S. Diefenbach, N. Henze & M. Pielot (Hrsg.): Mensch und Computer 2015 Tagungsband, Stuttgart: Oldenbourg Wissenschaftsverlag, 2015, S. 233-242.

Usability of Gesture-based Mobile Applications for First-time Use

Ben Heuwing, Ina Köller, Viktor Schanz, Thomas Mandl

Institut für Informationswissenschaft und Sprachtechnologie, Universität Hildesheim

Abstract

Gestural touch interaction is increasingly being employed when designing interaction for mobile devices. This study compares the usability of manipulative touch gestures to that of an established interaction style based on buttons and menus in a realistic setting during first-time use of two mobile productivity applications. Usability is measured as task success, time on task, error rates, and user satisfaction ratings. While results over all tasks indicate a negative influence of gestures during first-time use, they also show that users are able learn new gestures in a shorter time span than expected. However, this does not ensure users' overall satisfaction with a gestural interface.

1 Introduction

Gestural interaction on mobile devices has been introduced by many applications and touch gestures have become standardized aspects of interacting with mobile operating systems. While this interaction style finds passionate supporters as well as opponents, it is not obvious how interfaces that rely on gestures as their primary mode of interaction are understood and perceived by general users.

Previous research has mainly concentrated on the sensomotoric aspects of gestural interaction and on finding standard gestures for common actions. However, there are additional challenges to usability because gestures often do not rely on visible clues that can be interpreted, which makes it more difficult to discover the affordances of interface objects (Norman 2010). Because of this, users will often have to explicitly learn about the gestures that are available, e.g. from instructions in manuals or tutorials.

This study concentrates on usability criteria of primarily gesture-based interaction during first-time use of an application, compared to a more traditional interaction style based on (virtual) buttons and menus, which uses taps on objects visible in the interface as its primary input mode. As examples of these interaction styles, two prominent mobile productivity applications were selected which offer basic functionality to manage to-do lists.

2 Research on Touch Gestures

In general, gestures on mobile devices provide the advantage of saving space on small screens. At least in theory, gestures also provide a potentially unlimited number of possible interactions. In addition, there are claims that gestures allow for more intuitive interaction, i.e. to apply prior knowledge unconsciously, with less mental effort resulting in faster interaction (Blackler et al. 2010). Intuitive interaction therefore is influenced by the design of an application as well as by the experiences users already have with similar interactions. Gestures also have the potential to support inclusive design and accessibility of mobile devices. Gestures for the direct manipulation of an interface element appear to be easier to use than symbolic gestures, where users have to draw a free-form icon, a symbol or an alpha-numeric character (Stößl & Blessing 2010). Touch gestures have been demonstrated to close the age gap between user groups in terms of their ability to successfully complete simple gestures on touch panels compared to desktop interaction (Findlater et al. 2013). Empirical studies have also found that some types of gestures are easier to use than virtual buttons when users are in a mobile context, e.g. while walking or being distracted (Bragdon et al. 2011).

However, more complex gestures may confront users with difficulties when performing complex movements on the screen and have been criticized due to the lack of feedback while performing a gesture (Stößl & Blessing 2010). In addition, and perhaps even more critical, offering gestures as a central means of interaction will usually result in a lack of cues about the actions that can be performed (Norman 2010). Considerable efforts have already been made to create taxonomies of understandable gestures for mobile devices, e.g. in user studies (e.g. Stößl & Blessing 2010) and in online experiments (e.g. Poppinga et al. 2014), collecting the proposals of users of gestures for different actions.

However, because no standardized gestures exist for most domains and many general actions, usability of gestural interfaces will predominantly be affected by the possibility to perceive or discover the interactions which are available. For desktop-applications, Appert & Zhai (2009) demonstrated that users easily learn gestures from cues in the interface as additional short-cuts. Alpha-numeric touch gestures have been successfully employed in the context of an application for mobile search and were evaluated in a study based on touch interactions of users recorded online (Li 2010). However, this type of study does not offer the possibility to directly compare alternative interaction styles.

3 Pilot Study: Difficulty Levels of Gestures

As a pilot study, the usability of multiple types of gestures on mobile devices was compared in order to gain a basic understanding of the perception and the use of gesture-based interaction in mobile applications. This qualitative user study compared the use of manipulative touch gestures with a connection to an object on the screen with symbolic touch gestures. Twelve participants have been recruited from two age groups: six younger users aged 24 to 26 years (avg. 25) and six older users (54-70 years, avg. 62). From each group, three participants considered themselves to be experienced using smartphones. Participants were recruited using personal contacts. All of the tests were recorded for documentation and analysis. Users had to perform seven tasks using gestures with a selection of different applications on a smartphone, with instructions provided for more complicated gestures after the first unsuccessful attempt.

Results indicate that gestures which were not established more likely lead to problems. This had a larger effect on their usability than the complexity of gestures, i.e. manipulative gestures compared to more complicated iconic and symbolic gestures. This led to the question of the impact of previously unknown gestures on usability, being used with only minimal instructions in the context of a realistic scenario during first-time use.

4 Research Questions and Hypotheses

We assume that not all users want to be instructed on how to use an application, but instead will initially explore the application on their own. Therefore, the study presented here compares the influences of a gesture-based interaction style (manipulative touch gestures) to that of button and menu-oriented touch interactions on usability during first-time use. The latter are still pre-dominantly used in mobile applications. Measures for usability included task success, error-rate, time on task and ratings of user satisfaction as dependent variables. Interactions were not tested in isolation, but in the context of the same scenario formulated as goals from users' point of view. As participants in a pre-test and in the pilot study had severe difficulties when discovering gestures on their own, a basic set of instructions about the gestures was made available to the participants. The influences of age and experience on the use of gestures were considered as additional independent variables.

Based on the results of the pilot study, gestures were expected to have a negative impact on task success, error-rate and time on task during first-time use compared to a button-oriented interaction style. User satisfaction was expected to be higher for the gestural controls as a result of a more original experience. It was also assumed that age may have a negative impact on the effectiveness (measured as task success) and efficiency (measured as task time) of the use of gestures.

5 Methods

The goal of this study is to compare different interaction styles on mobile devices in realistic tasks and settings to find whether the interaction method had an influence on the user's performance as well as on the perceived usability of the application. The user tests therefore took place in a stationary setting, on site at the home or at the workplace of the respective participant. If more convenient for the participants, they could decide to take the test at a location provided by the test conductor. Using mobile devices at home has become an important use case (Church & Oliver 2011). Because of this, the test-setting of this study allows to balance the external validity of the context of use with a greater control of its influence compared to a more mobile context.

5.1 Test-Objects and Tasks

Two mobile applications with the same general functionality but different means of control have been selected to be compared for this study (Figure 1). The two applications share the purpose of creating lists of their users' liking, e.g. shopping lists or to-do lists, and are available on Android as well as iOS devices. This makes it possible to provide a familiar environment for most of the participants.

The basic structure of the two applications and the actions that are offered are very similar: To add lists, to add items to a list, to delete lists and items, and to change the order of lists and items. The tasks that the participants were asked to complete were the same for both apps (Table 1). For the actions that the tasks of the experiment consisted of, the main difference between the two applications is the use of gestures instead of buttons. In the case of the gesture-based application, *Clear*, no buttons are offered and all interactions have to be performed using gestures, while for the button-based application, *Wunderlist*, the only gestures used are established ones such as swiping for scrolling and as a means to switch between screens.

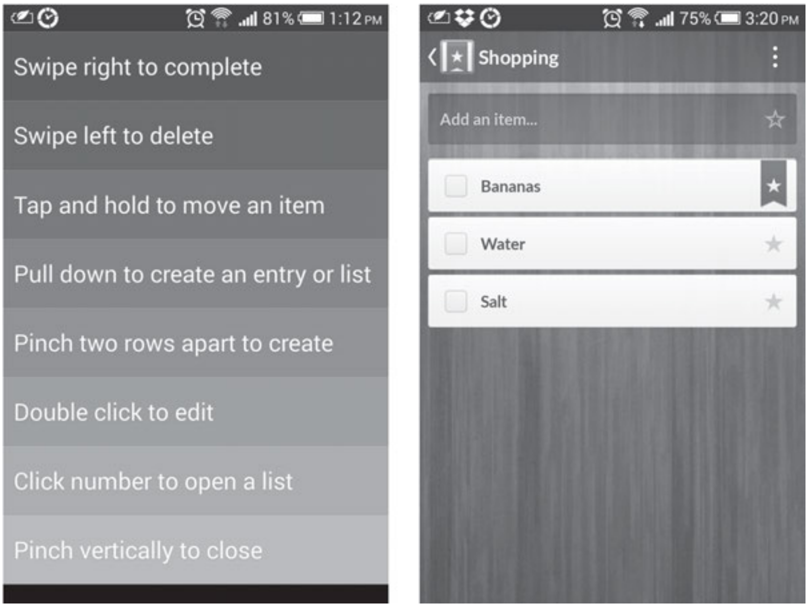

Figure 1: (left) gesture based application Clear with a list of instruction on possible gestures, (right) button- and menu-based application Wunderlist showing a list from task 3

#	Task description	Actions in Clear	Actions in Wunderlist
1	Open the app.	*App button* in phone menu	*App-button* in phone menu
2	Create a new list, named „shopping“.	*Pull down* in the list overview	*“+”-button*, confirm with *“✓”-button*
3	Put three items on the list: “bananas”, “water”, ”salt”.	Open the list by tapping on it, *pull down within the list* to create entries	Tap into text field “*Add an item…*”, confirm with *“✓”*
4	Delete the list “wishlist”.	*Pinch together* to get back to the list overview, swipe the list to the left	*Back-button*, select wishlist. *Menu-button*,*“Delete List”-button*
5	Delete “water” from the shopping list.	Tap list to open it, *swipe* the entry to the left to delete	*Back-button,* select entry, tab *check box* on its left or click the check mark in the entry description
6	Change the order of the items, so “salt” appears above “bananas”.	*Tap and hold* on of the entries, then *move it over or under* the other entry	*Star-button* on the right of the entry
7	Add “water” back on the “shopping” list.	*Pull down* to create	*“Show Completed Items”-button,* Tab in the *check box*
8	Delete the “shopping” list before you hand back your friend’s phone	*Pinch together* to get back to the list overview, *swipe* the list to the left	*Menu-button, “Delete List” -button*

Table 1: Task descriptions presented to the participants with explanations of the gestures needed to complete the task in the gesture-based application Clear.

5.2 Participants

Since one focus for this study was age, participants from different age groups had to be recruited. It proofed easy to find students to take the test using postings in appropriate online groups. Participants older than that were mainly approached through several personal contacts in the age group of around 50 years. As a result of the recruiting process, a convenience sample of 30 persons from an age range of 14 to 59 years participated in the user tests. The average age was 30.9 years. The experience of the participants using smartphones ranged from 0 to 5 years (avg. 2.5 y). None of the participants was familiar with either of the two applications tested. 17 of the test participants were female and 13 male.

5.3 Test-Setting

For the test, participants were allowed to choose their preferred platform (Android or iOS) to eliminate problems based on users’ expectations and differences regarding standard interactions between platforms, such as the use of the back-button. An iPhone with a 3.5” display (iOs) and a Sony Xperia with a 4” display (Android) and comparable screen resolutions were available as test devices. Two cameras were used to record the session, one camera directed at the face of the participant and the other at the screen of the phone.

The gesture-based application *Clear* usually comes with a tutorial and a pre-programmed list explaining how to perform possible actions. Because of the results of the pilot study and a pre-test, this list of gestures was offered to the participants as an optional support. During the sessions, most participants consulted the list. However, most of them only used these instructions later in the session when they encountered problems, which is possibly mirrors a realistic initial behavior of many users. At the beginning of a test session, the candidates filled out a questionnaire about their personal data and previous experience with smartphones. In the main testing phase, every participant used both applications (within-group design). The order of the two applications was alternated in order to avoid bias. Each set of tasks together constitutes a congruent series of interactions within a realistic usage scenario: To experiment with the respective application on the mobile phone of a friend. Immediately after completing a scenario with one of the applications, the participants answered the Software Usability Scale (SUS, Brooke 1996), a standardized usability questionnaire.[1] After both scenarios were completed, participants answered questions about which application they preferred, what they did and did not like about them, and whether they generally preferred buttons or gestures for touchscreen interaction.

6 Results

As the independent variable of this experiment, two interaction styles are compared, one predominantly based on buttons and one based on manipulative gestures on objects. In addition, the age of the participants and their experience using mobile phones were collected. Several quantitative, observational usability measures were derived from the recorded user videos: completion time, the number of errors, and the frequency of task success. The attitudes of the participants regarding the apps and the two interaction styles were collected using surveys and a short interview after the tests.

6.1 Task Success

Task success was categorized using the following scale: *completed successfully*, *completed with help* (external advice or help was given), and *failed to complete*. Overall, task success (without assistance) provides an indication regarding the effectiveness of the respective interaction styles during the first use of the applications. Out of the 240 tasks (8 tasks x 30 participants), users successfully completed 213 tasks using *Clear* and 225 using *Wunderlist*.

1 German translation of the SUS-Scale by Ruegenhagen, E. and Rummel, B. from http://www.sapdesignguild.org/resources/sus.asp - last accessed 3/30/2015

However, on the task level significant differences are only likely with tasks 2 and 4 (Figure 2, confidence intervals calculated using adjusted Wald, α=0.05).

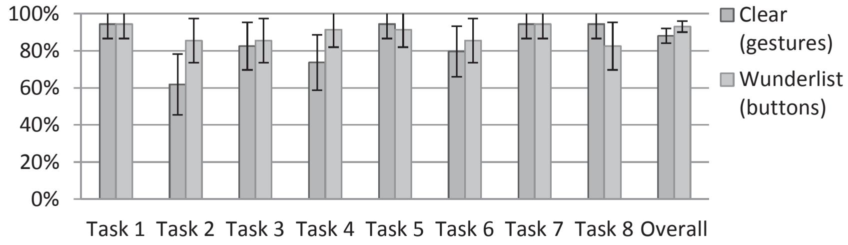

Figure 2: Rates of task success, n=30, confidence interval (α=0.05)

6.2 Errors

In addition, the errors arising in the process of completing the task were counted. An action was considered to be an error only if it was difficult to recover from, i.e. if the user unintentionally left the app, deleted an entry by accident, or inadvertently created a list or a list entry. In contrast, minor problems, which were not considered to have an influence on task success, were not counted as errors, i.e. either when a wrong screen was opened or when a wrong gesture was performed without consequences. These actions were considered to be normal activities while exploring a new application for the first time.

Error rate was higher for the gesture-oriented application *Clear* with a total of 50 errors compared to 19 for the button-based application *Wunderlist* (Table 2). However, in the course of the test scenario, the total number of errors declined for *Clear*, while a higher number of errors occurred for *Wunderlist*, with both applications accounting for approximately the same number of errors in the last five tasks.

Task	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**SUM**	**SUM 5-8**
Clear (gestures)	0	23	8	6	3	7	0	3	**50**	13
Wunderlist (buttons)	0	5	1	1	1	5	0	6	**19**	12

Table 2: Error counts

Because the difference in the number of errors may be based on learning effects during the scenario, the reasons that led to the errors were analyzed in more detail. When employing gestures, accidently deleting a list or an item was the most common problem. Examples of such accidental use of gestures mainly happened while exploring the interface during the first tasks, at a point when most of the participants had not consulted the list of instructions. The problem was aggravated by the fact that it was difficult for users to revert this action. While two participants managed to complete all tasks with *Clear* without consulting the instructions, the other 28 users referred to this list at some point during the test. However, none of the participants had problems creating lists or items during later tasks, as the gesture to accomplish this had been used and thus learned in earlier tasks. In the fourth task, the majority of users had problems to understand the pinching gesture for going back in *Clear* and needed assistance during the task. With the button-based application *Wunderlist*, the main problems resulted from the fact that the delete-button was often hidden by the keyboard of the device when it was needed during the task. This recurring problem accounts for the increase in errors for tasks 6 and 8. In addition, there was no obvious button to change the order of list entries.

6.3 Time on Task

When calculating completion time, unsuccessful tasks were not included, and neither was time spent reading instructions between tasks. The geometric mean was used as an average to compare completion times as it gives less weight to outliers (Sauro & Lewis 2010), which were observed for most tasks. Average completion time as a sum over all tasks was 262 seconds using the gesture-based application compared to 193 seconds using the button-based application, an increase of 35%. The difference is significant at a level of α = 0.05 (two-sided t-test, based on log-normalized completion times). Nevertheless, when considering the level of single tasks (Figure 3), the difference can only be considered as significant for tasks two and three. In the last task, average completion time for the gesture-based application becomes significantly shorter.

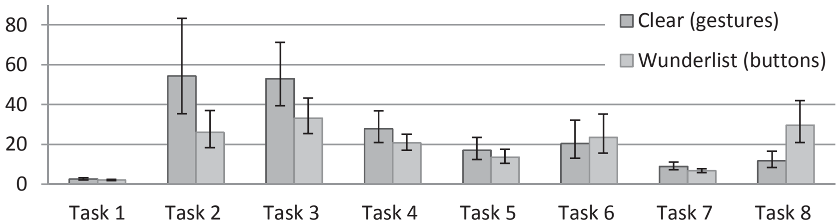

Figure 3: Time on task: Geometric mean (seconds) - confidence intervals (log-transformed) at α=0.05, n=30

6.4 Satisfaction

The results of the System Usability Scale demonstrate a trend towards the button based app. Scores in the SUS questionnaire can take values between 0 and 100. With an average score of 67.2 (stdev. 23.3), the usability of *Wunderlist* was judged higher than that of *Clear* (avg. 55.1; stdev. 25). The difference of 12.1 points is significant at the selected alpha-level of 0.05 (p=0.031, Wilcoxon Signed Ranks test). This preference is also supported by the question asking participants to compare the two apps directly: Support for the button-based app was higher (18/30 participants, 60%) than for the gesture-based app (7/30, 23%). Asked about their preference regarding buttons and gesture-based interaction in general, 16 of the 30 participants (53%) preferred buttons, compared to nine participants (30%) preferring gestures, while five participants (17%) remained undecided. Preferences for the input method were closely related to the preferences for the respective application.

6.5 Influence of age and experience

Regarding the proposed influence of age and experience, we tested their correlation (Pearson's R) with the number of errors, with time on task (log-normalized), and with user satisfaction. Age and experience using mobile devices were considered independently, as they have been shown to have different impacts in the pilot study. In the group of participants, there was no direct relationship between age and experience (r=0.12). Results show a moderate to strong relationship for age with completion time (log-normalized) for both interaction styles (gestures: r=0.61, buttons: r=0.57), and a moderate relationship between age and number of errors, but only for the button-based application (r=0.46). Correlations of experience with these measures did not reach an adequate level of significance to be reported.

7 Discussion

Results show that difficulties with gestures may have a negative impact on the usability during first-time use. Differences are significant for overall completion times and for user preferences based on SUS and direct preferences. Therefore, while usability during first time use was indeed lower for the gesture-based app during first-time use than for the button-based app, the assumption that users would prefer gestural interaction after learning the basic gestures needed for interaction was not supported. It was observed that most errors were not a product of problems at the sensomotoric level but resulted from choosing the correct gesture for an intended action. While it is possible that these differences may be the result of other characteristics of the two applications and not of their primary interaction styles, the difficulties observed during the experiment clearly indicate that this was the most important influence. Nevertheless, it is possible that the applications selected for this experiment, even though being popular, may not be representative implementations of the interaction styles tested. On a task level, differences between the usability of the interaction styles are likely to be only significant for earlier tasks. It is possible that learning gestures from prior tasks may have occurred more rapidly than expected. However, because measuring the influence of learning and experience was not the goal of this study, the claim will have to be examined in further studies on long-term use.

8 Conclusions

Even though gesture based touch interaction may provide advantages on a sensomotoric level when tested in isolation, e.g. for older users (Findlater 2012) and in mobile situations (Bragdon et al. 2011), this study demonstrates that even simple manipulative gestures may have a negative impact on usability during first-time use. The results support the claim (e.g. Norman 2010; Stößl & Blessing 2010) that gestural interfaces may lead to problems for exploring and learning. Apparently, it is difficult for users to transmit prior knowledge to this kind of interface. However, participants in this study appeared to be learning rapidly while interacting with the app. Therefore, it is crucial to improve support for users while they explore the actions available in a gestural interface. We suggest that it may be more useful to develop standardized cues which help to discover touch gestures, as proposed by Appert & Zhai (2009) for desktop applications, than to create standardized libraries of gestures. For example, the possibility to move an object with a manipulative gesture can be conveyed using spatial indicators such as closures and shadows. Most importantly, immediate feedback on the goal of a gesture is needed before the action is completed to avoid the accidental use of gestures. In addition, an interface should of course provide an easy and consistent way to undo unintended actions. These guidelines can encourage users to explore the possibilities that are available for interaction, especially for those who are not willing to go through a tutorial or who use a certain application only infrequently.

Further investigations will have to assess if the findings of this study can be replicated with a larger number of applications from different domains and with a larger number of tasks. In addition, strategies employed by users when exploring an unknown gestural interface could be inspected in more detail. This would provide the foundation to develop support for the learning behavior of different user groups and to recognize when alternative means of interaction are necessary in order to enable inclusive access to applications that are promoting innovative styles of interaction.

References

Appert, C. ; Zhai, S. (2009): Using Strokes As Command Shortcuts: Cognitive Benefits and Toolkit Support. In: *Proceedings SIGCHI 2009.* ACM, pp. 2289–2298

Blackler, A., Popovic, V. &, Mahar, D. (2010). Investigating users' intuitive interaction with complex artefacts. In: *Applied ergonomics* vol. 41, Nr. 1, pp. 72–92

Bragdon, A., Nelson, E., Li, Y. & Hinckley, K. (2011). Experimental analysis of touch-screen gesture designs in mobile environments. In: *Proceedings SIGCHI 2011*. ACM, pp. 403–412

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, Nr. 189

Church, K. & Oliver, N. (2011). Understanding mobile web and mobile search use in today's dynamic mobile landscape. In: *Proceedings MobileHCI 2011*. ACM, pp. 67–76

Findlater, L., Froehlich, J. E., Fattal, K., Wobbrock, J. O. & Dastyar, T. (2013). Age-related differences in performance with touchscreens compared to traditional mouse input. In: *Proceedings SIGCHI 2013*. ACM, pp. 343–346

Li, Y. (2010): Gesture Search: A Tool for Fast Mobile Data Access. In: *Proceedings of the Symposium on User Interface Software and Technology 2010*. ACM, pp. 87–96

Norman, D. A. (2010). Natural User Interfaces are not Natural. In: *interactions* vol. 17, Nr. 3

Poppinga, B. ; Sahami Shirazi, A. ; Henze, N. ; Heuten, W. ; Boll, S. (2014): Understanding Shortcut Gestures on Mobile Touch Devices. In: *Proceedings MobileHCI 2014*. ACM, S. 173–182

Sauro, J. & Lewis, J. R. (2010). Average task times in usability tests: what to report? In: *Proceedings SIGCHI 2010*. ACM, pp. 2347–2350

Stößel, C. & Blessing, L. (2010). Mobile Device Interaction Gestures for Older Users. In: *Proceedings NordiCHI 2010.* ACM, pp. 793–796