Ephemeral Adaptation: The Use of Gradual Onset to Improve Menu Selection Performance

Leah Findlater[*], Karyn Moffatt[*], Joanna McGrenere, Jessica Dawson
Department of Computer Science
University of British Columbia, Vancouver, Canada
{lkf, kmoffatt, joanna}@cs.ubc.ca

ABSTRACT
We introduce ephemeral adaptation, a new adaptive GUI technique that improves performance by reducing visual search time while maintaining spatial consistency. Ephemeral adaptive interfaces employ gradual onset to draw the user's attention to predicted items: adaptively predicted items appear abruptly when the menu is opened, but non-predicted items fade in gradually. To demonstrate the benefit of ephemeral adaptation we conducted two experiments with a total of 48 users to show: (1) that ephemeral adaptive menus are faster than static menus when accuracy is high, and are not significantly slower when it is low and (2) that ephemeral adaptive menus are also faster than adaptive highlighting. While we focused on user-adaptive GUIs, ephemeral adaptation should be applicable to a broad range of visually complex tasks.

Author Keywords
Adaptive interfaces, personalization, abrupt visual onset, menu design, user study, interaction techniques.

ACM Classification Keywords
H.5.2 [User Interfaces]: Evaluation/methodology, interaction styles.

INTRODUCTION
Adaptive graphical user interfaces (GUIs) automatically tailor features to better suit the individual user's needs. To date, these interfaces have tended to rely on one of two forms of adaptation: spatial or graphical. Spatial techniques reorganize items to reduce navigation time and, to a lesser degree, to aid visual search [2,12,14]. An adaptive split menu, for example, moves or copies the most frequently and/or recently used items to the top of the menu for easier access [14]. Graphical techniques, on the other hand, reduce visual search time, for example, through changing the background colour of predicted items [7,8,18]. Some techniques use a combination of both spatial and graphical elements [18,19].

As an alternative to spatial and graphical adaptation, we propose the use of a temporal dimension and introduce

CHI 2009, April 4–9, 2009, Boston, Massachusetts, USA.

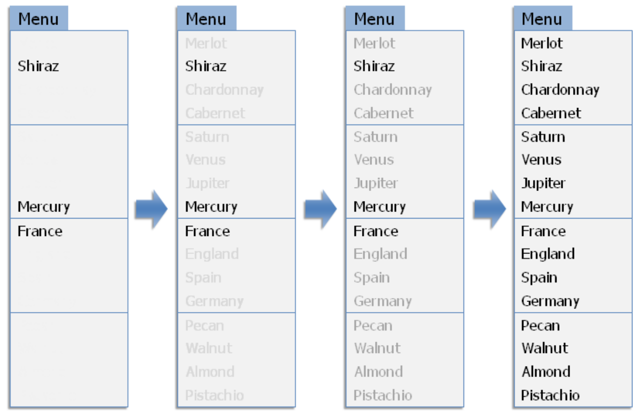

Figure 1. Ephemeral adaptation applied to menus: predicted items appear immediately, while remaining items gradually fade in.

ephemeral adaptation as a new adaptive interaction technique that uses this dimension to reduce visual search time. Ephemeral adaptive interfaces use a combination of abrupt and gradual onset to provide initial adaptive support, which then gradually fades away. The goal is to draw the user's attention to a subset of adaptively predicted items, in turn reducing visual search time. Figure 1 applies ephemeral adaptation to a menu: adaptively predicted items appear abruptly when the menu is opened, after which the remaining items gradually fade in.

Ephemeral adaptation maintains spatial consistency, thus addressing one of the main drawbacks of spatial adaptation techniques [2]. An adaptive menu that reorganizes features, for example, by promoting the most frequently used ones, offers theoretical performance benefits over a traditional static menu. In practice, however, spatially adaptive interfaces are not often faster than their static counterparts because the user needs to constantly adapt to the altered layout, wiping out any potential gains [2,4,5,12]. Successes have tended to occur only when the adaptive approach greatly reduces the number of steps to reach desired functionality, for example, through a hierarchical menu structure [8,10,19], or when limited screen real estate necessitates scrolling [5].

Similarly to ephemeral adaptation, graphical techniques also maintain spatial consistency and focus on reducing visual search. Several researchers have proposed techniques to highlight predicted items with a different background

[*] The first two authors are equal contributors to this work.

colour [7,8,18] but no performance results comparing colour highlighting to a static control have been reported. While ephemeral and graphical adaptation are similar in that both aim chiefly to reduce visual search time, there is some evidence in the human perception literature that abrupt onset may be a stronger attention cue than colour [16]. This suggests ephemeral adaptation may provide a performance benefit over highlighting.

The primary contribution of this paper is the introduction of ephemeral adaptation as a technique that adapts the interface along a previously unexplored temporal dimension. To demonstrate the benefit of ephemeral adaptation, we applied the technique to pull-down menus and conducted two controlled lab studies with a total of 48 users. Our results show that when the accuracy with which the adaptive algorithm predicts the user's needs is high (79%), ephemeral adaptation offers performance and user satisfaction benefits over traditional static menus and a performance benefit over an adaptive highlighting technique (based on [8,18]). Moreover, there is little overall cost to using ephemeral adaptation when adaptive accuracy is low, since ephemeral adaptive menus were not significantly slower than static menus at low accuracy (50%). We also show that adaptive highlighting is not a promising approach for improving performance: although subjective response was positive, highlighting was not found to be faster than static menus even at a high level of adaptive accuracy.

Our results show that ephemeral adaptation is a viable interaction technique to improve visual search time in complex interfaces. While our focus has been on user-adaptive GUIs, the use of the temporal dimension for abrupt and gradual onset should be applicable to a broader range of applications, including guiding attention within visually complex web pages and for information visualization tasks.

RELATED WORK ON ADAPTIVE INTERFACES

Adaptive interfaces have appeared both in commercial applications and in research prototypes. The first level of the Microsoft Windows XP Start Menu, for example, provides a small set of both user-specified ("pinned") and adaptively chosen programs. Early research showed promise for adaptive interfaces in restructuring menu-based access to an adaptive telephone directory [10]. More recent research on adaptive GUIs, however, has been less successful [2,4,5,12].

One technique which has received a lot of attention is the adaptive split interface, where items are split into separate static and adaptive sections: those items predicted to be the most useful are moved or copied to the adaptive section [8]. Apart from the original split menu approach, the evaluation of which used a menu layout that was predetermined before use (rather than adapting during use) [14], there has been no evidence to show that single-level pull-down split menus are faster or preferred to static menus for desktop applications [2,4,5], and in some cases (e.g., when the accuracy of adaptive predictions is low) they can even be slower [5]. They have, however, found success in a few specific contexts. For example, Gajos et al. showed that an adaptive split toolbar that copies top-level items and promotes submenu items to the top level is faster and preferred to a static toolbar [8]. Here, the adaptation reduced the number of mouse clicks needed to reach the target item. In another example, Findlater and McGrenere showed that adaptive split menus can provide a performance benefit over static menus for a small screen interface, where they significantly reduced the scrolling needed to find items [5]. Commercially, split interfaces have been used successfully for font menus, where the large number of items necessitates scrolling.

As an alternate technique, several researchers have included adaptive highlighting in their designs. Tsandilas and schraefel compared highlighting of items in a traditional menu to highlighting in a fisheye menu, but since highlighting was included in both we cannot draw conclusions about how it would compare to a static menu [18]. Gajos et al. subsequently applied adaptive colour highlighting to a graphing calculator [7] and a toolbar [8]. Performance results were not reported in either case but user feedback was negative. One difference between these interfaces and an adaptive menu is that changes are immediately visible (rather than only being visible upon opening the menu); users reported the change to be disorienting [7]. Finally, Tsandilas and schraefel used colour highlighting in bubbling menus, a technique which combines both spatial and graphical elements and shows some promise for improving selection of deeply embedded items in a cascading menu [19]. Since highlighting was only one aspect of the technique it is not possible to draw conclusions about highlighting alone.

Recently, researchers have begun to isolate characteristics which impact the effectiveness of adaptive GUIs. Spatial consistency is one important aspect: copying rather than moving items from the static section to the adaptive section of a split interface increases spatial consistency and results in higher user satisfaction [7]. Similarly, the frequency with which the set of predicted items changes (stability) may impact performance [2]. Another important factor is adaptive accuracy, the accuracy with which the adaptive algorithm can predict the user's needs. Higher adaptive accuracy results in faster performance and higher user satisfaction with both split menus and toolbars [5,8,9] and highlighting approaches [18]. Predictability of the adaptive algorithm from the user's perspective has also been shown to impact user satisfaction [9]. Finally, a cost/benefit trade-off of adaptivity has been shown to be important: factors such as screen size [5] play a role in determining the usefulness of adaptation.

EPHEMERAL ADAPTATION

Ephemeral adaptive menus are designed to reduce selection time by guiding the user's attention to adaptively predicted menu items through a combination of abrupt and gradual onset. In contrast to the spatial and graphical techniques described in the previous section, our goal was to design an adaptive mechanism that utilizes a temporal dimension. The adaptation is thus ephemeral and not as intrusive as many adaptive techniques: adaptive support is provided initially but then fades away, returning the interface to normal. This maintains spatial consistency of user interface elements and should reduce visual search time.

Abrupt Onset and Potential Benefit for Adaptive GUIs

Yantis and Jonides [20] demonstrated that an item with an abrupt onset (sudden appearance) is visually processed first among a set of items, even in the absence of an explicit attention set (i.e., the subject has not been told to interpret the stimulus in any specific way). This behaviour results in fast identification of abrupt onset stimuli compared to stimuli without an abrupt onset. In addition, abrupt onset is fairly unique in this regard [11]. Colour can also capture attention, but only when the subject has been instructed to attend to it [6]. Even when such instruction is given, abrupt onset may still be better than colour at drawing attention [16], but this might depend on the particular colour used. Finally, the attention-capturing behavior of abrupt onset can occur below the threshold of subjective awareness [13], suggesting that abrupt onset can be used unobtrusively.

Based on this discussion we predicted that abrupt onset will provide stronger adaptive support than graphical methods for visually distinguishing items, such as background colour highlighting. Highlighting should require the user to actively adopt a strategy of looking for the highlighted elements, whereas abrupt onset in ephemeral adaptation should draw users automatically.

Even though abrupt onset has the ability to draw attention automatically, research has shown that the response is not involuntary: people can override it if motivated to do so [15, 21]. Thus, if the user knows the abrupt onset stimulus is irrelevant or false, is looking for a different type of stimulus, or knows the location of their target, then an abrupt onset will not distract them. This suggests that using abrupt onset for adaptive predictions should not force the user to give priority to the adaptively predicted items. Thus, when predictive accuracy is low, or the user already knows where the target item is located, the user should not find the ephemeral adaptation approach distracting.

Pilot Testing of Early Designs

We initially tested a design that used two abrupt onsets: adaptively predicted items appeared immediately when the user opened a menu, followed by the abrupt appearance of the non-predicted items after a short onset delay (we piloted onset delays of 25ms, 50ms and 100ms). However, piloting with four participants (using similar methodology to Study 1, described below) was not encouraging. Preliminary analysis of the performance data gave no indication that the adaptive technique reduced selection time, and may even have increased it for some onset delays. Moreover, 3 out of 4 participants preferred the static control condition.

One possible explanation for this result with the shorter delays (25ms and 50ms) is that users did not have enough time to respond to the predicted items before the appearance of the non-predicted items. In contrast, with the longer delay (100ms) users may have too often been left waiting for the non-predicted items to appear. Variability between trials and participants may have meant that no delay length offered an adequate compromise.

Final Technique

Non-predicted items faded in gradually over a delay period (as depicted in Figure 1). The onset delay for this technique is the elapsed time from opening the menu until all items reach their final foreground colour. The non-predicted items begin as the same colour as the background of the menu (a light grey), and then darken through a series of 10 linear increments until they are the same colour as the predicted items; this gradual appearance is visually smooth for the onset delays we used (250ms to 1000ms).

The rationale for this approach is that, unlike abrupt onset, gradual onset does not draw attention [20]. Moreover, because the non-predicted items become legible after only 1 or 2 darkness increments, but the predicted items remain visually prominent until the last 1 or 2 increments, this approach leaves a wider window for the interaction and should allow for more variability among user abilities.

To evaluate this design, we conducted two controlled lab studies. In Study 1, we tuned the onset delay, and as a first step, compared ephemeral adaptation to a static menu. In Study 2, we looked more closely at the performance benefits of ephemeral adaptation by comparing it to a colour highlighting approach. The final design, used in Study 2, has an onset delay of 500ms.

STUDY 1: INTRODUCING EPHEMERAL ADAPTATION

The first goal of Study 1 was to determine whether ephemeral adaptation can offer a performance benefit over static menus for a basic selection task. This benefit should be seen when adaptive accuracy is high, and due to the spatial consistency of the menus, there should not be a significant performance hit when accuracy is low. We used pull-down menus since adaptive approaches have been extensively applied to them, facilitating comparison to previous research.

A second, though equally important, goal of this study was to explore different onset delays. Previous research has suggested that 200–300ms should be sufficient to prevent the capture of attention caused by abrupt onsets [1,17]. However, the task used in that work was quite different from ours (subjects only needed to detect the presence of a stimulus), suggesting a longer delay may be more appropriate for a selection task. Thus, we examined a range of onset delays starting from 250ms, namely 250ms, 500ms, and 1000ms. Early pilot participant feedback for

1000ms was that the delay was too long, so we only looked at the 250ms and 500ms onset delays in Study 1.

Methodology

Menu Conditions

The three menu types we tested were:

1. *Control*: Traditional static menu.
2. *Short-Onset*: Ephemeral adaptive menu, where non-predicted items gradually appear over a 250ms delay.
3. *Long-Onset*: Ephemeral adaptive menu, where non-predicted items gradually appear over a 500ms delay.

The static control condition (Control) consisted of 3 traditional pull-down menus with 16 items in each menu. Items were separated into groups of 4 semantically related items (e.g., Merlot, Shiraz, Chardonnay, Cabernet). The adaptive conditions were identical to Control, except for the delayed onset of non-predicted items. Menu contents were randomly generated for each participant and condition.

Both adaptive menu conditions used the same adaptive algorithm to predict a set of 3 items that were likely to be selected next by the user; only the onset delay differed. A set size of 3 has been used previously in adaptive split menu research [2,5,14] and is the same number of predictions that are highlighted with similar-length bubbling menus [19].

Adaptive Accuracy Conditions

Adaptive accuracy is the percent of trials where the item the user needs to select is included in the set of predicted items for that trial. We used two levels of adaptive accuracy:

1. *Low*: 50% accuracy, on average.
2. *High*. 79% accuracy, on average.

To achieve two different levels of adaptive accuracy, we follow the two-step process used by Findlater and McGrenere [5]. First, for each participant we randomly generated a selection sequence (see the Task section for more detail) and applied an adaptive algorithm to predict a set of 3 probable items at each selection in the sequence; this algorithm calculated predictions based on the items that had been recently and frequently selected and resulted in prediction accuracy of 64.5% on average for all participants. Second, for Low accuracy we randomly adjusted 18 trials so that they were no longer correct, and for High accuracy we randomly adjusted the same number of incorrect predictions to be correct. This resulted in the accuracy conditions listed above.

These accuracy levels are in line with previous research in the area [5,8,18,19]. There is no definitive answer in the research community on achievable levels of accuracy for these types of selection tasks, but limited real world data suggests that a predicted item set size of 3 can result in accuracy levels near 90% for Microsoft Word [4].

Task

The experimental task was a sequence of menu selections from an experimental system. A prompt across the top of the screen displayed the name of the item to be selected and the menu in which it was located. Three menus were positioned just below the prompt. Once the participant had correctly selected the target item, the prompt for the next trial would be displayed.

To mitigate the effect of an individual selection sequence, the same underlying sequence was used for all conditions and task blocks for a given participant, but the location of the menus was permuted for each condition to reduce learning across conditions. For example, if the first selection in the first condition was Menu 1, Item 3, then in the second condition it would be Item 3 of either Menu 2 or Menu 3. The underlying selection sequences were then masked with different menu item names in each task block and condition. Each menu was generated by randomly selecting 4 groups of 4 semantically related items from a set of 72 such groups, so that each group appeared only once.

To generate a task selection sequence we followed the approach taken by Cockburn, Gutwin and Greenberg [2]: a Zipf distribution (Zipfian R^2=.99) over 8 randomly chosen items from each menu was used. Thus, within a menu the relative selection frequencies of items per block were 15, 8, 5, 4, 3, 3, 2, 2, and the final selection sequence consisted of 126 randomly ordered selections per block. Each participant completed two different task blocks per condition.

Quantitative and Qualitative Measures

Speed was measured using the median selection time, calculated as the time from opening the menu to selecting the correct item. The median was used to reduce the influence of outlier trials. We used an implicit error penalty in the speed measures; that is, participants could not advance to the next trial until they correctly completed the current trial. For completeness, we also recorded the error rate. Finally, subjective data was collected using 7-point Likert scales on difficulty, satisfaction, efficiency and frustration. At the end of the study, a questionnaire asked for comparative rankings of the menu conditions.

Apparatus

A 2.0GHz Pentium M laptop with 1.5 GB of RAM and Microsoft Windows XP was used for the experiment. The system was connected to an 18” LCD monitor with 1280x1024 resolution and the experiment was coded in Java 1.5. The system recorded all timing and error data.

Participants

Twenty-four participants (12 females) were recruited through on-campus advertising. All were regular computer users, were between the ages of 19–45 (M = 25.5) and were reimbursed $10 per hour to participate.

Design

A 2-factor mixed design was used: adaptive accuracy (Low or High) was a between-subjects factor and menu type

(Control, Short-Onset or Long-Onset) was a within-subjects factor. Order of presentation was fully counterbalanced and participants were randomly assigned to conditions.

Procedure

The procedure was designed to fit into a single 1-hour session. Participants were first given a background questionnaire to collect demographic information. Then, for each condition participants completed a short 8-trial practice block of selections to familiarize themselves with the behavior of the menus before completing two longer 126-trial task blocks. Short breaks were given in the middle of each block and between blocks. After both task blocks, participants completed a questionnaire with the subjective Likert scale questions for that condition. Once all experimental tasks were complete, a comparative questionnaire was given.

Before each adaptive menu condition, participants were given a brief description of the adaptive behavior: they were told that some of the items would appear sooner than others, and that these were the items the system predicted would be most likely needed by the user. However, participants were not told the level of prediction accuracy.

Hypotheses

Given the exploratory nature of Study 1, we did not make any formal hypotheses for the relationship between Long- and Short-Onset. We did, however, expect an effect of onset delay on performance and preference, and possibly an interaction between onset delay and accuracy. Establishing these effects was one of our main goals for this study.

Since the goal of our study design was to compare the menu types to each other, Control was included within each level of accuracy. Thus, accuracy was not fully isolated in the design and we did not hypothesize a main effect of accuracy on performance.

H1. Speed.

1. *For High accuracy: at least one of Long-Onset or Short-Onset will be faster than Control*. No formal hypotheses for Long- versus Short-Onset.
2. *For Low accuracy: both Long-Onset and Short-Onset will be no worse than Control*. Ephemeral adaptation maintains spatial consistency of the menu items, thus we predict that performance should not be significantly hindered when accuracy is low.

H2. User Preference.

1. *For High accuracy: at least one of Long-Onset or Short-Onset will be preferred to Control.* Corresponds to the speed hypothesis.
2. *For Low accuracy: Control will not be preferred to either Long-Onset or Short-Onset*. Corresponds to the speed hypothesis.

Results

We ran a 2x3x6 (accuracy x menu x presentation order) repeated measures ANOVA on the dependent variable selection time. As expected, there were no significant main or interaction effects of order, so we omit these results from the presentation. All pairwise comparisons were protected against Type I error using a Bonferroni adjustment. We report on measures that were significant ($p < .05$) or represent a possible trend ($p < .10$). Along with statistical significance, we report partial eta-squared (η^2), a measure of effect size. To interpret this value, .01 is a small effect size, .06 is medium, and .14 is large [3]. Not including break times, the experimental tasks for each condition took on average 10.5 minutes to complete (SD = 1.5).

One participant was removed from the analysis. This participant's behavior clearly showed that he was not following our instructions to go as quickly and as accurately as possible, and his comments at the end of the study indicated that he was confused about the task, particularly in the control condition. Performance-wise, he was more than 2 standard deviations away from the mean on the sum of selection times for all task blocks, and 23% slower than the next slowest participant in his condition (high accuracy). We report on data from 23 participants.

Overall Speed

Selection times are shown in Figure 2. Speed was impacted by menu type (main effect: $F_{2,22} = 3.80$, $p = .038$, $\eta^2 = .257$) and by the combination of menu and adaptive accuracy (interaction effect: $F_{2,22} = 3.73$, $p = .040$, $\eta^2 = .253$). As expected there was no significant main effect of accuracy.

At High accuracy, Long-Onset was fastest, and at Low accuracy, ephemeral was not slower. Pairwise comparisons on the interaction effect showed that at High accuracy, Long-Onset was faster than both Short-Onset (p = .018) and Control (p = .047). No significant difference was found between Short-Onset and Control (p = .854). For Low accuracy no differences were found between the three menu types (p = 1.000 for all comparisons).

Speed of Selecting Predicted and Non-Predicted Items

While our main measure was overall selection time because it encompasses both the benefit and cost of using ephemeral adaptation, we performed two separate analyses (2x3x6 RM ANOVAs) to better understand this cost/benefit breakdown:

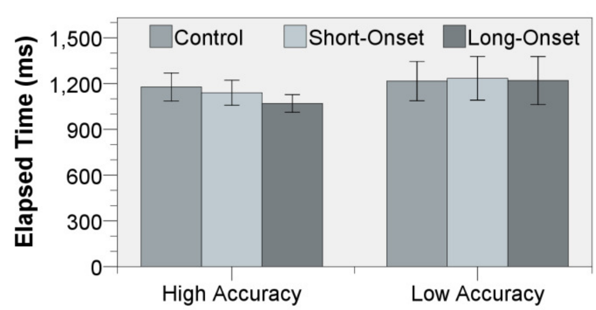

Figure 2. Average selection time per trial for Study 1 (N = 23). Error bars show 95% confidence intervals (CI).

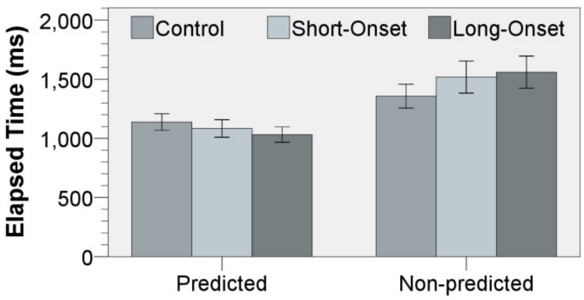

Figure 3. Study 1 average selection times on predicted and non-predicted trials, collapsed across accuracy level (N = 23). Error bars show 95% CI.

(1) speed for trials that were correctly predicted; (2) speed for trials that were not correctly predicted. There were no adaptive predictions in the Control condition, but since each participant's underlying selection stream was the same for every condition, the corresponding Control trials can be compared to the Short-Onset and Long-Onset trials.

Note that we would expect selection times for the non-predicted trials to be longer than for the predicted trials even in the Control condition. This is because the adaptive predictions are based on recently and frequently selected items; thus, non-predicted items are typically those items with which the user is least familiar, and, correspondingly, should be slower selecting. Results are shown in Figure 3.

Long-Onset was fastest for predicted trials. For predicted trials there was a significant main effect of menu on speed ($F_{2,22} = 18.5$, $p < .001$, $\eta^2 = .627$). As expected, based on the overall results, pairwise comparisons showed that Long-Onset was faster than both Short-Onset (p = .004) and Control (p = .001). A trend also suggests that Short-Onset was faster than Control (p = .062). No significant main or interaction effects of accuracy were found.

Control was fastest for non-predicted trials, suggesting a cost for ephemeral adaptation when items are not correctly predicted. For non-predicted trials there was also a significant main effect of menu on speed ($F_{1.35,14.9} = 10.1$, $p = .001$, $\eta^2 = .479$; Greenhouse-Geisser adjusted). Pairwise comparisons showed that Control was faster than Long-Onset (p = .008) and Short-Onset (p = .047) for non-predicted trials. No significant difference was found between Long- and Short-Onset (p = .495).

Errors

The speed measure included an implicit error penalty but we report on error rates for completeness. Error rates ranged from 1.9%–2.5% of trials on average for participants in the High accuracy conditions and from 0.5%–1.0% on average for the Low accuracy conditions.

Subjective Findings

When asked to rank the menu types based on overall preference, 10 out of 11 High accuracy participants and 9 out of 12 Low accuracy participants chose one of the adaptive conditions. For High accuracy, preference was skewed towards preferring Long-Onset over Short-Onset (7 vs. 3 participants). In contrast, preference was more evenly split in the Low accuracy condition (4 and 5 participants for Long-Onset and Short-Onset, respectively).

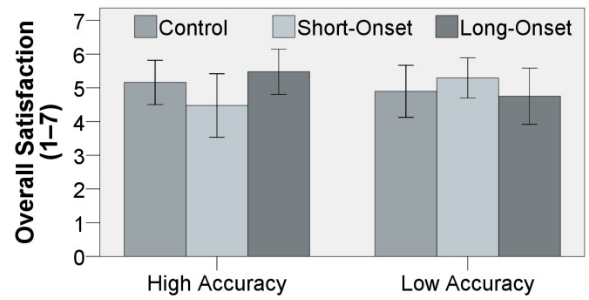

Figure 4. Satisfaction ratings for Study 1 (N=23). Higher values indicate higher satisfaction. Error bars show 95% CI.

A single *overall satisfaction* measure (shown in Figure 4) was constructed by collapsing the four Likert scale questions asked for all conditions. A reliability analysis showed that they were likely measuring the same construct (Cronbach's alpha = .852). Friedman tests within accuracy levels showed no significant impact of menu on overall satisfaction. This could be due to low statistical power. Interestingly, in the High accuracy condition the mean rating for Short-Onset was lowest, whereas in the Low accuracy condition it was highest.

Summary and Discussion

We summarize our results according to our hypotheses:

H1. Speed.

1. *For High accuracy: at least one of Long-Onset or Short-Onset will be faster than Control.* Supported. Long-Onset was faster than Control, but no difference was found between Short-Onset and Control.
2. *For Low accuracy: both Long-Onset and Short-Onset will be no worse than Control.* Supported. No differences were found in overall speed.

H2. User Preference.

1. *For High accuracy: at least one of Long-Onset or Short-Onset will be preferred to Control.* Although overall satisfaction results were unclear, preference rankings suggest a preference for ephemeral adaptation, which further investigation would need to confirm.
2. *For Low accuracy: Control will not be preferred to either Long-Onset or Short-Onset.* Though there was no indication that Control was preferred (supporting our hypothesis), the lack of clear preference results overall suggests this too should be examined further.

Although satisfaction was split between the two ephemeral adaptation approaches, only Long-Onset showed a performance benefit: it was faster than Control at High accuracy and was not slower at Low accuracy. The cost/benefit breakdown of the predicted and non-predicted selections shows that, not unexpectedly, there is a cost to using ephemeral adaptation when items are incorrectly predicted. That this cost did not result in a significant negative impact at Low accuracy suggests it is relatively small in contrast to approaches that do not maintain spatial consistency, such as adaptive split menus [5]. However, only for Long-Onset at High accuracy was the benefit for predicted items large enough to provide an overall gain.

STUDY 2: EPHEMERAL ADAPTATION VERSUS ADAPTIVE HIGHLIGHTING

Study 2 extends the results from Study 1, comparing the best onset delay condition from that study to an adaptive highlighting approach. We chose highlighting as an appropriate comparison because, like ephemeral adaptation, it maintains the spatial layout of GUI elements and provides only a visual change. A secondary goal for Study 2 was to evaluate the performance of adaptive highlighting. Although adaptive highlighting has been previously studied in the context of different levels of accuracy, it has not been compared effectively to a control condition.

Methodology

Study 2 used the same methodology as Study 1 with the exception that Highlight replaced the Short-Onset condition, and to increase the likelihood of finding differences between the menu conditions, we examined only one level of adaptive accuracy. We chose the higher accuracy level from Study 1 because at lower accuracy there was no difference between ephemeral adaptive menus and the static control. Using a high level of accuracy also increased the likelihood of finding a benefit for adaptive highlighting (see Discussion section for more detail). The following sections describe the impact of these differences.

Menu Conditions

Study 2 compared three menu types, where both the adaptive menus (Ephemeral and Highlight) used the High accuracy adaptive condition from the previous study:

1. *Control*. The same as Control in Study 1.
2. *Ephemeral*. The 500ms onset delay (Long-Onset) condition from Study 1.
3. *Highlight*. Shown in Figure 5, Highlight emphasizes predicted items by changing the background colour to light purple (the same colour as used in [18]). It uses the same adaptive algorithm as Ephemeral.

Participants, Measures and Design

For Study 2, we recruited 24 new participants (10 females). All were regular computer users between the ages 19–33 (M = 25.3). A single-factor design was used with menu type (Control, Ephemeral or Highlight) as the within-subjects factor. Order of presentation was fully counterbalanced and participants were randomly assigned to an order. We collected subjective data using the same questionnaires as in Study 1, plus, for the adaptive conditions, 7-point Likert scales on distinctiveness, helpfulness, and distraction.

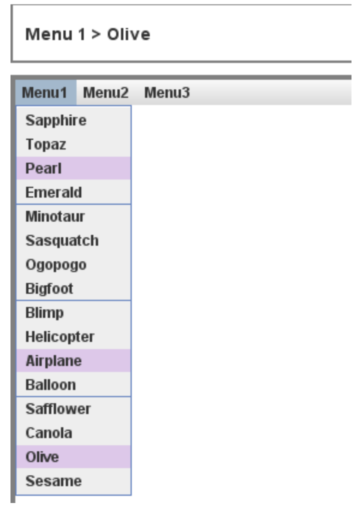

Figure 5. Our experimental interface showing the Highlight technique: predicted items have a light purple background.

Hypotheses

H1. Speed.

1. *Ephemeral will be faster than Control*. This hypothesis is based on our results from Study 1.
2. *Ephemeral will be faster than Highlight*. Abrupt onset has been shown to be a stronger cue than colour [11,16]. Although this was shown in a different context, we predict the relationship will also hold here.
3. *Highlight will be faster than Control*. Previous research has not provided definitive results, but has suggested that colour highlighting should offer a performance advantage [7,18].

H2. Satisfaction/Preference.

1. *Ephemeral will be preferred to Control*. Although there were no statistically significant results in Study 1, the descriptive statistics suggested that with a larger sample we may see this result.
2. *Control will be preferred to Highlight*. This hypothesis is based on previous findings [7].
3. *Ephemeral will be preferred to Highlight*. Based on the above two hypotheses, Ephemeral should also be preferred to Highlight.

Results

We ran a 3x6 (menu x presentation order) repeated measures ANOVA on selection time. As with Study 1, there were no significant main or interaction effects of order. A Bonferroni adjustment was used on all pairwise comparisons. We report on measures that were significant ($p < .05$) or represent a possible trend ($p < .10$). Not including breaks, the experimental tasks for each condition took on average 10.8 minutes to complete (SD = 1.8).

Speed

Ephemeral was the fastest menu type overall. There was a significant main effect of menu on speed ($F_{2,36} = 13.4$, $p < .001$, $\eta^2 = .427$). Results are shown in Figure 6 (Overall).

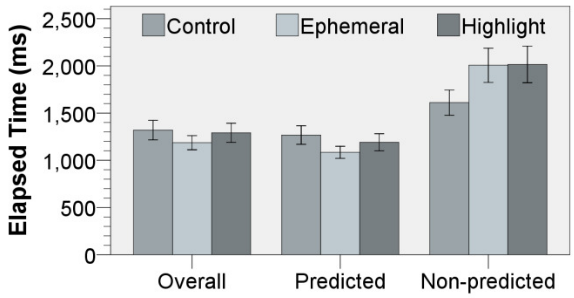

Figure 6. Study 2 selection times overall, and for predicted and non-predicted trials. Error bars show 95% CI (N = 24).

Pairwise comparisons showed that Ephemeral was significantly faster than both Control ($p = .001$) and Highlight ($p = .004$). No significant difference was found between Control and Highlight ($p = .581$).

Speed of Selecting Predicted and Non-Predicted Items

As with Study 1, we performed a secondary analysis, breaking down selections into those that were adaptively predicted and those that were not. Figure 6 shows the selection times for predicted and non-predicted trials. There was a significant main effect of menu on speed for both predicted ($F_{2,36} = 21.7$, $p < .001$, $\eta^2 = .547$) and non-predicted $F_{2,36} = 25.7$, $p < .001$, $\eta^2 = .588$) trials.

Ephemeral was the fastest menu type for predicted trials. Pairwise comparisons showed that Ephemeral was faster than both Control and Highlight for correctly predicted selections ($p < .001$ for both). Highlight was also faster than Control ($p = .043$).

Control was the fastest for non-predicted trials. Pairwise comparisons showed that Control was faster than both Ephemeral ($p < .001$) and Highlight ($p < .001$) when the predictions were incorrect. No difference was found between Ephemeral and Highlight ($p = 1.000$).

Errors

Error rates were uniformly low, at 2.0%, 2.2% and 2.3% on average for Ephemeral, Highlight and Control, respectively. Errors were indirectly accounted for in our speed measure.

Subjective Findings

Highlight was preferred overall by 12 participants, while 8 preferred Ephemeral and only 4 preferred Static. Most common reasons cited for preferring one of the adaptive conditions included making the task easier or faster.

Ephemeral and Highlight were more satisfying to use than Control. We calculated an overall satisfaction measure similarly to Study 1 (Cronbach's alpha = .837). Ratings were from 1 to 7, where 7 indicated strong positive agreement. A Friedman test showed there was a significant main effect of interface on satisfaction ($\chi^2_{(2,N=24)} = 12.9$, $p = .002$). To understand the source of this effect, we performed pairwise comparisons using Wilcoxon Signed Ranks Tests and applied a Bonferroni adjustment. The average satisfaction rating for Control was 4.3 (out of 7), which was less than Ephemeral's average rating of 5.0 ($z = -2.63$, $p = .027$) and Highlight's average rating of 5.0 ($z = -2.67$, $p = .024$). No significant difference was found between Ephemeral and Highlight ($p = .871$).

For the two adaptive conditions we also asked participants three additional Likert scale questions. While no statistically significant differences were found (using Wilcoxon Signed Ranks Tests), the ratings were overall positive. Participants felt that both the Ephemeral and Highlight adaptive behaviour helped them distinguish predicted items ($M = 5.5$, $SD= 1.4$) and helped them find items more quickly ($M = 5.1$, $SD = 1.4$). Also, participants responded neutrally to whether or not the adaptive behaviour was distracting ($M =3.6$, $SD = 1.5$).

Summary

We summarize our results according to our hypotheses.

H1. Speed.

1. *Ephemeral will be faster than Control*. Supported.
2. *Ephemeral will be faster than Highlight*. Supported.
3. *Highlight will be faster than Control*. Not supported. No difference was detected between Highlight and Control for overall performance.

H2. Satisfaction/Preference.

1. *Ephemeral will be preferred to Control*. Supported.
2. *Control will be preferred to Highlight*. Not supported. Contrary to previous results, Highlight was preferred to Control.
3. *Ephemeral will be preferred to Highlight*. Not supported. While more participants preferred Highlight over Ephemeral, no significant differences on overall satisfaction were found between the two conditions.

DISCUSSION

Our ephemeral adaptation approach, which employs temporal adaptation, shows promise in terms of both performance and user satisfaction. Results show that when predictive accuracy is high (79%), ephemeral adaptation helps users to find adaptively predicted menu items faster than either a static control condition or an adaptive highlighting condition. Another encouraging finding is that, in contrast to previous research on adaptive split menus [5], the ephemeral conditions did not perform worse than the static control condition when predictive accuracy was low (50%). This suggests that the consistent spatial layout provided by ephemeral adaptation allows its adaptive support to degrade more gracefully with lower accuracy than a split menu. Moreover, users were receptive to the ephemeral adaptive menu and rated it more highly than the static menu. Although further research is required to refine the technique, these combined results suggest that ephemeral adaptation is a viable option for distinguishing adaptive predictions in a visually complex interface.

The adaptive highlighting technique we tested did not show a performance benefit over static menus even though predictive accuracy was high; that is, the small advantage on correctly predicted trials did not translate to a significant overall improvement. However, Highlight was rated by users as comparable to the ephemeral adaptive menu and, in contrast to previous research [7,8], better than the static menus. One possible explanation is that our implementation was simply more subtle than Gajos et al.'s highlighting techniques (both [7] and [8] used brighter colours). Another possibility is that since Gajos et al.'s evaluations used more complex applications, the highlighting would have competed with other uses of colour in the interface and may have been perceived as more distracting than in our experimental interface. This would suggest that it may be difficult to design visually attractive real world interfaces that use highlighting, and even harder to add colour highlighting to an existing interface.

Any technique that vies for visual attention, including both adaptive highlighting and ephemeral adaptation, will need to compete with other visual elements in the interface. This underscores the need to explore the effectiveness of ephemeral adaptation within the context of a real application where other visual elements, such as animation, may detract from the efficiency and satisfaction benefits seen in our experiment.

In Study 2 we only considered one level of accuracy; thus, we do not have a complete understanding of how colour highlighting compares to ephemeral adaptive menus and static menus. It has, however, already been established for colour highlighting that performance worsens when adaptive accuracy drops [18]. Thus, we would expect that the highlighting technique would at best perform comparably to the static condition when adaptive accuracy is low, and possibly would perform worse. Since the ephemeral adaptive menus were not found to be slower than static menus at low accuracy in Study 1, they likely would not be slower than adaptive highlighting either.

APPLICATIONS FOR EPHEMERAL ADAPTATION

Though we focused on pull-down menus, ephemeral adaptation has broader application to a range of interfaces. Most clearly it could be applied to drop-down or tabbed (as in the Microsoft Office Ribbon) toolbars, but it could also be applied to other interface elements that have a point of onset. Conversely, ephemeral adaptation would not be appropriate for visually persistent toolbars.

It could additionally be applicable in contexts that are not necessarily user-adaptive but are visually complex. For example, ephemeral adaptation in a busy webpage like the New York Times homepage could help guide users to content the site designer deems to be important (Figure 7). While font size and bolding are commonly used techniques to guide the user's attention and help structure the page, ephemeral adaptation would cause main content to appear abruptly, with the other elements fading in gradually. Of course, a challenge here would be deciding which content should be featured. Similarly, ephemeral adaptation may be useful for guiding the user's attention during information visualization tasks: briefly providing high-level structure before the remaining visualization appears could improve visual processing time.

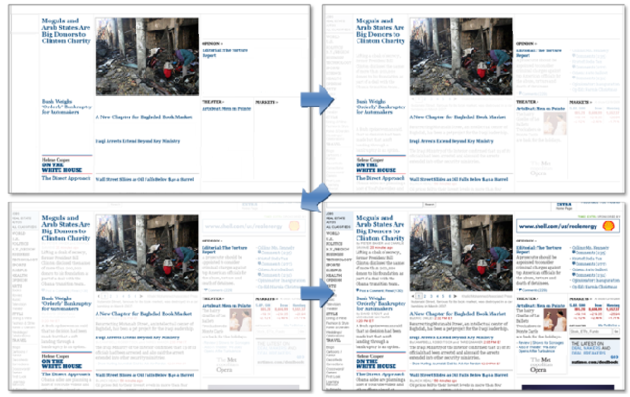

Figure 7. Ephemeral adaptation applied to a news website; the headlines appear before the content text and advertisements.

FUTURE WORK

A few possibilities for refinement of the ephemeral adaptation technique exist. In Study 2, some participants reported that they would prefer either a longer or shorter onset delay, suggesting that further tuning is needed. In Study 1 and piloting beforehand we determined that the optimal delay is between 250ms and 1000ms, but this range could be narrowed further. Another possible modification is to change the gradual onset function. In all of our ephemeral adaptation conditions, we used a linear darkening algorithm, but other options exist, such as transitioning more slowly through the lighter increments but speeding up for the darker ones, or vice versa.

The optimal onset delay may also depend on the level of adaptive accuracy. We did not see any indication of this in the Study 1 performance results, but we only examined two onset delays thoroughly. Moreover, the satisfaction scores do provide a preliminary indication of an interaction between accuracy and delay length, with the lower accuracy group reporting slightly higher satisfaction for the shorter delay, and the higher accuracy group choosing the longer delay. This makes sense because a longer delay costs more when the prediction is wrong, and this cost could begin to dominate as accuracy falls. Future work could seek to confirm this interaction. If it is true, one possibility to manage the cost/benefit would be to dynamically change the delay onset based on the observed accuracy of the adaptive algorithm.

Future work could also compare ephemeral adaptation to adaptive split menus. However, the similarity of our experimental task to the desktop condition and accuracy levels used by Findlater and McGrenere [5] suggests that ephemeral adaptation will be faster than an adaptive split menu (since their comparison showed that split menus were not faster than static menus). If anything, since the menus we used were shorter than those used by Findlater and McGrenere, the reduction in movement time that could be offered by adaptive split menus would be even less.

CONCLUSION

We have introduced *ephemeral adaptation*, a new technique that uses a temporal dimension to reduce visual search time of GUI elements while maintaining spatial consistency. Ephemeral adaptation uses a combination of abrupt and gradual onset to draw the user's attention to the location of adaptively predicted items: when applied to a pull-down menu, predicted items appear abruptly when the menu opens, after which non-predicted items gradually fade in. In contrast to spatial and graphical techniques, which have tended to only find success when adaptation greatly reduces the number of steps to reach desired functionality, ephemeral adaptation shows promise even for relatively short, single-level pull-down menus.

Through two controlled laboratory studies, we showed that ephemeral adaptation results in both performance and user satisfaction benefits over a static control condition when adaptive accuracy is high, and is no slower when adaptive accuracy is low. We also showed that, at high adaptive accuracy, ephemeral adaptive menus were faster than a colour highlighting technique and both adaptive techniques were preferred to static menus. There was no performance difference between highlighted menus and static menus. The fact that highlighting was liked was surprising considering the lack of performance results and the negative response it has received in previous studies [7,8].

Combined, our results show that ephemeral adaptation is a promising technique for guiding visual search in complex interfaces. It should be applicable to a broad range of applications because the adaptive support disappears after only a brief delay, allowing for standard interaction with the interface. Ephemeral adaptation may also be useful for visually complex tasks such as scanning a busy web page or navigating information visualizations.

ACKNOWLEDGMENTS

We thank Ron Rensink for invaluable early feedback on this research and Joel Lanir and Tony Tang for comments on the paper. We also thank NSERC for funding.

REFERENCES

1. Breitmeyer, B. and Julesz, B. The role of on and off transients in determining the psychophysical spatial frequency response. *Vision Research 15*, (1974), 411–415.
2. Cockburn, A., Gutwin, C., and Greenberg, S. A predictive model of menu performance. In *Proc. CHI'07*, (2007), 627–636.
3. Cohen, J. Eta-squared and partial eta-squared in communication science. *Human Communication Research 28*, (1973), 473–490.
4. Findlater, L. and McGrenere, J. A comparison of static, adaptive and adaptable menus. In *Proc. CHI'04*, (2004), 89–96.
5. Findlater, L. and McGrenere, J. Impact of screen size on performance, awareness, and user satisfaction with adaptive graphical user interfaces. In *Proc. CHI'08*, (2008), 1247–1256.
6. Folk, C.L., Remington, R.W., and Johnston, J.C. Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance 18*, 4(1992), 1030–1044.
7. Gajos, K., Christianson, D., Hoffmann, R., Shaked, R., Henning, K., Long, J.J., and Weld, D.S. Fast And Robust Interface Generation for Ubiquitous Applications. In *Proc. UBICOMP'05*, (2005), 37–55.
8. Gajos, K.Z., Czerwinski, M., Tan, D.S., and Weld, D.S. Exploring the design space for adaptive graphical user interfaces. In *Proc. AVI'06*, (2006), 201–208.
9. Gajos, K. Z., Everitt, K., Tan, D. S., Czerwinski, M., and Weld, D. S. Predictability and accuracy in adaptive user interfaces. In *Proc. CHI'08*, (2008), 1271–1274.
10. Greenberg, S. and Witten, I. Adaptive personalized interfaces: A question of viability. *Behaviour and Information Technology 4*, 1(1985), 31–45.
11. Jonides, J. and Yantis, S. Uniqueness of abrupt onset in capturing attention. *Perception and Psychophysics 43*, 4(1988), 346–354.
12. Mitchell, J. and Shneiderman, B. Dynamic versus static menus: An exploratory comparison. *SIGCHI Bulletin 20*, 4(1989), 33–37.
13. McCormick, P.A. Orienting attention without awareness. *Journal of Experimental Psychology: Human Perception and Performance 23*, 1(1997), 168–180.
14. Sears, A., and Shneiderman, B. Split menus: Effectively using selection frequency to organize menus. *ACM TOCHI 1*, 1(1994), 27–51.
15. Theeuwes, J. Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception and Psychophysics 49*, 1(1991), 83–90.
16. Theeuwes, J. Stimulus-driven capture and attentional set: Selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance 20*, 4(1994), 799–806.
17. Tolhurst, D.J. Reaction times in the detection of gratings by human observers: A probabilistic mechanism. *Vision Research 15*, (1974), 1143–1149.
18. Tsandilas, T. and schraefel, m.c. An empirical assessment of adaptation techniques. In *CHI'05 Extended Abstracts*, (2005), 2009–2012.
19. Tsandilas, T. and schraefel, m.c. Bubbling menus: A selective mechanism for accessing hierarchical drop-down menus. In *Proc. CHI'07*, (2007), 1195–1204.
20. Yantis, S. and Jonides, J. Abrupt visual onset and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance 10*, 5(1984), 601–621.
21. Yantis, S. and Jonides, J. Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance 16*, 1(1990), 121–134.