

===== REVIEWS =====

----- Submission 453, Review 4 -----

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

Reviewer: AC  
Overall rating: 4.5 (scale is 1..5; 5 is best)

Contribution Type Specific Rating (Meets Contribution Type Specific Criteria Well)

4.5 . . . Between agree and strongly agree

Overall Rating

4.5 . . . Between possibly accept and strong accept

Expertise

4 (Expert )

Contribution to HCI

Demonstrates that "ephemeral adaptation", where predicted menu items are displayed immediately while unpredicted ones fade in, can improve menu item selection performance. There are several strong contributions: the ephemeral adaptation interface strategy is interesting, novel and practical; it is demonstrated to outperform item highlighting; ephemeral adaptation performance parameters are explored; and the paper is an exemplar for its review of related work and presentation.

The Meta-Review

Congratulations. The reviewers and I agree that this is a fine paper. The presentation is exemplary, as is the review of related work. The idea of controlling visual onset is original and well motivated, and the empirical work is sound. I will be arguing that the paper be accepted at the program committee meeting.

Furthermore, I would like to nominate the paper for the best paper award. For this, however, I would like your rebuttal to address some questions from the reviewers, as follows.

**\*Outlier removal\***

R1 and R3 raise concerns about removing participants from the data set. Like R3, I have not previously seen removal at 2s.d. from the mean (although 3sd from the mean is common). Does including this participant radically affect the results? Did anything other than the speed of their performance mark them as outliers?

**\*Novice versus expert performance\***

R2's main concern lies with expert performance: "one could argue that its benefit for frequent users would be relatively small". Indeed, for expert users (who know the location of items), gradual onset may harm

performance when the system's prediction is incorrect. R2 asks that you "comment on how the results changed from the first to second block". [AC's comment: Given that you used a Zipfian distribution (with some items appearing very often), I suggest you analyse performance with inexperienced (say, less than 5 selections) and experienced items (more than 5). Is there an interaction between interface type and experience? If your technique is only useful for novices, it's still worthwhile; but less so than if users benefit throughout. ]

**\*Adaptive accuracy rates\***

As R1 requests, can you provide a rationale for selecting 50% and 79% for the accuracy conditions?

The reviewers also provide many further comments and recommendations that you make like to comment on in your rebuttal.

-----

Thanks for your detailed rebuttal. Reviewers commented that it further increased their already high estimation of the work, recommending that it be nominated for the best paper award.

Associate Chairs Additional Comments

----- Submission 453, Review 1 -----

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

Overall rating: 4.5 (scale is 1..5; 5 is best)

Contribution Type Specific Rating (Meets Contribution Type Specific Criteria Well)

4.5 . . . Between agree and strongly agree

Overall Rating

4.5 . . . Between possibly accept and strong accept

Expertise

4 (Expert )

Contribution to HCI

This paper elaborates on how a "reverse" highlighting of interface widgets/objects can be used to draw the user's attention to interesting parts of the interface. This reverse highlighting means that instead of highlighting widgets of interest these are presented in a standard way and in standard colour but the uninteresting interface parts are at first invisible and then, over a short time, gradually fade in until fully visible. The paper describes how the new technique in combination with prediction functionality can be applied to pull-down menus to create

"Ephemeral adaptive menus" (prediction functionality is not a research topic further discussed within the paper, it is assumed and serves to achieve adaptability). But it is also pointed out that, given a prediction functionality, the "reverse" highlighting ought to work on other interface widgets as well and that the menu implementation serves as a demonstration example. In the Ephemeral menu, the prediction functionality identifies the three menu items which are most likely to be selected next. The predicted items are presented as normal when the menu opens up, all non-predicted items have invisible labels at first. The labels of the non-predicted items then start to fade in, and after a short period of time (500ms) their labels have reached the same colouring as the labels of the predicted items, and now the Ephemeral menu looks and behaves as a standard pull-down menu. The hypothesis is that this functionality will help the user to localise the most probable item(s) and thus make the menu interaction easier and faster than when using a standard menu or an adaptive menu that either reorganises or colour highlights the items according to selection probability; colour highlighting supposedly fails to attract attention, reorganisation causes an unfavourable spatial instability in the interface. The effect of the new highlighting strategy is demonstrated in two user experiments where performance using an Ephemeral adaptive menu is compared with a standard menu and an adaptive menu which colour highlights predicted items.

As the authors state themselves, the main contribution of this paper is the introduction of the ephemeral adaptation approach which rely on a time component to visually pronounce/communicate adaptive effects/changes. This is of highly valuable for researchers working on adaptive interfaces. Since the technique seems applicable in other non-adaptive contexts the value is increased.

## The Review

First of all: it was hard to find anything to "complain about" in this paper. Nice work.

Strong aspects:

- \* The presented technique makes use of findings from human perception research: the effect of an 'abrupt onset' as attention-catcher. The usage has (to my knowledge) never been explored in an explicit way like this in HCI before.
- \* Previous/related work on adaptive interfaces is adequately reviewed.
- \* The problem set out to solve was: how to maintain the advantages of user adaptation in an adaptive GUI and at the same time avoid the disadvantages existing with current adaptive GUIs (spatial inconsistency, obtrusive or ineffective colouring of GUI objects). This problem is worth solving.
- \* The proposed technique is straightforward, description is easy to understand.

Somewhat less strong aspects (unordered):

- \* The validation is rigorous but the results show only a gain in a best-case situation (where prediction accuracy is as high as 79%). This gain might be of practical significance: according to my interpretation/calculations using the provided Bar graphs, the technique seems to reduce menu selection time by 5 to 8%. It seems possible that the technique might also yield gains when applied on other GUI-widgets or

in visually complex tasks, as suggested by the authors. However, since I regard the main contribution of this paper to be the introduction of the 'abrupt onset tool' to HCI, the not overwhelming (and hard to find) performance increase in this menu example is less "damaging".

\* Since the main contribution is the abrupt onset, it would have been beneficial if little bit more space had been dedicated to its application in non-menu/tool selection (Ribbons, toolbars etc.) contexts. The authors mention web-interfaces (is elaborated a bit further at the end of the Discussion section) and information visualisations as possible other contexts, but the reader is left with her/his own imagination (I know, hard to find space, but for example, most of the Bar graphs could be squeezed down without a loss in readability, the less important pilot-section and/or the lengthy repetitions of the hypotheses in the test-summaries could be shortened).

\* It would have been good with a short note on how realistic a prediction accuracy of 79% is. If there has been any research on this in particular, that reference is missing. I am not 100% convinced that such high accuracy is realistic. One might, somewhat boldly, reason and question whether or not users could in real-world usage be helped by this adaptive feature. Why? When starting using a particular menu and the functions behind its menu items, the user would be suitably served by an Ephemeral menu that helps to localise the needed items, however at the beginning with less expressive usage patterns to work with, can high enough prediction accuracy be achieved? Twist-around: when a high prediction accuracy seems more feasible, i.e., at later stages with stronger usage "history", has not the user by then learned where the different items are located, in particular the most frequently used ones? Here/now the high prediction accuracy seems less useful.

\* I was wondering about what was happening in the tests of the two persons who's results were excluded from the data analyses. Were they "wild-hitters" going for speed rather than accuracy? Software errors? From another population group (in the statistical sense)? This should have been explained.

\* I was also wondering about the differences between the results of the first and the second test. I would have expected that the mean time for the ephemeral+non-predicted condition in test 2 would be about the same as the mean time for long-onset+non-predicted in test 1. But reading and scrutinizing the corresponding Bar graphs, the mean time in test 2 appears to be about 1950-2000ms and in test 1 only about 1600ms, i.e., roughly 15-20% faster in test 1. This seems to be a lot, in particular when taking into account that the data from test 1 come from both low and high accuracy trials, as I understand.

\* First I was struggling a bit to understand "there was an implicit error penalty in the speed measure..." After a while I understood where this penalty was concealed: by using the median the influence of the outliers was reduced – presumably an outlier was in most cases a trial where the first item selection was erroneous and had to be repeated – and at the same time the slow times of the "negative" outliers were shifting the median upwards and so the "penalty" manifested itself. I kind of like this approach. In some way it makes the speed measure and the results more expressive compared to when error frequencies and speed measures are

reported as two separate measures that have to be weighted against each other to judge the speed-accuracy trade-off inherent in this kind of user task. However, a more explicit explanation about the implicit penalty would support many readers I guess.

Nevertheless, the paper's contribution to HCI is significant: the paper provides researchers that are trying to improve interfaces and interaction techniques with a new promising "tool" to further explore.

#### Areas for Improvement

Excellent presentation! All sentences and passages clearly written using easy to understand language for non-native English speakers. Figures are serving their purpose, no more no less, are appealing and correctly placed at the top or bottom of columns. No obvious reference is missing, everything suitably backed up. Presentation very near to perfect! Six minor "typos":

- \* according to CHI proceedings format, references in the reference listing should not have the publishing year directly after the author list, instead it should follow the name of the publisher right before the page numbers.
- \* some inconsistencies in writing proceedings' abbreviated names, e.g. ref 2, 4, and 5.
- \* the label for the y-axis in Figure 6 could be changed to Trial Selection Speed (Time->Speed) to be consistent with Figure 2, 3, and 7.
- \* page 5, last paragraph: (2) speed for those trials there were not... there->that
- \* top of page 9: "predictive accuracy is high (78%)" 78->79 (?)
- \* test 1 error rate for Control is missing.

----- Submission 453, Review 2 -----

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

Overall rating: 4.5 (scale is 1..5; 5 is best)

Contribution Type Specific Rating (Meets Contribution Type Specific Criteria Well)

4.5 . . . Between agree and strongly agree

Overall Rating

4.5 . . . Between possibly accept and strong accept

Expertise

4 (Expert )

Contribution to HCI

This paper introduces ephemeral adaptation -- an approach for adapting user interfaces at run-time based on the user's current task. Ephemeral adaptation---where the promoted items are presented abruptly while others are faded-in gradually---has never been characterized before. The

authors make two contributions: 1. they demonstrate that the approach can have positive impact on user satisfaction and performance and, 2. for the duration of the onset delay---the main tunable parameter of this adaptation method---the authors explore a range of values and find one that strikes a good balance between enhancing visual salience of promoted items and minimizing the delay in accessing the non-promoted items. This is a solid and useful contribution to the effort to systematically explore the design space of adaptive user interfaces.

## The Review

This paper makes a contribution to the systematic study of the design space of user interfaces that dynamically adapt to the user's current task. Specifically, the paper introduces Ephemeral Adaptation, empirically evaluates appropriate settings for the parameters guiding the behavior of such adaptation, and it provides evidence that this type of adaptation can improve users' satisfaction and performance. I am fairly familiar with the literature in this area and I believe that the contributions of this paper are entirely novel.

I think the paper is important and timely: "intelligent" (or machine learning-driven) interactions are attempted more and more frequently in our community. Yet, little work has systematically explored the challenges of designing robust interactions with intelligent systems.

I find the scope of the contribution and the quality of the execution to be perfectly appropriate for CHI.

My main concern with the proposed approach (which I hope the authors can discuss in the final version of the manuscript) is the following: adaptation driven by a recency- or frequency-based algorithm, will be most helpful for frequent users of the system, and will be particularly helpful for interacting with elements that the user has used before (perhaps even frequently). The adaptive mechanism proposed in this paper helps reduce visual search so one could argue that its benefit for frequent users would be relatively small. I thus wish that the authors had designed their study such as to give participants more time to develop familiarity with the interfaces, so that we could have a better sense of how this approach impacts experienced users. Given that each participant completed 2 blocks of trials (presumably with the same interface) in each condition, could the authors comment on how the results changed from the first to second block?

**\*\* After Rebuttal \*\***

Thanks for addressing the novice versus expert performance in your rebuttal. Makes the paper even stronger.

## Areas for Improvement

Introduction, p.2: when you first mention high and low accuracy levels, mention the accuracy numbers

data figures: it would help if you found a way to visually indicate significant pairwise differences

Figure 3: differences in (a) are very small compared to those in (b) -- this suggests that at 50% accuracy levels, the differences for non-predicted trials should dominate the overall results (thus making control faster than the other two conditions) but that's not the result we observe. What am I missing in interpreting Figs 3 and 2?

Preference rankings: is there a reason why you didn't look for statistical significance of preference rankings in either of the studies? Friedman test would be a good choice here.

Split menus: because the original version and the commercial deployments of this concepts differ in that the former moved elements while the latter copy them, it would be helpful if you could be more explicit in your discussion of related work as to which approach the evidence pertains to; arguably, few people complain about the commercial version of split menus.

----- Submission 453, Review 3 -----

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

Overall rating: 4.5 (scale is 1..5; 5 is best)

Contribution Type Specific Rating (Meets Contribution Type Specific Criteria Well)

4.5 . . . Between agree and strongly agree

Overall Rating

4.5 . . . Between possibly accept and strong accept

Expertise

3 (Knowledgeable)

Contribution to HCI

The authors present an innovative adaptive menu performance technique to enhance users' awareness of predicted added elements to a menu based on context of menu use. The authors motivate this subtle but effective technique of ephemeral onset awareness by an excellent review of related literature both in HCI and in psychology. While there are a few addressable problems in the study presentation, the work is solid, and the way the work is presented is exemplary, offering a dual contribution both of the technique itself, and the way its presented.

The Review

In terms of the quality of the research, experimental methodology and results analysis, the paper is excellent, with the exception of some possible problems discussed below.

A real strength of the paper is the fact that the paper has thoroughly taken into account all the previous experience in the evaluation of adaptive user interfaces, resulting in a very sound experimental design.

The only weakness of the paper is the extent of its innovation, and it may be criticized for that. It introduces one more menu technique, which is not really exciting in terms of design and results in small only speed benefits (<10%, although we shouldn't neglect the subjective preferences of participants).

However, the new technique is not a random choice and has been extensively tested. Also, it has been well motivated in terms of previous research coming from Experimental Psychology. The extensive reference of the paper to results from Psychology is a strong part of the paper and its contribution. Although not very exciting, such results are important for helping researchers and designers assess the potential of adaptation techniques, use them appropriately and avoid mistakes of the past.

#### Areas for Improvement

Below, are a few comments on problematic points and suggestions for thought and discussion.

- Removing the results of a participant altogether (both in Study 1 & 2) sounds a bit suspicious. This shouldn't be done without good reasons for that, e.g., if it was observed that a participant could not follow instructions or had a particular problem that did not qualify him/her as a regular user. I would expect some further explanation. Also note that many researchers suggest 3 rather than 2 standard deviations away from the mean as a cutoff point for outliers. At least, shouldn't these participants' subjective preferences and satisfaction rankings be included in the analysis? Why were they ignored?

- Error rates in Study 1 (as opposed to results in Study 2) are not reported in detail, i.e., per technique. Although the inflation of error rates in the High accuracy condition is not huge, it would be good to know its cause. As the borders of the predicted items might not be well perceived from the beginning of a fade-in process, it might be the case that some of these errors were generated when participants tried to click on targets of Ephemeral menus that had not yet been faded in. If this is the case, the problem could be possibly fixed by always showing the borders of the the predicted items. However, I'm not sure that this was a cause of errors, since in Study 2, no difference in errors was found. It might be also the case that errors were only caused by the short-onset Ephemeral technique, e.g., because users were enforced to click predicted items more rapidly.

- It's interesting to see that the Highlight technique did not improve performance, but it was, however, preferred by participants (is it possible that highlighting reduces cognitive overhead even when it does not improve performance?). I believe that highlighting is more useful when users have not yet learned the position of items in a menu, as it can help them identify items faster. As the user learns the position of items, its role becomes weaker. In the experimental setting described by



the paper, users had to repeat the same tasks many times one after the other, so I guess that learning was relatively fast. I'm also wondering whether, in addition to attention, ephemeral adaptation has a positive effect on learning (see also Grossman et al. for comparisons of various visual effects in menus to assist learning of hotkeys). In any case, it would be interesting to check whether there was any significant interaction effect of block or think about future tests.

Ref: T. Grossmann et al. Strategies for Accelerating On-line Learning of Hotkeys. CHI 2007

----- Submission 453, Review 5 -----

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

Reviewer: 2AC

Overall rating: 4.5 (scale is 1..5; 5 is best)

Contribution Type Specific Rating (Meets Contribution Type Specific Criteria Well)

4.0 - Agree

Overall Rating

4.5 . . . Between possibly accept and strong accept

Expertise

3 (Knowledgeable)

Contribution to HCI

Presents a form of animated feedback for adaptive menus that is shown to outperform previous approaches.

2AC Review

I agree with the reviews: this is a truly impressive piece of work in terms of both methodology and presentation. The findings are non-trivial and have significant implications for HCI, especially in the field of adaptive user interfaces. It will be a perfect CHI paper.