Exploring Recipe Interface Layouts:

A Comparative Study of Baking Applications

Philipp Niksch, Lukas Probst, Emilia Szymanska, Mauro Vogel, Yutong Ye, Xinran Yu

ETH Zurich

Fall 2022, Human Computer Interaction

1 Introduction

Human Computer Interaction (HCI) [1], as a subfield of computer science, is the interdisciplinary studies aimed at finding the intrinsic relationship between computing systems and human behavior. To better adopt the concept of HCI and discover suitable user-centered designs, this study focuses on how different mobile application designs could affect users' performance under specific circumstances. Being a process consisting of actions that demand using utensils and ingredients to perform continuous, precise, and complex movements, baking presents itself as one of the most desirable scenarios to examine the viability of using smart devices to assist with human tasks [2][3]. Thus, in order to facilitate users with the process of baking and compare the usability of interfaces, an application with two different recipe layouts was developed.

Prior to this study, desk research across several existing cooking assistant applications, case studies, as well as low and mid-fidelity prototype designs have already been conducted and produced. These prior research suggested that two interface layouts, waterfall vs. step-by-step, are both potentially suitable for presenting recipes to users while they are performing tasks in the scenario of cooking (examples of each layout are shown in Fig. 1).

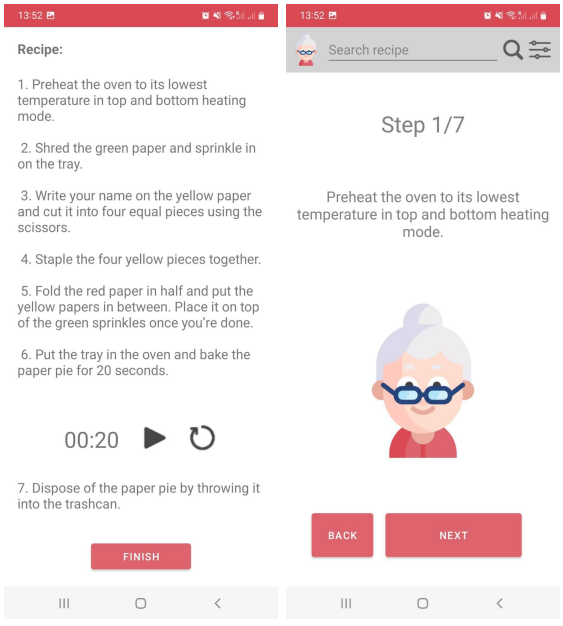

Figure 1: Example layout of waterfall vs. step-by-step view.

The "waterfall" recipe layout displays the entire demonstration on one page, allowing the user to scroll through the recipe from top to bottom. On the other hand, the "step-by-step" recipe layout displays each step of the recipe in order, demonstrating the steps in a more disconnected way, and the user navigates through the steps by clicking the "back" and "next" buttons. These layout structures are also commonly adopted in various commercialized mobile cooking applications, such as *FOOBY* [4] and *Tasty* [5]. However, no particular research in the past has shown which layout design is preferable and ergonomically friendlier for the user to operate during the baking process.

To explore the usability and user experience of these layouts, the following chapters explain the procedure of our investigation on these aspects, in the order of Study Design, Result, Discussion, and Future Work.

2 Study Design

2.1 Study Setting

The study is set to be in a kitchen, where the participant has access to common cooking devices/utensils such as an oven, a cutting board, and a trash can. To better simulate the process of baking, the participant is provided with artificial ingredients, such as papers, small bricks, poker cards, and dice, to perform certain tasks. Tools such as scissors, a pen, a stapler, and a box are also provided to participants to reproduce these baking-related procedures.

An Android device is required to run the application of this study. In addition, a camera needs to be correctly positioned in a way that the entire process of the user study, including the participant's movements and interactions with the smartphone, is thoroughly filmed.

2.2 Variables & Hypotheses

The independent variables of this study are the two interface layouts for prototypes A and B: prototype A has a waterfall scrollable interface, whereas prototype B has a step-by-step slidable interface.

There are a total of four dependent variables: (1) User's screen time; (2) Number of user's screen touches; (3) Number of user's screen glances; (4) User's satisfaction rating on a scale of 1 to 5. The first three variables are objective measures obtained from video recordings of each user study, while the last variable is a subjective measure obtained from the questionnaire that each participant filled out when completing the user study.

Based on the four dependent variables, the null hypotheses are stated as follows:

1. Null Hypothesis #1: there will be no effect of waterfall view vs. step-by-step view on user's screen time.

2. Null Hypothesis #2: there will be no effect of waterfall view vs. step-by-step view on the number of times the user touches the screen.
3. Null Hypothesis #3: there will be no effect of waterfall view vs. step-by-step view on the number of times the user glances at the screen.
4. Null Hypothesis #4: there will be no effect of waterfall view vs. step-by-step view on user's satisfaction rating on a scale of 1 to 5.

2.3 Experimental Procedure

During the study, the participants experimented with two different tasks (1 or 2) with two different interfaces (A or B). That is, instead of performing the same task twice on two different interfaces, they would work on different tasks on each interface, which minimizes the bias of repetitive tasks.

Upon arriving at the experiment location, the participant would first be assigned with a unique integer participant ID. If the ID is odd, the participant would be experimented with *Prototype A: Recipe 1* as well as *Prototype B: Recipe 2*. If the ID is even, the participant would be experimented with *Prototype A: Recipe 2* as well as *Prototype B: Recipe 1*.

Next, the participant would fill out a pre-questionnaire that contains sociodemographics and relevant app-related questions. Afterwards, the participant would be given instructions on things to bear in mind throughout the experiment and be presented with all the tools and ingredients. He/she would then be guided towards the training mode for Prototype A and be asked to spend time exploring the interface. After being familiarized with the interface, the participant would start to work on the designated recipe (1 or 2) and filming would also begin. When all the tasks have been completed, the participant would fill out a questionnaire to review different aspects of the interface and provide their first hand feedback. The exact same process would be performed once again for Prototype B, where the participant would first go through a training mode and then perform the actual tasks as indicated by the recipe. The entire process would be filmed and a questionnaire would be filled out by the participant upon completing his/her designated task (1 or 2).

After the participant had completed the tasks for both prototypes, the experimenter would chat informally with the participant to gather additional suggestions for the application. In the end, the questionnaires would be submitted and the experiment would be closed.

2.4 Participants

Due to the nature of this study, participants were mainly recruited from family, friends, and neighbors.

The participants had a mean age of 29.38 years. The age range of participants spread from 15 to 67 years, but the majority of participants were students aged between 21 and 28 years (Fig. 2), resulting in a high SD of 15.04 years. With regards to their occupation, 62.5% of the participants were college students, while there were also middle school students, apprentices, working professionals, housewives, and retirees involved. The distribution of participants' gender was rather balanced, with 43.8% of the participants being male and 56.3% being female. Regarding their ethnicity, 68.8% of the participants were White and 31.3% were Asian.

The pre-questionnaire also inquired about participants' frequency of recipe usage, their familiarity with smartphone operations, as well as level of English proficiency. 62.5% of the participants stated that they rarely use cooking recipes, whereas 25% used them around weekly. More than half of the participants (56.3%) stated that they are extremely familiar with smartphone operations. Lastly, almost half of the participants (43.8%) claimed that they are proficient in English.

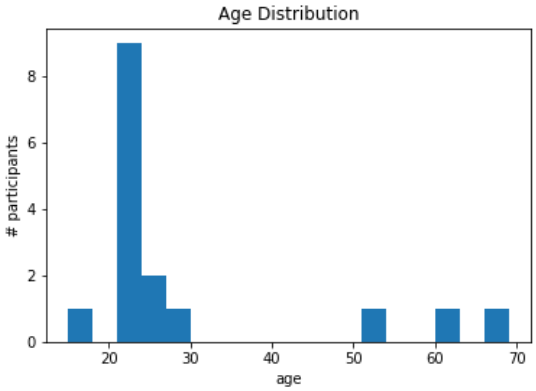

Figure 2: Average screen glances for both recipes and prototypes.

3 Results

3.1 Quantitative Results

3.1.1 Screen Time

Participants had about the same screen time for recipe 1 with just 4 seconds more on average for prototype A. With prototype A for recipe 2, participants on average exhibited the least screen time (87 seconds), but their screen time was still around the same magnitude as other prototype-recipe combinations. However, for prototype B with recipe 2, participants on average spent much more time looking at their screen (108 seconds) in comparison to other prototype-recipe combinations (Fig. 3). This is an indicator that the combination of prototype B with recipe 2 was much slower to use than all other prototype-recipe combinations.

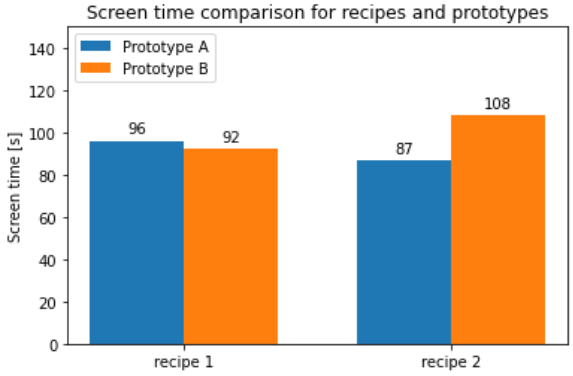

Figure 3: Average screen time for both recipes and prototypes.

3.1.2 Screen Touches

For recipe 1, screen touches were rather the same for both prototypes with just one more screen touch on average for prototype A. For recipe 2, on the other hand, participants touched the screen for prototype B almost twice as often as for prototype A (Fig. 4). These results show that participants interacted much

more with prototype B on recipe 2, which aligns with our expectation given the step-by-step nature of prototype B.

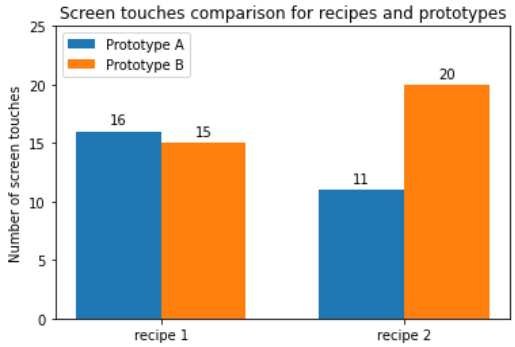

Figure 4: Average screen touches for both recipes and prototypes.

3.1.3 Screen Glances

With recipe 1, participants on average needed 10 screen glances for prototype A and 12 screen glances for prototype B to complete the tasks. With recipe 2, participants on average needed 13 screen glances for prototype A and 16 screen glances for prototype B (Fig. 5). The difference in the number of screen glances could be explained by the more interactive step-by-step view which requires more attention from users than the regular waterfall recipe view. These results indicate that information could be grasped more clearly and absorbed more quickly with prototype A, in comparison to prototype B.

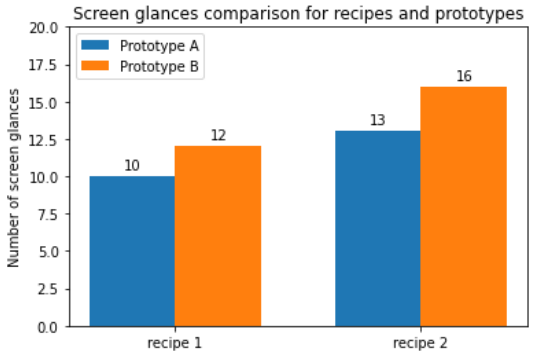

Figure 5: Average screen glances for both recipes and prototypes.

3.2 Qualitative Results

3.2.1 Participants' Feedback

While conducting the user study, we collected some valuable feedback from our participants. Some of them mentioned that their interface preference strongly depended on the recipe and the overall time used. Prototype A was preferred by 62.5% of participants, because it gave them an overview of the length and complexity of the recipe and saved them the effort of switching back and forth to read over the steps again - the act of switching was said to be overly tedious if only some minor detail of the step was forgotten. The remaining 37.5% of participants preferred prototype B instead. Their main reasoning was that they did not have to digest much information at one time and that an individual might have less patience to read through the steps for the waterfall view over and over again. Using an image to complement each step might help participants better track their task completion status, as suggested by some participants.

Some additional improvements were also suggested by participants. One design-specific feedback for the step-by-step interface was that the font size of the text could be bigger, since a lot of space was left "wasted" on each screen, which defeats the purpose of having a step-by-step view. To have a more structured recipe for the waterfall view, certain participants also suggested adding checkboxes to each step of the recipe, in the same way that ingredients and utensils could be ticked.

3.2.2 User Experience

Comparing the user experience for both prototypes, prototype A got a rating greater than or equal to prototype B in all six different categories (Fig. 6). Concluding from the statistics, the performance of prototype A is better than that of prototype B.

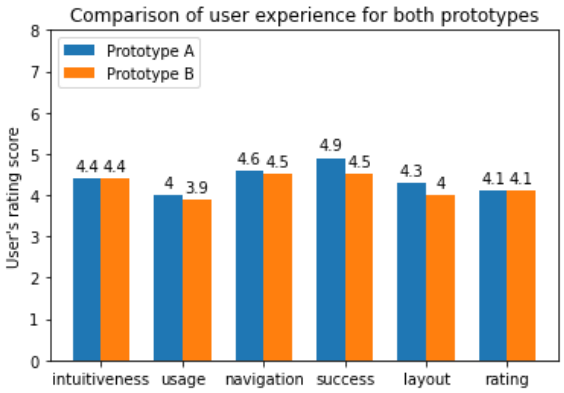

Figure 6: User experience for both prototypes.

3.3 Statistical Testing

For our quantitative measurements of screen time, number of screen glances, and screen touches, we decided to perform statistical tests to investigate the differences between the two prototypes. The samples do not depend on each other as we varied the recipes for each user. Therefore, we chose a two-sample t-test for evaluating the results. Considering the two recipes compared on both prototypes, p-values and decisions on the significance for each experimental variable are presented in Tab. 1-3.

screen time	p-value	significance
recipe 1	0.8176	no
recipe 2	0.2052	no

Table 1: Results for two-sample t-test for screen time.

screen touches	p-value	significance
recipe 1	0.5744	no
recipe 2	0.0126	yes

Table 2: Results for two-sample t-test for screen touches.

screen glances	p-value	significance
recipe 1	0.2748	no
recipe 2	0.3261	no

Table 3: Results for two-sample t-test for screen glances.

4 Discussion

The results show that in the case of recipe 2, prototype A performed significantly better in terms of screen touches in comparison to prototype B, whereas for recipe 1 this was not the case. Taking the qualitative feedback from participants into consideration, we also notice that the waterfall-based prototype was preferred by the majority of the users, although the user experience ratings for different metrics of both prototypes do not vary remarkably.

Taking our statistical tests into consideration, we could determine whether we have the confidence to reject or not reject the Null Hypotheses formulated in *2.2 Variables & Hypothesis* within the Study Design section using an alpha of 0.05. By looking at p-values, we could state that for both recipes, the difference in screen time and screen glances between prototypes A and B was not statistically significant, and thus we fail to reject Null Hypothesis #1 and Null Hypothesis #3. Whether we reject Null Hypothesis #2 or not depends on the recipe (Tab. 2), and therefore we also cannot state the final decision given our current data. Null Hypothesis #4, can not be rejected, given the similar quantitative data as shown in Figure 6.

It should be noted that the results of our user study heavily depend on which participants we selected as well as the order in which the prototypes were tested. For instance, we never started the experiment session with prototype B, which could have had an effect on the results.

5 Limitations

The most notable limitations of this study are the low number of participants (16) and their poor distribution. Most of the participants are college students aged between 21-25 (62.5%). Almost all participants are familiar or very familiar with smartphone operations. On the other hand, however, 75% rarely or never use cooking recipes. Only 25% use them at least around weekly.

Certain limitations follow directly from the experiment design. First of all, our experiments were step-by-step task completion. The tasks were built so that they would resemble realistic baking procedures as much as possible, but after all, they are not real baking steps, so the ability to make inferences and draw conclusions based on these pseudo tasks is inherently limited. The participants also did not face the problems of having dirty hands, and the tasks were much shorter than realistic baking recipes, which could take hours to complete from start to end.

To ease the process of data collection, the participants were asked to lay the phone face down on the kitchen table whenever they were working on the tasks, which precluded them from performing the tasks while looking at the phone screen. Even though this requirement allowed us to obtain more unambiguous data, it was not exactly how people would intuitively use their phone in the process of baking and thus has potentially disrupted the normal flow of movements during baking.

The experiments were conducted by five different individuals from our project group. Since the study protocol was clearly formulated and followed by everyone, the experiment procedure looked similar. Nevertheless, there might still be some minor differences in instructions and thus influence data collection depending on the individual who conducted and evaluated the experiment. There was also not a single person that oversaw all experiments, which could have resulted in the detection of existing inconsistencies.

6 Future Work

For the recruitment and selection of participants, it is clear that a larger number of more diverse participants would be needed in order for the study to generate coherent results with more generalizability. By doing so, we could then better evaluate the influence of age, occupation, smartphone usage, as well as baking experience on participants' performance on Interfaces A & B. On a different note, the selection of participants would become more obvious and intuitive if we were to only focus on a specific user segment (i.e. elderly). Having such an objective in mind would also allow us to produce a potentially more targeted interface. Regarding experimental design, to ensure the consistency of data collection, one person should be designated to oversee the process of all experiments, or produce a sample video tutorial for other experimenters to follow. If time permits, it would also be valuable to have participants complete real baking tasks and then are asked to rate their satisfaction with the interfaces, which would improve the accuracy of the data collected for Null Hypotheses #4.

7 Conclusion

In this study, a thorough statistical analysis comparing two different mobile layout designs for a baking application has been performed. Despite the relatively small number of participants and their uneven demographic distribution, the study has proved that the difference in recipe layout could have an influence on users' navigation of the baking process, even though most of such differences were not proved statistically significant. Nevertheless, despite the small size of data, the difference in the number of screen touches still proved to be statistically significant for one of the recipes, which shed light on potential future areas of research regarding interface usability and auxiliary feature implementation.

References

[1] Dix, Alan, et al. Human-computer interaction. Pearson Education, 2003.

[2] Khurana, Urja and Melissa Tjhia. Cookin': Interactive Cooking Assistant. 2017.

[3] Hashimoto, Atsushi et al. Smart Kitchen: A User Centric Cooking Support System. 2008.

[4] "FOOBY," Coop, 11 02 2017. [Online]. Available: https://fooby.ch/en/apps.html.

[5] "Tasty," BuzzFeed, 14 08 2018. [Online]. Available: https://play.google.com/store/apps/details?id=com.buzzfeed.tasty.