

**Reviewer 4 (Primary AC)**

Submission title Exploring Methods to Improve Pen-based Menu Selection for Younger and Older Users

**Overall Rating**

4 (Borderline: Overall I would not argue for accepting or rejecting this paper.)

**Expertise**

3 (Knowledgeable)

**Contribution to HCI**

Describes the design and evaluation of two interfaces that help some users reduce errors in menu selections with stylus input devices. The specific error under study is 'missing just below', where users erroneously acquire the menu target immediately below the intended one. One of the two designs introduces a dead-space gap between items, and the other reassigns, or shifts, the motor-space associated with each item slightly downwards.

Experiments are conducted with novices in two age groups: 19-30 and 66-81.

**The Meta-Review**

It is very difficult to form an overall recommendation for this paper. The reviewers' recommendations are quite varied (two 5s, one 2, one 3), and from my own read I can see both the strong merits and the strong limitations. I would like to recommend acceptance, but without a strong rebuttal addressing the paper's primary weaknesses (below), I will be unable to do so.

First, the strengths.

- The paper is extremely well written (R1, R2, R3, R4), and the clarity of all aspects of the paper (motivation, related work, design, methodology, results analysis and discussion) is to be generally applauded (with a few minor niggles as identified in the detailed reviews).

- Evaluation with two user groups (young and old participants) is also highly laudable. Few researchers go the extra mile to involve participants away from their university/lab, so the inclusion of relatively elderly participants is a major strength of the work. Three of the five contributions that R1 identifies stem from your investigation of different participant pools.

- The data generated by your evaluation is richly presented and summarised. As R2 states "what I find remarkable, is how the authors turn an unexpected and potentially negative result into something valuable and ultimately of significance".

However, all reviewers raise important concerns regarding the paper's significance. Reviewer 1, an expert in the area, provides a particularly deep and insightful review, including an important analysis of the prior work by Moffatt and McGrenere, which you use as a platform for launching this research.

To summarise the limitations, which I would like rebutted:

1) Significance of the problem: Frequency of errors (motivation and results) You motivate the research partly by emphasising the prevalence of 'missing just below' errors. Specifically, you state that Moffatt and McGrenere's data shows that "a selection along the top edge of a menu item is 11 times more likely to be intended for the item above it". This is well written and compelling. But it also masks the real nature of the problem, which is, as noted by R1, the frequency of error. R1's analysis of Moffatt and McGrenere's data shows that only 2.6% of selections were erroneous, and that roughly half of these errors were made by a small subset of participants. How frequently does the problem arise?

The problem of absolute error rates applies to the results as well as the motivation. R1 and R4 comment on the problem of details overriding the large-scale meaning. While the results analysis carefully dissects the subtleties of the techniques for different users, the "big picture" of how often these errors occur is absent (R1 does a good job of piecing this information together, and his/her ultimate finding is disappointing).

In your rebuttal please comment on the absolute error rates, both motivating your study and measured in your experiments. What are the absolute error rates for each of the conditions in the study? Is there evidence that you don't just shift missing just below errors into missing just above ones?

2) Significance of the problem: Novices, and Error Cost. As you note, the cost of errors will influence user performance. With experience, users are likely to notice if they frequently miss menu items, and they will therefore take more care, possibly implementing a human control mechanisms to avoid the bottom of menu items, largely equivalent to your current "reassignment" solution (e.g. "I always aim high to make sure I hit the target"). R1 notes that you might add "Novice" into your title.

In your rebuttal please comment on whether you think users will naturally accommodate their performance if they exhibit high error rates.

Another aspect of concern (not mentioned by the other reviewers) is that it seems your methodology implicitly encourages participants to optimize for speed rather than accuracy---they advanced to the next task regardless of correctness. Although you invited participants to proceed "as quickly possible while remaining accurate", it seems likely that after 1.5 or 2 hours of trials your participants will migrate towards optimisation for speed to complete the evaluation more quickly. That is, errors had no cost for participants during the experiment.

Comments?

3) Comparison with other approaches. Review 3 raises the concern that other approaches, previously described in the literature, might outperform your solution.

Comments?

Overall, as AC I am required to focus primarily on the reviewers' assessment of the paper's contribution to HCI, its significance, and the benefit others can gain from it. Unfortunately, the weakness that R1 (in particular) and R2 note are directly targeted at these aspects of the paper, and their arguments are strong. I hope you can provide an equally strong rebuttal.

**Reviewer 1**

Submission title Exploring Methods to Improve Pen-based Menu Selection for Younger and Older Users

Review type External

**Overall Rating**

3 (Possibly reject: The submission is weak and probably shouldn't be accepted, but there is some chance it should get in.)

**Expertise**

4 (Expert)

**Contribution to HCI**

The paper presents two novel menu designs for pen-based interfaces which aim to tackle the "problem" users are having in hitting slightly below the desired menu item instead of correctly selecting the target item. The new designs did however only marginally improve performance (which is perhaps not such a positive finding for the authors, but is of course a valid research finding). However, I am not convinced this is a common problem and therefore this contribution is in my view significantly reduced (see below). To me, the main contributions of this paper are instead that it: 1) clearly shows that there are individual differences in the tap distributions among users of pen-based menu selections (high-hitters, low-hitters and neutrals). This indicates that designers of such interfaces need to include this aspect in their designs. 2) clearly shows that older inexperienced pen users with limited computer experience have more difficulties operating a menu using a pen than younger inexperienced pen users with more computer experience have. 3) provides an easy to replicate menu experiment with pen input and so extends the sparse literature in this domain. 4) extends the sparse literature on age related issues in interaction design. 5) shows (after close consideration and reading between the lines) how important the design of the tasks used in such experiments are, and how hard it is to identify the (true) interpretations of the experimental results if the tasks are not properly designed.

**The Review**

I enjoyed reading this very well-written paper. It uses a clear, and for a non-native speaker easy to understand, language. Its structure is suitable and its figures and tables are nicely formatted and are informative and understandable. The design of the presented experiment is adequately described and so could be easily replicated by others. The results and analyses are easy to comprehend and the statistics are sound.

From my point of view, this paper has one major weakness, namely the limited contribution to HCI researchers, and in a further step, the findings and their implications do not seem very likely to be very valuable for future users of pen-driven interfaces. That is, although the authors nicely present and evaluate two possible ways to assist (pen) users during menu selection, the problem they aim to overcome/tackle with their designs is somewhat restricted. I am no tablet-pc user but use a PDA (without a “hovering” functionality), having said that, I am not convinced that the error type “missing just below” is so frequently occurring and that it is such a big problem for the users that severely impede their overall menu performance. After reading the Moffatt&McGrenere 2007 Assets paper (here referred to as [19]) I was surprised that the authors chose to try to tackle the “missing just below” problem! To me, the “drifting” problem described in [19] seems to be a much more critical phenomena (as it “did have a significant negative impact on speed.” And “did not decrease with learning”)! Regardless of that, if my interpretations of the information given in [19] and if my calculations of the provided error statistics are right, only a total of 135 errors were made (=2,6% from n=5184) and 60 of these errors were made by three participants. 58 of the 135 errors were “missing just below” errors (i.e., in about 1.1% of all 5184 trials, a “missing just below” error was registered) and probably many of these errors were made by the three “error producing participants” (how many of the “missing just below” errors that were committed by these three participants are not reported in [19]). This leads me to question whether the frequency of “missing just below” errors really motivates all the design work and experiment work, and in the end a publication at CHI. Furthermore, both in the present study and in [19] all participants were novice tablet-pc users, having no or very limited pen-experiences, but varying computer, and consequently also varying menu selection experience. In the present study, I can not find any information about the distribution of hits just below the target item across blocks – was there a learning effect, i.e. did the number of hits just below the target item decrease with practice or not? I would suspect that over a longer time of usage (perhaps not visible in this experiment, which lasted for about 1.5 hours, with 648 trials + an unknown number of practice trials), a user gets accustomed to operating the pen and learns how to handle it and how to correct and adjust the operation to its and the touch-screen’s peculiarities. That is, I am not sure whether or not the user will after a while notice his or her tendency to tap below the desired menu item (if it can be called a tendency when 36 users together make 58 such selections out of 5148 trials in [19]) and therefore over time “automatically” and “manually” correct for this “beginners habit” during menu selection. Anyway, it would be nice if the authors could provide a note on this issue and clarify that it is unknown whether or not more experienced pen-users also make this type of error. Consequently, I would argue for changing the title of the paper to: Exploring Methods to Improve Pen-based Menu Selection for Novice Younger and Older Adults.

Please also provide the absolute numbers of hits made in the item below resp. above the target item. Based on Figure 6 (showing the tap distributions) using a ruler and by some approximations I count about 110 taps below the target item for the low hitters (about 3% of the 3684 trials) and about 360 taps in the item above the target item for the high hitters (about 6% of the 6238 trials). According to my calculations the overall error rate would

be about 3.1%  $[(110+360)/(432+(23*648)) = 0.03064]$ , again a rather low error rate as in [19] which report an error rate of 2.6%. Well, the low error rate in the present study could not be known a priori by the authors, but perhaps anticipated / suspected based on the results reported in [19]. Furthermore, as stated by the authors, the task design might also have influenced the tap distribution. In [19] a discrete task was used with the start button placed in the middle of the screen, i.e. below the menu target items, in the present study the task was a continuous one somehow “promoting” taps at the top of the target items. In [19] we see a “tendency” to hits below, and here we see a “tendency” to hits above.

To sum up, the low error rates, the participants’ low experience with pen-interaction, the unknown (not reported) error distribution over blocks for the errors in the presented paper and for the “missing just below” errors in [19] (a possible learning effect with decreasing errors over time), the likely possibility that the majority of the “missing just below” errors in [19] was produced by only 3 of the 36 participants, only 6 out of 24 participants in the present study were actually identified to have a tendency to hit in the lower part of the target item (about 3% of the trials from this group was too far below and would cause an erroneous selection), and the possibility that the chosen task designs in [19] and here had an influence on the tap distributions indicate that we have a paper trying to provide solutions for an interaction problem that does not seem to be very common, I am not convinced that the problem even exists at all! This considerably reduces the importance of this paper.

I was also wondering whether or not the “low-hitters” would be “high-hitters” (and reverse) in menus such as the start-menu in Windows XP (or menus popping up when clicking on some of the system-tray icons) or on PDAs or smart phones running Windows Mobile OS where the pull-down menus often are “pull-up” menus placed along the lower border of the screen. Would the new designs in such situations use a reassignment/deactivation of the 10% of the item above?

It would be nice with an early discussion on this issue since I think it would bring us closer to the true “causes” why some people tend to hit below or above. What are/were the authors' hypotheses / explanations on why so many (well, see comment above) people had a tendency to hit below in the study presented in [19]? What were the explanations presented by Moffatt&McGrenere on this issue? Are such patterns caused or influenced by the menu scanning process (i.e., menus anchored at the top border of the screen would have a top-down scanning direction whereas menus anchored at the lower border would have a bottom-up scanning order) or are the patterns more a product of the way the pen is operated and held (i.e., used angle to the surface, the grip, the distance between the pen and the surface when the pen is unused (i.e., when reading or waiting for system response), etc.)? Furthermore, could it be a matter of left/right handedness (I could not find any information in the participants section about which hand the participants used, were all right-handed? This would be more interesting to know than how much the participants were rewarded for participating)?

After weighting this major weakness (i.e., the attempt to solve a non-existing problem) of the paper against its positive aspects and its contributions, I rate it as "Possibly reject:

The submission is weak and probably shouldn't be accepted, but there is some chance it should get in".

### **Areas for Improvement**

Some minor typos and formatting issues:

- \* Page 1, second column, fourth row: the white space is too wide between “edge” and “of”
- \* Page 2, first column, middle of third paragraph: “However, Moffatt and McGrenere noted [19] that users typically do not....” I would put the reference number before “noted”
- \* Page 2, first column, fifth row of fifth paragraph: the apostrophe used for “individuals” has another font than the rest of the text (compare the apostrophe in the word “target’s” in the last paragraph on page 2).
- \* Page 2, top of second column: I could only spot ONE (i.e., the task design) identified reason for the different tap distributions in this study compared with the study in [19].
- \* Page 5, middle of last paragraph in second column: a faulty punctuation is made after the word benefit!
- \* Page 9, the last sentence in the fourth paragraph has no punctuation.
- \* References on page 10, some titles have capital letters ([19, 22, 28]) the other have not.
- \* Figure 8 would be better placed at the end of the first column on page 9.
- \* According to the submission format: references should not have the year in braces after the title, instead the year in braces should follow the publisher. CHI’04 should read CHI 2004. For example, [18] should read: Kurtenbach, G. and Buxton, W. The limits of expert performance using hierarchic marking menus. In Proc. CHI 1993, ACM Press (1993), 82-487.

### **Reviewer 2**

Submission title Exploring Methods to Improve Pen-based Menu Selection for Younger and Older Users  
Review type External

### **Overall Rating**

5 (Possibly accept: Possibly above the line, but I wouldn't want it to edge out stronger submissions.)

### **Expertise**

3 (Knowledgeable)

### **Contribution to HCI**

This Paper introduces and studies two method aimed at improving the selection of menu items in pen-computers for both younger and older adults. In particular the authors address the problem of users missing a target because their pointing becomes downwardly shifted. In addition, the authors present an empirical study that finds great variability of pointing distributions among both younger and older adults, thus defining three groups: downwardly shifted clicks (low hitters), upwardly shifted clicks (high hitters), and unshifted clicks (neutrals). This paper’s study also reveals how the

characteristic of the two proposed methods affects the performance of these three user groups.

### **The Review**

This paper follows a very traditional structure. It states the problem very clearly, it proposes solutions to it and then proceeds experimentally to test them so as to confirm a number of hypotheses. And it does all these things beautifully: the problem is simple and of potential significance, since its solution has the potential to impact a user population in need. The related work section references the appropriate literature. All explanations are extremely well written and clearly presented.

The experimental section is also very well presented in a way that satisfies the statistically inclined but remains accessible to readers only looking for insightful data interpretation. I really liked how the authors turn an unexpected and potentially negative result into something valuable and ultimately of significance. Bravo!

That being said, I inclined to believe that the contributions of this paper to HCI are limited: How often is the problem a problem? It seems from Figure 6 that the percentage of misses on low-hitters, neutrals and high-hitters is marginal. In this particular case, it does not seem that the cost of making an error is high enough to make such a marginal number of misses a serious issue. I welcome the authors to use their rebuttal to clarify as to the significance of this problem and hence of the paper's overall contribution.

Overall, I believe that the superb presentation of this paper is in itself a contribution and makes me rate it possibly above the line. Still I also believe that the paper is not stronger than submissions with a clear contribution to the field.

### **Areas for Improvement**

There is very little I can suggest to improve the quality of this paper. One thing the authors might want to consider is to include an image that describes with a storyboard what pointing interactions look like for each of the proposed techniques (i.e., reassigned & deactivated).

Another small detail is to clarify the monetary unit for the compensation (e.g. USD?)

### **Reviewer 3**

Submission title Exploring Methods to Improve Pen-based Menu Selection for Younger and Older Users

Review type External

### **Overall Rating**

2 (Reject: I would argue for rejecting this paper.)

### **Expertise**

3 (Knowledgeable)

## **Contribution to HCI**

This paper presents two designs for making the menu selection task robust to missing-just-below errors. A missing-below error occurs when the target item is missed by selecting the item below the target. Missing-below is also characterized by a significantly smaller frequency of reaching the top-edge of a menu target item. The paper investigates the issue of missing-just-below errors in a pen-based environment, with a focus on demographics separated by age (older vs. younger). The two designs based on 1) reassigning item space and 2) deactivating pixels are compared and results do not show much difference between designs for the group of individuals where missing-just-below appears frequently. While these contributions were derived in a very systematic manner, and with good experimental rigour, in my view the approach did not include (or discuss) competing designs that might be clear winners in contrast to those proposed.

Furthermore, the paper does not convince the reader that reassigning or deactivating pixels work sufficiently well. As a result, I am left with the impression that this paper is an incremental contribution to the work carried out by Moffatt and McGrenere (2007) that brings into question its publication at CHI.

## **The Review**

The authors carry out a good review of prior work. However, the moment I was introduced to the problem in the Introduction (which was very well carried out), I immediately created expectations for seeing designs that might take on a variation to the delphian desktop in which motor control prediction could be based on calibrations for individual user profiles and made to work sufficiently well for reducing missing-below errors. I am still left wondering whether such designs could work in the scenario in this paper. Also, the authors missed a reference to the bubble-menus presented at CHI 2007, a technique which I felt might work sufficiently well (if tweaked) on pen-based input and with the problem presented here.

I am not entirely convinced by the proposed solutions for several reasons. (1) After reading the paper very carefully a couple of times, I still could not find reference to the low level pixel information about the reassignment and deactivated space on the menu items (I finally concluded that I missed it somewhere in the paper rather than this being an omission from the authors). However, the amount of space reassigned needs careful consideration. On tablets I am always challenged by the parallax between pen tip and interface widget. The problem is even more serious when I switch from portrait to landscape view, in which I usually have to recalibrate the pen. So to me it seems as though the space allotted in the redesigned menu interfaces would suffer seriously from parallax as well as reoriented devices. If it does not then it would be helpful to get insight from the authors as to why it is not the case. (2) I am also not convinced whether either design is really that robust, as deactivated pixels would force more taps (as presented in the results) and reassignment could lead to significant frustrations. The latter solution (reassignment) does not seem to be adequately tested as in a real-world setting (ecological validity), as even suggested by the authors, the user triggers a selection and moves onto the next step of their task hierarchy (such as moving to click a button on a newly opened up dialog box). However, the continuous selection task does not seem to

test this issue adequately and the experimental task might have benefited from the inclusion of a second step. In doing so, I believe the analysis of Cost of Re-Tapping might have led the researchers to a significantly different conclusion than the one presented.

Overall, I believe the authors have tackled an interesting topic (and which could be more strongly motivated), but have not presented a convincing case for adopting either of the two designs presented in the paper.

### **Areas for Improvement**

References to some of the prior work is missing, such as bubble menus (2007), Shift (2007) in place of Offset Cursor.

But generally very well written, with very few typos or errors.

### **Reviewer 5**

Submission title Exploring Methods to Improve Pen-based Menu Selection for Younger and Older Users

Review type External

### **Overall Rating**

5 (Possibly accept: Possibly above the line, but I wouldn't want it to edge out stronger submissions.)

### **Expertise**

3 (Knowledgeable)

### **Contribution to HCI**

Presents an empirical evaluation of two new techniques to improve menu selection in pen-based interfaces.

### **The Review**

I really enjoyed reading this paper. It is not only well written; their study has a very clean, clear design and the analysis is exhaustive, smart and, in my opinion, statistically careful. It is rare to see papers this precise and clear in the analysis and presentation of the results.

There are only two things that I missed in the paper: a more detailed discussion on how to classify the users into high- low- and neutral- hitters, and a more explicit explanation of the repercussions of the differences found between techniques. The former is important because without it, it would be difficult to replicate their results. I think that as the field matures, "individual user differences" will take an increasingly important role, but it will also difficult to agree on how to make meaningful clusters of users; therefore we need clear explanations of the divisions, and a justification of the choices made.

The other missing element is a clear interpretation of their results in terms of real use. After so many F values and post-hoc comparisons, we got a pretty good idea of how techniques compare to each other, but it was difficult for me to imagine what the real benefit of each of the new techniques is. This connects with what I think could be the most serious problem of the paper: the size of the contribution.

The difference between the techniques is already pretty subtle (we are talking about just a couple of pixels), and I wonder also how subtle the advantage of these techniques is. The devil is in the details, the authors will claim, and I agree, but it is difficult to know if all this carefully crafted research will be just forgotten when we find more radical ways to solve the problem of selection when using pen-based interfaces. In fact, some of the new techniques such as expanding targets or certain kinds of menus might already have a much larger impact on performance than just a couple of pixels up or down.

I was also surprised that the authors did not discuss at all the problem of calibration/parallax in pen based devices. Is it possible that some of the high and low hitters were so because of the particular errors of the calibrations they performed?

I think this paper should not take precedence over papers that take more radical or more risky approaches to solve the same problem, but I still think it is valuable and might lead to important generalizations in the future. Therefore, even at the risk of publishing just incremental research, I am in favour of its publication.

### **Areas for Improvement**