerably more attrition due to data errors, tracking hardware problems, analysis software limitations, and other complications. It is not surprising that the use of eye tracking data in HCI is somewhat rare, given the challenges involved simply gathering and examining such data. However, once gathered, the data can contribute an interesting counterpoint to the “traditional” RT and accuracy data. It seems clear that both are important in understanding human behavior with computing systems.

To summarize the results, the RT and accuracy measures would suggest that in dialog boxes the “correct” button (i.e., the most likely choice, often “OK”) should be placed on the left side of the set of response buttons. This arrangement (typical of Windows dialog boxes) allows the user to read through the dialog box text, then read the first button label, and since it is the correct or desired option, click on that button immediately. This should (and does, here) lead to faster reaction times than if the “correct” button were located on the right. In that case, the user would have to read through the “incorrect’ button label before arriving at the “correct” button. It should be noted, however, that having the correct button placed in the left-most position only leads to a difference of about 100 ms in RT, on average.

Continuing with the summary of results, the consideration of eye movement provides information above and beyond traditional usability metrics, and suggests design recommendations contrary to RT and accuracy measures. First, users do tend to start by looking at the left-most button, then read left-to-right in the *read-flow* direction. This principle is supported by our finding that the left-most button will be seen first and that placing the correct button in the left-most position results in a faster reaction time. In contrast, however, the finding of more regressions when the “correct” button is on the left suggests that users prefer to be fully informed before they submit their response. This finding also suggests that users analyze the visual interface in a manner similar to reading. Specifically, they read the button options from left to right, passing over the “correct” button on the left in order to make sure they have read the label on the right-most button before backtracking (regressing) and clicking on the left button. That is, user behavior seems to support the principle of *informed submit* because they check the other possible responses before making a response decision. The time and eye movement that is required to read through all the options, then backtrack to the correct response, may be considered wasteful. Thus, the simpler, more straightforward design approach (at least in terms of eye movements, and presumably risk of confusion) might be to have the “correct” button located to the right, such that the user can read through all of the options, and upon arriving at and reading the final button, click on it without having to backtrack or move the eyes back to some previous button.

Implications

Regardless of the metrics providing divergent recommendations, it may be argued that the sizes of the effects on performance in the present study are quite small (e.g., 100 ms difference in RT), perhaps not “practically significant,” regardless of their statistical significance. In that case, the eye movement data may allow for a meaningful discrimination of interface approaches in cases where RT or accuracy might not produce a clear design recommendation.

Consideration of eye movements allows designers to employ measures of “usability” and “user-friendliness” that are quantitatively based, yet not limited to focusing on saving time on the order of milliseconds, which has questionable practical significance. It allows us to consider how people want their information arranged, how they choose to read and investigate interface elements as simple and ubiquitous as dialog boxes, and allows us to factor that in to our designs.

This study showed that deeper levels of behavior analysis bears new and interesting information. By showing that eye movements are distinct measures within the context of a visual interface, this study has shown that tracking eye movements is a rich source of information for usability research that should continue to be investigated. This study provides theoretical guidelines for future research and shows that researchers cannot assume that accuracy and reaction time give complete information.

Finally, there are some clear empirical answers provided by this research that should not be overlooked. First of all, the heuristic often used by designers that assumes that users interact with a graphical user interface in a similar manner to reading was supported. More specifically, the results here suggested a “directed” visual search that follows a reading pattern. The eye tracking may also help us to objectively infer other aspects of the user’s experience, such as the user’s satisfaction with the interface and mental effort. This is why eye movements have been considered an important aspect of the interaction between humans and machines since Fitts’ developed his widely used laws. Considering the success of his law in guiding contemporary design, the opportunities that his methods provide are promising.

Limitations

HCI research is often fraught with difficulties, for many reasons. In this study, there are limitations that need to be noted with respect to both the RT/accuracy measures and the eye movement data. In simple tasks such as this one, accuracy data must be interpreted with caution, due to the low task difficulty: the task may not have been hard enough to resolve differences in accuracy. Despite increased task difficulty, however, accuracy may remain largely at ceiling, given that users tend to figure out the correct action, even if it takes more time. Eye tracking data also needs to be considered carefully: Due to the binary nature of the response choices (Yes, No), the participant could look at the first button and know immediately if it was correct or not. If the choices on the buttons were potentially different on each dialog box instance (i.e., more closely resembling the situation in real human-computer interactions) we would expect even more of a need to read all the options before making a fully informed submit action. That is, in a more variable task situation, we would predict even more regressions if the “correct” button were always on the left. It remains to be seen how such a variable set of button options would affect RT. Thus, both RT and eye movement data would be key to consider.

It should also be said that the degree to which these results can be generalized to other situations depends heavily on whether the user is interacting with a dialog box that he or she is familiar with, or one that is novel. Most of the text in our stimuli remained the same across trials, with exception to the words “buy” or “sell” and the dollar value of the stock. This predictability may have influenced how the participants interacted with the dialog box, since they did not have to read through the whole line of text before making their response. Their looking strategy seems to have evolved over the course of their trials from a strict left-to-right, top-to-bottom reading pattern to a visual search process (looking specifically for the stock price immediately), although this remains anecdotal, and not statistical, at this point. In a natural setting, the dialog boxes would be likely to change every instance.

Another aspect of experience is a user’s familiarity with a given operating system. In the present case, this becomes another complicating factor, both in terms of user experience and in terms of simply analyzing and understanding the results. For example, nearly all of our participants were Windows users, so their familiarity with the “left-is-correct” layout of dialog boxes needs to be factored in. It remains to be seen what would happen in this experiment with a number of Macintosh users who are used to the opposite layout. Again, even if one or the other layout were shown to be clearly superior, it is not clear that making a change to an operating system that has been in place for years would be warranted. Regardless of the design that may be considered “best” (whatever that is defined to be), it is important to consider installed base and existing experience as a factor in determining whether or not to make any changes to any interface.

Final Conclusions

The results of this study were not really intended to give a simple or clear cut answer to the question of “which design is best?” Rather, given the methodological subject matter of this paper, it creates more questions than answers. The eye tracking data support read-flow and informed submit, and thus placing the correct button on the right. The more typical reaction time behavioral measure, however, suggests exactly the opposite. The important message of the present research, as we have noted, is that the use of the traditional measures can be done in conjunction with the more challenging but equally rewarding eye tracking data collection. Eye tracking means gathering more participants, due to its inefficiency in data collection. Typically 20 percent of participants eyes’ will not track at all (Goldberg & Wichansky, 2000). Furthermore, data that is collected requires a lot of post-hoc work, due to the data reduction that has to take place. In short, understanding eye-tracking data is much more time, effort, and money-intensive than many quantitative measures such as accuracy, response time, and mouse logging. But we contend, and the data presented here support the statement, that it is no longer sufficient to simply count the number correct, or determine the “faster” design. It has been said that the eyes are a window to the soul. For HCI, they may also be a window to the thoughts, preferences, and feelings of users, and in a user centered design environment they should be given their due consideration.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Paul Corballis for his expertise and lab resources, as well as the valued assistance of Nate Parks in the development and setup of this study.

REFERENCES

Bodart, F., Hennebert, A.M., Leheureux, J.M., & Vanderdonckt, J. 1994. Towards a dynamic strategy for computer-aided visual placement. *Proceedings of InterChi,* 78-87.

Galitz, A.B. 1989. *Handbook of Screen Format Design.*Wellesley, MA: QED Information Sciences, Inc.

Goldberg, J.H., & Wichansky, A.M. (1993). Eye tracking in usability evaluation: A practitioner’s guide. *In The Mind’s Eye: Cognitive and Applied Aspects of Eye Movement Research,* 493-516.

Jacob, R.K., & Karn, K.S. (1993). Eye Tracking in Human-Computer Interaction and Usability Research: Ready to deliver the promises. *In The Mind’s Eye: Cognitive and Applied Aspects of Eye Movement Research,* 574-605.

Kim, W.C., & Foley, J.D. (1993). Providing high-level control and expert assistance in the user interface presentations design. *Proceedings of InterChi,* 430-437.

Mullet, A.B. & Sano, A.B. 1995. *Designing Visual Interfaces*. Mountain View, CA: Sunsoft Press.

Cohen, J. & Cohen, P. 1983. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* Mahwah, NJ: Lawrence Erlbaum Associates.