DidYouMean: A Novel, Fast Error Recovery Mechanism for Mobile Web Navigation

Stefan Hadjis[1] **Courtney Hill**[2] **Mohammad Edghaim**[2]

[1]Department of Engineering Science, University of Toronto
[2]Department of Computer Science, University of Toronto
{stefan.hadjis, courtney.hill, m.edghaim}@mail.utoronto.ca

ABSTRACT
We present *DidYouMean*, a novel mechanism for recovering from mis-clicked links when navigating in mobile web browsers. We conducted experiments in which users navigated through realistic web pages on a smartphone browser and observed that *DidYouMean* achieves a 48% reduction in error correction time against Android Chrome and Firefox. Additional analysis including total time vs. mis-clicked link frequency and qualitative user experience feedback is provided. All data, software and high-resolution images can be found at the following URL: ow.ly/uQ6CK

Author Keywords
Mobile web navigation; error recovery; fat finger problem;

ACM Classification Keywords
H.5.2. User Interfaces: Input Devices and Strategies, Interaction Styles; D.2.2: User Interfaces

1. INTRODUCTION
While mobile versions of websites are becoming more common, navigation on mobile devices is still a problem for many websites. Some have their own apps but it is burdensome for users to download an app for every website. Indeed, Tossel et al [7] report that only 25% of participants prefer smartphones over desktops for web navigation, and that some users avoid smartphone web usage due to poorly optimized content and browsers.

Figure 1 shows a screenshot of reddit.com. Links are small and close together, requiring users to either manually zoom or click the vicinity of a desired link and hope that the intended link is selected. This challenge of limited precision with touch-screen devices is known as the *fat finger problem* and much recent research focuses on addressing the problem of unintended selection [1, 3, 4, 5, 6]. Vogel et al. [1] developed an interface for giving users increased precision prior to making a selection, at the expense of increased selection time. Widgor et al. [3] explored a different approach, providing no increased precision but giving users instant feedback when an invalid selection is made. Chen-Hsiang et al. [6] also identify the fat finger problem and limited display size as primary challenges for mobile websites, and introduce speech recognition as an alternative navigation technique. However, this raises navigation time significantly.

Commercial navigation techniques for mobile browsers follow Vogel's work. On Android smartphone devices, Google's Chrome browser has a built-in navigation mechanism which magnifies nearby links (Figure 2). However no browsers emphasize error *correction* rather than prevention, and this motivates a study of whether navigation time can be reduced not by giving users increased precision but rather allowing users to make errors and correcting as quickly as possible.

In this work we present DidYouMean, a novel error correction method for mobile web navigation, and compare it against existing methods for web navigation in terms of total navigation time and number of mis-clicked links. The remainder of the paper is organized as follows: Section 2 summarizes the methods tested and lists our hypotheses. Section 3 describes our experimental design. Section 4 presents results and Section 5 provides analysis.

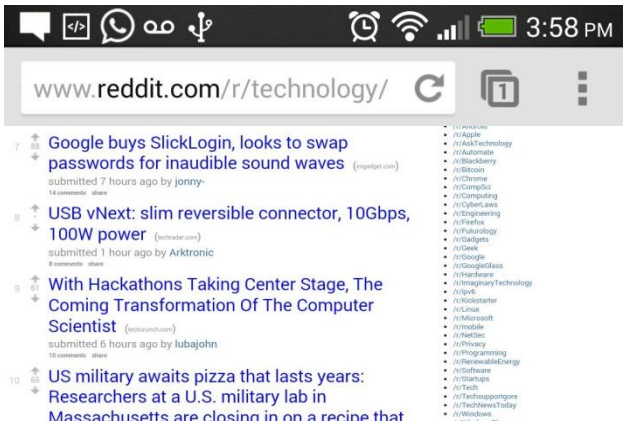


Figure 1. Partial screenshot of a mobile web browser displaying a non-mobile website (here, reddit.com). Small links make navigation challenging.

2. RESEARCH QUESTION AND HYPOTHESIS
This work introduces a novel mechanism to aid users in navigating mobile web pages with links that are close together. We compare three navigation techniques. They are introduced here and further described in Section 3.1.

2.1 Summary of Navigation Techniques
This work studies three navigation techniques:

1. Control (Mozilla Firefox)
Our control is the Android Firefox browser, which provides no aid to the user. If a user selects one of multiple nearby links and Firefox navigates to an unintended link, the user must manually go back in the browser and re-select.

2. Android Chrome's Built-In Magnification
On Android devices, when a user clicks in the vicinity of multiple links, the Google Chrome browser automatically magnifies the clicked area, allowing the user to select the desired link. This is shown in Figure 2.

Figure 2. A screenshot of Google Chrome's magnification tool for nearby links

3. Our novel DidYouMean method
Rather than magnify a region with multiple nearby links, DidYouMean navigates to a single link, just like the control. On the newly loaded page however, a prompt is then displayed asking the user if they meant a different page, and providing a link to that other page.

2.2 Research Question
When navigating in mobile browsers, how are navigation time (in seconds) and number of errors (mis-clicked links) by two navigation aid techniques? Specifically, the techniques studied are: (1) Chrome's link magnification tool, (2) Our DidYouMean error correction mechanism, and (3) No navigational aid mechanism (FireFox control).

2.3 Error Rate Hypothesis
$H0_{ERROR}$ - For web pages that contain close-together links in a mobile browser, there is no difference in the number of mis-clicked links when using various navigation aids, i.e. navigation error frequency is the same with Chrome's magnification technique, our "Did you mean…." technique, and the control (no technique).

For our alternative hypothesis, we predict that Chrome will have the fewest errors of the three methods. This is because Chrome provides magnification (navigation aid) prior to clicking a link. On the other hand, our method and the control (no aid) both navigate to the link the user clicked, even if their click was close to two links, and so we predict error rate to be similar in these two cases, although we predict our method to have fewer errors than the control because in the worst case with the control, if a link is incorrectly clicked then the user must go back and try again, and therefore it is possible to make a second mistake. For our method, this is not possible as only one other link is presented to the user.

$H1_{ERROR}$ – Error frequency (lowest to highest) will be: Chrome < Our method < Control.

2.4 Navigation Time Hypothesis
$H0_{TIME}$ - There is no difference in mobile browser navigation time when using various navigation aids, i.e. navigation time is the same with Chrome's magnification technique, our "Did you mean…?" technique, and the control (no technique).

For our alternative hypothesis, we predict that both our method and Chrome will beat the control in terms of navigation time, because with both methods going "back" in the browser is never necessary. We believe that our method will result in lower navigation time compared to the control because in the case that the user clicked the correct link the navigation time would the same, and if they clicked the incorrect link then they have a direct link to the other page (as opposed to going back and potentially clicking incorrectly again).

Comparing our method against Chrome, while we hypothesized earlier that Chrome would have fewer errors (because it magnifies prior to selecting a link), we believe that if there is a statistically significant difference in navigation time, our method will be faster. This is because our method eliminates the need for magnification and an extra click prior to link selection. In the event that the user clicked the correct page, then they finish sooner than with Chrome. In the event that they clicked the incorrect page, then they have a direct link to the correct page and that extra click corresponds to the extra click needed in Chrome. Therefore our method always involves the same or fewer inputs (clicks) from the user as Chrome.

One aspect of our method which may be slower is that the user may need to wait for the new page to load before realizing their error. While the DidYouMean link is presented to the user immediately (prior to page loading), the time to realize their error and correct it may add delay not present with Chrome. Because the performance of our method depends on this reaction time and error frequency, and therefore complex to predict, we believe the experiment is interesting and worthwhile to perform.

$H1_{TIME}$ – Total navigation time (lowest to highest) will be: Our method < Chrome < Control.

3. METHODOLOGY

3.1 Apparatus

Hardware

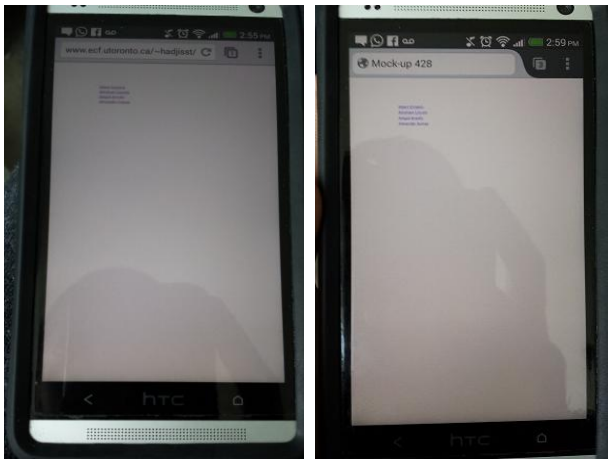
The experiment was conducted on the HTC One, running Android's 4.4 KitKat OS. The device has a quad-core 1.7 GHz processor. It has a (137.4mm x 68.2mm) display producing an effective resolution of 1080 x 1920 (approximately 469 ppi pixel density).

Browser Software

Two mobile browser applications for the smartphone were used, Google Chrome version 32.0 and Mozilla Firefox version 27.0. A photo of the smartphone running each browser is shown in Figure 3, with an example trial from our experiment.

Experimental Software

Full experimental procedures are described in section 3.4. In brief, the experiment consists of multiple trials where the goal of each trial is to select one of four nearby links. Rather than use a website that participants may be familiar with, we created our own webpage for the experiment. Viewing the webpage on the Firefox and Chrome browsers then allows us to test two of our three techniques (control and magnification respectively).



Google Chrome **Mozilla Firefox**

Figure 3. The HTC One Android smartphone and browsers used in this experiment

To implement the DidYouMean method, we elected not to create a third browser, but rather create a separate version of our experimental webpage which *simulates* a browser containing DidYouMean error correction.

DidYouMean functions like the control in that if two nearby links are clicked, the browser navigates to whatever page is clicked first. Additionally however, the browser stores the other link and briefly displays it, asking the user if they made an error. For example if the user intends to click on a link Abraham Lincoln, but accidentally clicks the link John F. Kennedy which is nearby, the prompt in Figure 4 will be displayed, presenting a link to Abraham Lincoln.

Figure 4. A cropped screenshot of our *DidYouMean* technique on mobile Chrome

To simulate this, in order to determine whether a user clicked on a link definitively or in between two links, the four links are placed within an html tag. Then, rather than simply select 1 of 4 links, the page calculates the y coordinate of the click within the div. Based on this y coordinate, the page determines whether a single link was clicked, or whether the click was ambiguous and could refer to two links. This is illustrated in figure 5. For each y-coordinate range, it shows (1) which page the user is redirected to, and (2) whether a DidYouMean prompt is displayed on the redirected page. All range intervals are ¼, ½ or ¾ the height of a link. For example, the second row of the table shows Kennedy/Lincoln: this means that clicking that ¼-height range would load Kennedy.html, but a prompt to Lincoln.html would be shown, as in Figure 4. This is illustrated by Figure 5.

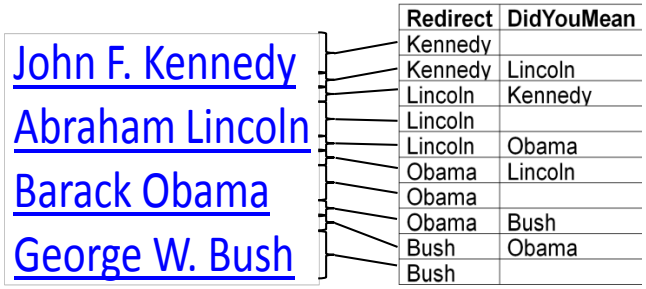


Figure 5. An illustration of how DidYouMean is simulated in HTML.

To the user, the div in Figure 5 looks like 4 links. The gray box outline is not included in the actual webpage. While the adoption of DidYouMean as a usable product would require its integration into a browser, to the user out experimental interface in Figure 4 provides an identical experience.

Data collection is built into this website and is further described in section 3.6.

3.2 Participants

The survey in Appendix 1 was administered to all participants prior to their participation. We selected 24 participants, 8 for each group. Our selection criterion mimics that of Vogel and Baudisch [1]. The twelve volunteers in their experiments ranged in age from 24 to 41 and all were from the same institution. Moreover their work claims that all participants had some experience with pen-based PDAs. Similarly, our work involved only participants with:

- At least 1 year of smartphone experience
- Age range 21-40
- Post-secondary students, or individuals having a post-secondary education
- Balanced male/female ratio in each group

Our choices are also similar to those made by Raneburger et. al. [2], in which 20 participants (10 for the iPodTouch and 10 for the Motorola Xoom) were hired of roughly the same age group (between 21 and 35 years old, with an average of 26 year) and of about the same level of education (students).

While this experiment tests smartphone browser navigation mechanisms and is independent of the smartphone OS, one method we are comparing against (built-in link magnification) is exclusive to Android phones. We therefore ensured that all users of Android were evenly balanced between our groups. Each group consisted of 4 Android users and 4 who had never used Android and had used iPhone. Due to the makeup of our participant pool each group had 3 females and 5 males and was balanced for age and education level. Participants were volunteers, and selected from within in classes and social groups of the experimenters.

3.3 Experimental Design

Our study contains one independent variable (navigation technique) with three levels. Each level is a different browser technique to aid navigation, each previously described in sections 2.1 and 3.1:

1. Chrome's built-in link magnification
2. No assistance technique (control)
3. Our novel *DidYouMean* technique

We used three groups, one with each technique (between subject design). There were 8 participants in each group. We justify this experimental design by comparing to the experiment in Vogel's paper with Microsoft Research [1]. They used 12 participants for a 2x3x6 within-subject design. As their design is more complicated and counter-balancing is needed for within-subjects, we justify that our experiment (3x1, between subjects) requires fewer per group.

3.4 Tasks and Procedures

Our first consideration is how participants hold the phone. Raneburger et. al. [2] had the devices fixed to the table, which while consistent across users does not reflect how most handheld devices are used. Vogel's [1] participants used two hands to operate the device, holding the device with their non-dominant hand and pointing with the index finger of their dominant hand. Two participants of Vogel and Baudisch [1] were left handed but no separate consideration was given to these participants. For external validity, we used the experimental design of Vogel and Baudisch [1] and instructed users to hold the device in their non-dominant hand while navigating using the dominant hand.

Participants were first shown a demonstration of 3 trials completed by the experimenters. For the participants in the groups using our novel method or the link magnification method they were each shown how the method works by the demonstrator purposely making a mistake to show them the technique of using the "Did you mean" link or the magnification.

Participants were seated on a chair, and instructed to hold the device as described above. They were then given a smartphone to complete a training of 3 trials. Once they had completed the training the experimenter then navigated to the start page and the smartphone was handed back to the participant.

The start page contains instructions to read and can be seen at our link in the Abstract. Once participants clicked "Begin", the experiment started. Participants were shown a screen like the one in Figure 6. This illustrates one of 24 trials. In each trial the following things vary:

1. The order of the 4 links
2. The position of the group of 4 links on the screen

To eliminate any unintentional independent variables, the 4 links (text) were always the same, and the goal was always to click Abraham Lincoln. The trial ended once Lincoln is clicked.

There are 24 permutations of the ordering of the 4 links. Therefore 24 trials were performed, with a 3 second break between each trial. Note that while the links are intentionally close together to mimic the navigation challenge in many websites, an existing site is not used but rather our own simple website (white background). This removed any outside factors, such as the test subjects' familiarity with the material. Another advantage of making our own website is that it allowed data recording to be built-in (see section 6).

Following each trial there was a three second pause, and the users were then presented with the next trial screen. Once Abraham Lincoln had been selected the 24th time, the experiment ended. We thanked the participants, explained the goals of our research, answered any questions, and give them the survey described in section 4.5 (survey is online). Note that users were able to elect to stop at any time, and were told this prior to the experiment beginning.

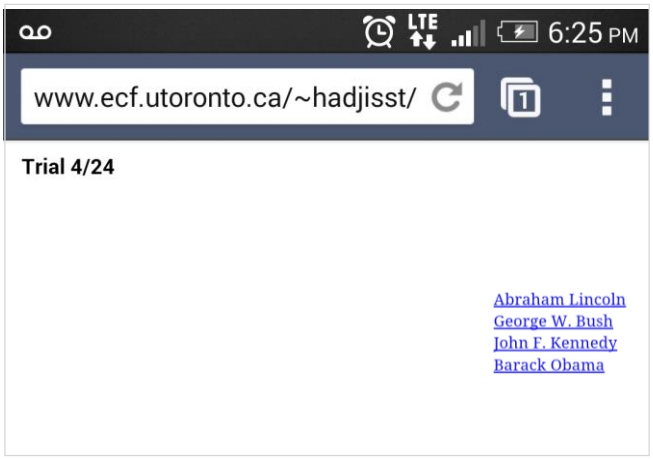


Figure 6. A cropped screenshot of one trial

3.5 Measures

Dependent variables must be useful in that they must give insight to the proposed research questions. The dependent variables chosen for this study reflect real-world usability problems of smartphone users, i.e. navigating through web pages quickly and without mis-clicked links. Drawing from our own experience, it is frustrating to click an incorrect link and have to navigate back. Also because many web pages are not suitable for mobile browsers and contain links which are close together, navigation time is often prolonged by the need for magnification, an extra step in the navigation process.

As a result, we measured two dependent variables: the total time to complete the task outlined in section 3.4 and number of errors (navigations to incorrect pages). We also recorded errors/time per trial, and administered a user experience survey upon completion of the experiment to gauge user frustration (see section 3.6).

3.6 Data Collection

Time and error were automatically collected for each trial and printed to a file. We measured time using javascript by recording the start and end time of each trial (see code at the URL we provided). We measured the number of mis-clicked links by incrementing a count every time a page was visited other than Lincoln. We also counted the number of times a trial was completed through a DidYouMean prompt.

Finally, we administered a survey upon the completion of the 24 trials. Each user was asked to rate the intuition of their experience, their frustration, and how satisfied they were with the navigation technique from 1 to 5, and to provide any additional comments. The survey is similar to that administered by Raneburger et. al [2].

4. RESULTS AND DISCUSSION

4.1 Statistical Analysis Procedure

Statistical analysis was performed to test $H0_{TIME}$ and $H0_{ERROR}$. As there is one independent variable and it has more than two levels (three in this experiment), a one-way ANOVA was used with Tukey's familywise error rate (FWER) for post-hoc comparison. A confidence interval of 95.0 was selected, and the ANOVA was performed using MINITAB Student Release 14.

Despite having tested 24 participants (8 in each group), currently neither the total time nor the error rate results are statistically significant, suggesting that in the coming week further testing must be done to increase statistical power. It was determined that the test was not statistically significant by observing the confidence intervals for each post-hoc pair-wise comparison. Because all comparisons had confidence intervals which contained 0, this means that none of the differences in means are significant, neither for time nor errors. Observing the data, it is clear that this is because there is so much variance, and this is discussed in detail below. Rather than select 8 participants from each group, we will aim for 16 from each for a total of 48 participants (doubling our current statistical power).

In spite of this however, interesting trends were observed which allow us to draw meaningful conclusions. Moreover, while the final paper will omit insignificant results, they are included here only if interesting conclusions or lessons can be drawn. The following section interprets our results, presents results which are statistically significant, and makes suggestions of how to improve the significance of results for $H0_{TIME}$ and $H0_{ERROR}$.

4.2 Hypotheses

Table 1 below summarizes mean experiment completion time for each of the three navigation methods. This is the time to complete all 24 trials, averaged across all 8 participants in that group.

Group	Control	Chrome	DidYouMean
Mean Time (s)	71.36	79.71	76.45
Std. Dev	29.51	28.49	35.20

Table 1. Mean experiment time for each group

These means are *not* statistically significant ($p = 0.842$), but are included to show the *variability* in the data: logically, DidYouMean should only ever perform better than the Control, as it is identical except for a shortcut link. DidYouMean could only take longer if the shortcut is mis-used (leading to additional errors), but this was never the case. This, along with the standard deviations above, suggests that any difference in time attributed to

the methods themselves is eclipsed by the noise in the data. Similar results were observed for errors (p=0.46):

Group	Control	Chrome	DidYouMean
Mean #Errors	7.38	4.50	6.00
Std. Dev	6.02	3.81	3.25

Table 2. Mean number of errors for each group

Once again the numbers themselves are not significant except insofar as they demonstrate the variance in the data. One result which matches expectation is that Chrome has the least errors, which makes sense because it provides pre-magnification (although this cannot be concluded from Table 2 because $p > 0.05$).

The variance in each group suggests either a flaw in our experimental design or the need for additional test subjects. Indeed, while testing we observed that despite our balancing for education, gender, and smartphone experience, the most important factors correlated with total time and errors were more subtle, and are explored further below.

4.3 Qualitative Observations

All participants were given clear instructions to complete the experiment as quickly as possible, and also shown demonstrations and allowed to attempt trials before beginning. Nevertheless, some users hesitated once the experiment began and were *intimidated* by the small size of the links and their inability to accurately select Abraham Lincoln (one participant said this verbally during her experiment).

Some users took their time selecting the links while others were careless and clicked as quickly as possible, making more errors. Participants were neither instructed to manually zoom or not to zoom, but overall 5/24 participants chose to zoom: 2 while using Chrome, 2 while using the Control and 1 while using DidYouMean.

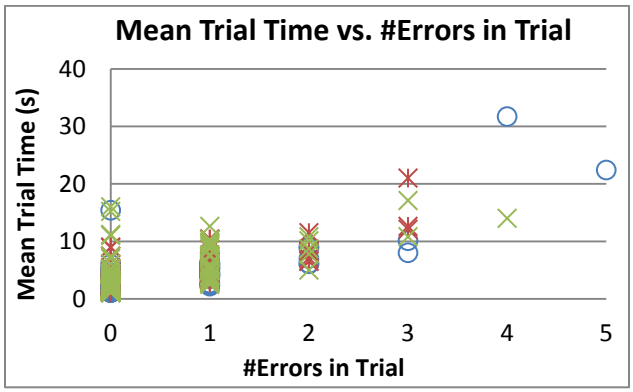
Another consideration was finger size. Although this did not appear in our survey, we noticed that two of the participants tested had fatter fingers, and those two trials had the most errors. The results for these two participants were removed from consideration.

4.4 Quantitative Observations

Statistical analysis concluded no difference between any of the three methods in terms of either error frequency or mean time. However, this is considering only the total times and total errors per participant. Interesting trends can also be observed by analyzing individual trials, and doing so increases experimental power.

Figure 7 shows a scatter plot of the number of errors in a trial vs. the time for that trial. As expected, there is a positive correlation between the number of errors in a trial and the time to complete that trial.

The interesting result comes from further examining which trials were completed using DidYouMean. As part of our automatic data logging, we stored when a trial completed through a DidYouMean Abraham Lincoln link. Of the 576 trials performed, one was thrown out because Lincoln was right at the bottom of the screen (a bug that was later fixed). Of the remaining 575 trials, exactly 100 contained 1 error, 18 contained 2 errors and 10 contained 3 or more errors.



O = Control, ✳ = Chrome, X = DidYouMean

Figure 7. Trial times vs. #Errors in trial (N=575)

Taking the 100 trials with one error (from all navigation techniques), these trials were grouped by whether Lincoln was reached through a DidYouMean prompt (11/100 overall), and the mean trial time was calculated for each group. This result *was* statistically significant ($F_{1,99}=19.55$, $p < 0.001$), with a mean trial time of **3.00s** for errors corrected by DidYouMean and **5.79s** for errors corrected without – a 48% reduction in error correction time. This is equivalent to *rejecting* the following null hypothesis, in favor of the DidYouMean method:

$H0_{TRIAL\ TIME}$ – For web pages that contain close-together links in a mobile browser, if a user navigates to an incorrect link there is no reduction in error correction time using the DidYouMean method. Error correction time is the time it takes to re-navigate to the intended link.

The results are also displayed in Figure 8, where the data is further split by navigation method. The statistically significant result is obtained by comparing columns 1-3 against column 4.

4.5 User Experience

Finally the results of the user experience survey are analyzed. While the means across the various methods are again not statistically significant for frustration, intuition or satisfaction, qualitative observations were made.

User experience varied widely, even with the same method, and is highly correlated with performance. For example using Chrome, one participant commented that they didn't like "having to click twice when [Chrome] zoomed in, even though [they were] certain [they] touched it properly." This user had a time of 87.9s, 87 errors (both above average) and rated the frustration 4/5.

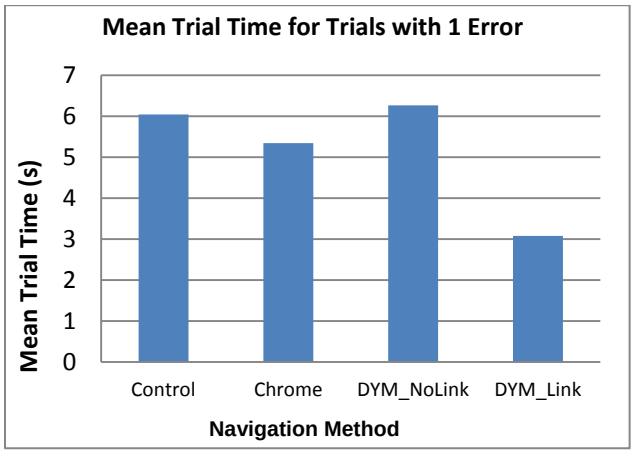


Figure 8. A comparison of trial times for trials with 1 error. The first three columns complete the trail normally, while the fourth column completes the trial using the DidYouMean technique. DYM_NoLink refers to the DidYouMean method being used, but with no link provided to the user (because they did not click in the viscinity of two links). DYM_Link refers to the DidYouMean method being successfully used to complete that trial. (N=100)

Another Chrome user finished in 55 seconds, had 0 errors, and rated frustration 1/5 (this user had never owned an Android device or used the magnification feature before, but simply found the method intuitive to use).

Figure 9 plots frustration against number of errors. This corroborates the notion that navigating on non-mobile websites can be challenging for all smartphone users, regardless of experience.

5. CONCLUSION

Our experiments show that DidYouMean, our novel error correction method, reduces error correction time during mobile web navigation by 48% ($p < 0.001$). We also showed that user frustration increases with number of mis-clicked links, and observed that some participants were displeased or intimidated by the link size in our experiment (of identical sizes to many commercial websites). This motivates future work in the field to continue to enhance error recovery and navigation in mobile web browsing.

6. ACKNOWLEDGMENTS

We would like to thank the course staff, Olivier St-Cyr and Dustin Freeman, for their advice.

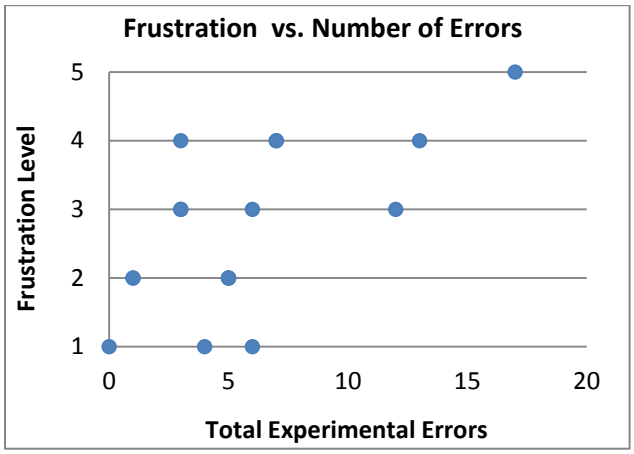


Figure 9. User frustration increases with the number of errors. (N=24)

REFERENCES

1. Vogel, D., Baudisch, P. (2007). Shift: A technique for operating pen-based interfaces using touch. In Proc. of CHI 2007, 657-666.
2. Raneburger, D., Popp, R., Alonso-Rios, D., Kaind, H., Falb, J. (2013). A user study with GUIs tailored for smartphones and tablet PCs. In Proc. of SMC 2013; Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics, 3727-3732.
3. Wigdor, D., Williams, S., Cronin, M., Levy, R., White, K., Mazeev, M., Benko, H. (2009). Ripples: utilizing per-contact visualizations to improve user interaction with touch displays. In Proc. of UIST 2009, 3-9
4. Benko, H., Wilson, A.D., Baudisch, P. (2006). Precise selection techniques for multi-touch screens. In Proc. of CHI 2006, 1263-1272
5. Potter, R., Weldon, L., Shneiderman, B. (1988). Improving the accuracy of touch screens: an experimental evaluation of three strategies. In Proc. of CHI 1988, 27-32.
6. Chen-Hsiang, Y., Miller, R. C. (2011). Enhancing mobile browsing and reading. In Proc. of CHI 2011, 1783-1788
7. Tossel, C C., Kortum, P., Rahmati, A., Shepard, C., Zhong, L. (2012) Characterizing web use on smartphones. In Proc. of CHI 2012, 2769-2778