A User Study with GUIs Tailored for Smartphones

David Raneburger[1], David Alonso-Ríos[2], Roman Popp[1],
Hermann Kaindl[1] and Jürgen Falb[1]

[1]Institute of Computer Technology, Vienna University of Technology, Vienna, Austria
`{raneburger, popp, kaindl, falb}@ict.tuwien.ac.at`
[2]Department of Computer Science, Faculty of Informatics, University of A Coruña, Spain
`dalonso@udc.es`

Abstract. Web-based graphical user interfaces (GUIs) are mostly not tailored for small devices with touchscreens, such as smartphones. There is little scientific evidence on the conditions where additional taps for navigation are better or scrolling. Therefore, we conducted a user study in which we evaluated different ways of tailoring a GUI for a smartphone. Each participant performed the same task with two different layouts of the same GUI. We collected quantitative data through measuring task completion time and error rates, as well as qualitative data through subjective questionnaires. The main result is that minimizing the number of taps is important on a smartphone. Users performed significantly better when they could scroll (vertically), instead of tapping on widget elements (tabs). This preference was also reflected in their subjective opinions.

Keywords: Usability, device-tailored GUI, small touchscreen, user study.

1   Introduction

Schmiedl et al. showed in an experiment that users completed tasks on small mobile devices 30–40% faster on Websites tailored for such devices than on Websites designed for desktop PCs [1]. Smartphones and PCs have *two* major differences in this respect: size and input method (i.e., touch vs. fine-grained mouse pointing). What is then the best approach for tailoring Web-based GUIs for touch-based smartphones? Web-based usability guidelines typically agree that unnecessary scrolling [2, 3], and in particular horizontal scrolling [4], should be avoided on desktop devices. Recently published guidelines for touch-based devices, in contrast, state that scrolling is part of such a device's user experience and, therefore, not bad at all [5, 6]. The different user experience originates in the more "natural" swiping gesture, in comparison to dragging a scroll-bar with a mouse, or turning its scroll-wheel. These guidelines, however, do not contain hints on whether vertical scrolling is preferable to tabs.

Since we are not aware of any empirical studies that investigate these questions, we designed and performed a user study evaluating an intentionally simple Web-based GUI using different layout approaches for a process-oriented application (flight booking). Our focus is on vertical scrolling vs. tapping, while two-dimensional scrolling and horizontal browsing is out of scope. Based on the collected data, we present both a statistical and a subjective evaluation of the results.

2 Related Work

Different GUI designs and input methods for touch-based mobile phones are investigated in [7], focusing on overall interface layout (scrollable vs. tabbed GUI), input mode (keyboard vs. tapping through a modal dialog) and menu access (device menu vs. context menu). The two GUIs under study in [7] differed in all these characteristics, which were investigated through an assigned task. Nevertheless, it is hard to prove that only the investigated characteristic accounts for slower performance and that there is no influence from other characteristics. Our study, in contrast, is based on GUIs that differ only in one characteristic – the layout. In addition, our GUIs are HTML pages rendered on Web browsers, which display them similarly on different devices. This helps to minimize the influence that familiarity with specific devices may have on users' perceptions of usability and correlated time-on-task [8].

Recently, Lasch et al. investigated touchscreen scrolling techniques and their effect on car driver distraction [9]. Their results indicate that *swiping* was less distracting while driving than *traditional scroll buttons* or *kinetic scrolling*. The latter uses the same gesture as swiping, but it accelerates the menu. The menu stops automatically after deceleration or if the user taps on the screen. So, this work investigates three different ways of scrolling in a context where distraction is a safety risk. Our primary concern, in contrast, is to achieve good usability in a context where the user concentrates on a specific task. Therefore, we investigated whether scrolling was preferable to tabbed panels, or vice versa. Nevertheless, the results of both studies complement each other.

Comparable user studies have also been performed in the past, but only with desktop devices [10–12], old cell-phones [13] or PDAs [14], so their results are not necessarily valid for current smartphones. Previous user studies on touch-based devices (e.g., PDA or smartphone) assumed that the user does not like to scroll and, therefore, primarily focused on investigating strategies to avoid or minimize scrolling [15–17]. Jones et al. [18] found that users preferred vertical over two-dimensional scrolling on small screens. Our results extend these findings for current touch-based devices through the comparison of two different layouts tailored for smartphones.
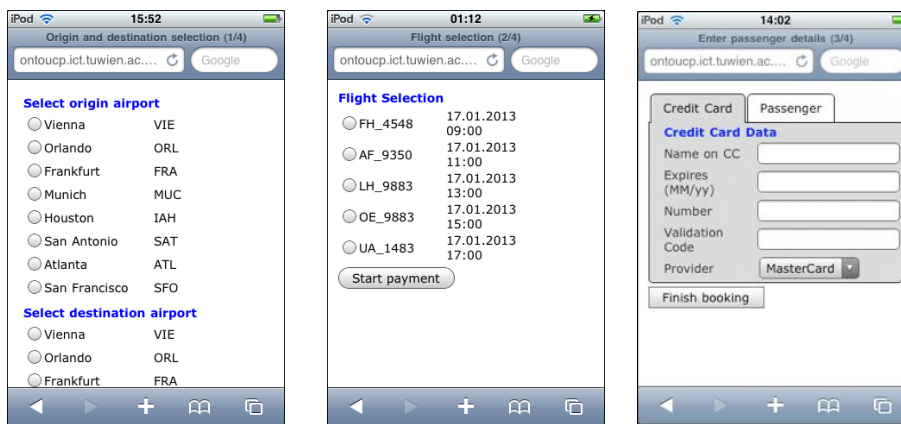
3 User Study Design

In order to evaluate the different GUIs, we combined quantitative data (task completion time and error rate) with subjective data (user questionnaires) like Jones et al. [15] and Buchanan et al. [13].

In our **study setup**, we tested two different layouts for an otherwise identical user interface. Both layouts were tailored for a small device each (a 320×480px iPod Touch / iPhone up to version 4S, which all show the GUI in the same way), more precisely in portrait mode. However, they followed different approaches (tabs vs. vertical scrolling). The application was a simplified flight-booking scenario consisting of HTML pages as follows:

- **Screen 1.** Origin and destination airports, flight date.
- **Screen 2.** List of flights available on that date.
- **Screen 3.** Billing information.

Fig. 1 shows screenshots of these screens, where Screens 1 and 3 are the ones tailored, while Screen 2 is identical for both layouts, as it is small enough to fit perfectly. Due to space limitations, Fig. 1 shows Screen 1 only in its form for vertical scrolling, and Screen 3 only in its tabbed form. During the test each layout was, of course, the same from the beginning to the end of the task – there were no changes of layout in the middle of a task.

(a) Vertical scroll-based layout (Screen 1)

(b) Both layouts (Screen 2)

(c) Tab-based layout (Screen 3)

Fig. 1. Screens tailored for iPod Touch / iPhone up to version 4S.

We split the **participants** into two groups, reversing the order in which the two layouts were tested. At the end of each test, we collected their subjective opinions through a questionnaire and, sometimes, a brief and informal interview. Overall, we hired 30 participants as summarized in Table 1.

Table 1: Participant Details.

Age:	18–39 years
Gender:	3 female and 27 male
Educational level:	28 students and 2 assistants with doctoral degrees
Use of mobile devices:	56.7% regular, 40% sometimes but never owned one, 3.3% never
Online flight booking:	20% regular, 40% sometimes and 40% never

At the beginning of each experiment we informed the participants about the content of the study and the **procedure**. We emphasized that the point was to test the

GUIs, not the skill of the participants. We asked them to fill in some background data and to give their consent to filming their hands operating the iPod Touch (for subsequent video annotation) and recording their voices. We fixed the iPod Touch to the table, setting it in portrait mode for the following reasons: the GUIs have been tailored for this mode; we wanted all the participants to use the same mode; filming would have been difficult if the participants were holding the iPod Touch.

We gave the following scenario description to the participants:

"Imagine it is Tuesday 14/02/2012, 11:55am and your boss Mr. Huber tells you to book a flight ticket for his wife as quickly as possible. Mrs. Huber is already waiting at the airport!

Book a flight from *Munich* to *Atlanta* on *02/14/2012* at *1pm* for Mrs. *Anna Huber* (age 47). Pay for it using her husband's (Max Huber's) *VISA Credit Card* with the number: *1258 8569 7532 1569* (validation code: *354*) and the expiration date *12/14*."

We left a printed version on the table, where it could be easily seen, so that participants did not have to remember the information. We emphasized the urgency of the task, so that the participants would concentrate on the task and finish it quickly and realistically.

After the participants completed both tasks, we collected their subjective opinions on several usability-related issues through a questionnaire and finally asked them if they had further comments, in which case we made a brief interview.

For our **analysis**, we were primarily interested in the correlations between task time / error rate and the type of layout. An independent variable was the GUI, with these possible values: tab-based (T-UI) and vertical scroll-based (V-UI). This variable is dichotomous, as we only compared two GUIs per run. A second independent variable was the order in which the two layouts were tested. The dependent variables were adjusted task completion time and error rate. For the *adjusted task completion* analysis, we measured the total time spent on a screen. We also measured the time needed for text input, screen loading, form submission, validation alerts and error messages, which are all layout-independent and may introduce bias into the results. Therefore, we subtracted all these times from the total time, resulting in *adjusted task time*.

For the *error rate* analysis we measured the number of errors each participant made on a given screen. Errors in our study occurred when the participant hit "submit" without entering some required information.

As the type of GUI is dichotomous and the calculated time is on an interval scale, we calculated the *point-biserial Pearson correlation coefficients*. Our tests also showed that the variables are nearly normally distributed and thus satisfy the prerequisites of the point-biserial Pearson correlation. Note that the difference to the more common *t-test* is that the correlation covers coherence, while the t-test deals with differences of means. The corresponding **null hypotheses** are correlation hypotheses as follows:

- $NH_{VT,T}$: *There is no statistically significant correlation (p-value = 0.05) between the adjusted task time and the type of GUI: V-UI and T-UI.*
- $NH_{VT,E}$: *There is no statistically significant correlation (p-value = 0.05) between the error rate and the type of GUI: V-UI and T-UI.*

In order to extract task completion times and error rates from our video recordings of the tests, we used the video-annotation tool Anvil[1], and converted the files to CSV in order to analyze the data with SPSS.

The participants were also given a **subjective usability questionnaire** for collecting information on how they perceived usability. With some adaptations, our definition of usability was based on [19], which aims to synthesize the best-known usability definitions in the literature. We discarded those usability characteristics that were either not relevant or identical on all layouts (e.g., security) and tried to phrase the usability criteria as short and self-explanatory questions, for which we also consulted the USE Questionnaire[2], the W3C's WAI[3], the Software Usability Measurement Inventory[4] (SUMI) and the Cognitive Dimensions framework [20].

For each question, the participant had to state the preferred layout on a Likert scale with preference being "extreme", "strong", "moderate", and "equal". The participants were explicitly asked to focus on the differences between layouts, rather than evaluate the usability of each layout independently. The reason is that for our study, which compares different layouts, their difference is relevant. Measuring usability on an absolute scale is more difficult and not necessary for our study.

4 Results

Table 2 summarizes the statistical results for the **adjusted task time analysis**. The positive correlations in Table 2 show that vertical scrolling performs better than the tab-based layout, with highly significant results. The sum of times for all screens (minus typing time, etc.; see above) is on average 25.81 seconds for the vertical scroll-based GUI (V-UI) and 36.79 seconds for the tab-based GUI (T-UI).

Table 2. Correlation between adjusted task time on a screen and its layout.

	Pearson Corr.	Sig. (1-tailed)	*V-UI* av. time	*T-UI* av. time
Time Screen1 × GUI	**0.35**	0.003	12.79s	17.98s
Time Screen2 × GUI	0.12	0.189	8.17s	9.50s
Time Screen3 × GUI	**0.43**	0.000	4.85s	9.31s

The null hypothesis $NH_{VT,T}$ can be rejected for this experiment, i.e., the adjusted task time using V-UI is significantly smaller than using T-UI in Screen 1 and Screen 3. In fact, it took 54% longer to operate T-UI.

Table 3 shows the **correlation between the error rate and the two layouts** (V-UI and T-UI), together with the average number of errors for each screen. For Screens 1 and 3, the error rate is significantly smaller. On Screen 2, however, there were no errors, thus no correlation could be calculated. So, the null hypothesis $NH_{VT,E}$ can also be rejected for this experiment. That is, the error rate for V-UI is significantly smaller

---

[1] http://www.anvil-software.org/
[2] http://www.stcsig.org/usability/newsletter/0110_measuring_with_use.html
[3] http://www.w3.org/WAI/EO/Drafts/UCD/questions.html
[4] http://sumi.ucc.ie/en/

than that of T-UI in Screens 1 and 3. Actually, there were no errors for V-UI – all the errors can be attributed to the participants hitting "submit" before changing tabs, so this could be an argument against using tabs.

Table 3. Correlation between the number of errors on a screen and its layout.

	Pearson Corr.	Sig. (1-tailed)	*V-UI* av. errors	*T-UI* av. errors
Errors Screen1 × GUI	**0.363**	0.002	0	0.23
Errors Screen2 × GUI	-	-	0	0
Errors Screen3 × GUI	**0.285**	0.014	0	0.2

Table 4 presents the **results of the subjective questionnaire**. In their overall assessment (i.e., the last questionnaire item), the majority (60%) of the participants preferred the vertical scroll-based GUI, but 30% preferred the tab-based GUI. The answers to the other questions varied much. There was only one criterion for which T-UI was widely preferred: visual attractiveness.

The error rate result was confirmed by participants in interviews, where they stated that it is easy to forget to switch tabs before hitting "submit" even when they were aware that there were two tabs. As for V-UI, the questionnaire shows that it was considered much more intuitive to navigate and to interact with, and slightly less demanding and more efficient to use in general. However, opinions were sharply divided on which GUI made information more visible and which GUI lent itself more to how the participants like to work. It is worth noting that, whenever opinion was divided, analyzing the data showed no clear relationship between the answers and the background of the participants, including whether they owned such a device or not.

5   Conclusion and Future Work

This paper reports on a user study with GUIs specifically tailored for smartphones. We used task time analysis, error rate measurement, subjective questionnaires and interviews. Participants performed significantly better when scrolling vertically, as opposed to tapping on widget elements (tabs). This preference was also reflected in their subjective opinions. The study also suggests that minimizing the number of taps is important on a smartphone.

We conjecture that our study results are not limited to such a simplified flight-booking application as the one used for two reasons. First, we defined a scenario that reduced the user task to entering pre-given data, without having to decide about actually booking a particular flight or not. Second, we used a very simple prototypical application that allowed only for exactly the interaction that was required to complete the given task, without alternatives. This also helped to keep the user focus on the interaction and allowed all users to complete the task successfully.

In future work, we plan to investigate related issues, and e.g., to study variable-sized lists, since it is not clear if vertical lists of *any* length are necessarily better than a tab-based layout. There may be a balance involved between motor performance and understanding the content.

Table 4. Subjective Questionnaire Results.

Question	V-UI preferred			equal	T-UI preferred		
	extreme	strong	moderate		moderate	strong	extreme
Which interface makes information more visible?	3.33%	33.33%	10%	6.67%	30%	13.33%	3.33%
Which interface makes interaction more intuitive?	13.33%	26.67%	26.67%	30%	3.33%	0%	0%
Which interface makes it easier to figure out what to do next?	20%	36.67%	20%	10%	10%	3.33%	0%
Which interface makes it clearer how to use it?	0%	13.33%	30%	53.33%	3.3%	0%	0%
Which interface lends itself more to how you like to work?	13.33%	23.33%	13.33%	23.33%	10%	13.33%	3.33%
Which interface requires less manual interaction?	6.67%	20%	26.67%	26.67%	13.33%	6.67%	0%
Which interface demands less precision on your part?	3.33%	13.33%	16.67%	56.67%	10%	0%	0%
Which interface demands less time from you?	13.33%	10%	26.67%	33.33%	10%	6.67%	0%
Which interface makes interaction more efficient?	10%	6.67%	26.67%	33.33%	16.67%	6.67%	0%
Which interface is more visually attractive?	0%	0%	6.67%	10%	40%	23.33%	20%
Overall, which interface would you use to book a flight?	16.67%	20%	23.33%	10%	10%	13.33%	6.67%

References

1. Schmiedl, G., Seidl, M., Temper, K.: Mobile phone web browsing: a study on usage and usability of the mobile web. In: Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services. MobileHCI '09, New York, NY, USA, ACM, 70:1–70:2 (2009)
2. Microsoft Corporation: User Experience and Interaction Guidelines for Windows 7 and Windows Vista. (Sept. 2010)

3. Nielsen, J.: Usability Engineering. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1993)
4. Bevan, N.: Guidelines and standards for web usability. In: Proceedings of HCI International 2005, Lawrence Erlbaum (2005)
5. Apple Inc.: iOS Human Interface Guidelines. (Aug. 2012)
6. Microsoft Corporation: Windows 8 User Experience Guidelines. (Aug. 2012)
7. Balagtas-Fernandez, F., Forrai, J., Hussmann, H.: Evaluation of user interface design and input methods for applications on mobile touch screen devices. In Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R., Winckler, M., eds.: Human-Computer Interaction – INTERACT 2009. Volume 5726 of Lecture Notes in Computer Science. Springer Berlin Heidelberg, 243–246 (2009)
8. Suzuki, S., Bellotti, V., Yee, N., John, B.E., Nakao, Y., Asahi, T., Fukuzumi, S.: Variation in importance of time-on-task with familiarity with mobile phone models. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '11, New York, NY, USA, ACM, 2551–2554 (2011)
9. Lasch, A., Kujala, T.: Designing browsing for in-car music player: effects of touch screen scrolling techniques, items per page and screen orientation on driver distraction. In: Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. AutomotiveUI '12, New York, USA, ACM, 41–48 (2012)
10. Harms, I., Schweibenz, W.: Usability engineering methods for the web: results from a usability study. In Knorz, G., ed.: Proceedings des 7. Internationalen Symposiums für Informationswissenschaft (ISI 2000), Saarländische Universitäts und Landesbibliothek, Univ.-Verl. Konstanz, 17–30 (2000)
11. Baker, J.R.: The impact of paging vs. scrolling on reading online text passages (Feb 2003)
12. Peytchev, A., Couper, M.P., Mccabe, S.E., Crawford, S.D.: Web Survey Design: Paging versus Scrolling. Public Opinion Quarterly 70(4), 596–607 (2006)
13. Buchanan, G., Farrant, S., Jones, M., Thimbleby, H., Marsden, G., Pazzani, M.: Improving mobile internet usability. In: Proceedings of the 10th international conference on World Wide Web. WWW '01, New York, NY, USA, ACM, 673–680 (2001)
14. Kim, L., Albers, M.J.: Web design issues when searching for information in a small screen display. In: Proceedings of the 19th annual international conference on Computer documentation. SIGDOC '01, New York, NY, USA, ACM, 193–200 (2001)
15. Jones, M., Marsden, G., Mohd-Nasir, N., Boone, K., Buchanan, G.: Improving web interaction on small displays. Comput. Netw. 31(11-16), 1129–1137 (May 1999)
16. Kaikkonen, A., Roto, V.: Navigating in a mobile XHTML application. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '03, New York, NY, USA, ACM, 329–336 (2003)
17. Churchill, D., Hedberg, J.: Learning object design considerations for small-screen handheld devices. Computers & Education 50(3), 881 – 893 (2008)
18. Jones, S., Jones, M., Marsden, G., Patel, D., Cockburn, A.: An evaluation of integrated zooming and scrolling on small screens. Int. J. Human-Computer Studies 63(3), 271–303 (2005)
19. Alonso-Ríos, D., Vázquez-García, A., Mosqueira-Rey, E., Moret-Bonillo, V.: Usability: A critical analysis and a taxonomy. Int. J. Human-Computer Interaction 26(1) 53–74 (2010)
20. Blackwell, A., Britton, C., Cox, A., Green, T., Gurr, C., Kadoda, G., Kutar, M., Loomes, M., Nehaniv, C., Petre, M., Roast, C., Roe, C., Wong, A., Young, R.: Cognitive dimensions of notations: Design tools for cognitive technology. In Beynon, M., Nehaniv, C., Dautenhahn, K., eds.: Cognitive Technology: Instruments of Mind. Volume 2117 of Lecture Notes in Computer Science. Springer Berlin Heidelberg 325–341 (2001)