"I Didn't Know, You Could Do That" - Affordance Signifiers for Touch Gestures on Mobile Devices

Emilie Lind Damkjær, Liv Arleth, and Hendrik Knoche(✉)

Aalborg University, Rendsburggade 14, 9000 Aalborg, Denmark
{edamkj14,larlet14}@student.aau.dk, hk@create.aau.dk

Abstract. As smartphones have become widely available at low prices, interface designers need to improve accessibility of mobile applications for e.g. illiterate users. We evaluated how effectively and efficiently different signifiers communicated the affordances of dragging and double tapping touch gestures in a text editor environment. We found that spatially represented signifiers for dragging (drag handles and drop shadow) communicated the affordance better than temporal signifiers for double tapping. The latter appeared only for a limited amount of time in the user interface and were less effective and efficient than even a control condition that provided no additional information.

Keywords: Touch interactions · Gesture signifiers · Affordance

1 Introduction

Smartphones with limited touch screen space for user interface controls have become widely adopted by both textually literate and illiterate users. Overloading controls with several affordances (possible actions) allows for leaner and more compact interfaces as several actions can be accessed within the same space through different gestures or sequence of microinteractions. The paper investigates ways to signify the afforded touch gestures of dragging and double tapping through signifiers that have either a lasting spatial representation or only appear for a limited amount of time in the user interface (UI). The gained understanding is relevant for both regular and interfaces for people with special needs.

2 Background

Applications on mobile devices need to be designed for limited screen space. Overloading interface elements and controls with several affordances (possible actions) allows for making more interactions available. Designers employ signifiers to help users at discovering or identifying these affordances. We use Norman's definitions of affordances and signifiers [6] throughout this paper. While

A. L. Brooks et al. (Eds.): ArtsIT 2018/DLI 2018, LNICST 265, pp. 206–212, 2019.
https://doi.org/10.1007/978-3-030-06134-0_23

single-, double-click, right-click, and drag are common in the UIs of desktop operating systems such overloading is less common in the UIs of mobile apps. For example, on a touch screen a button usually gets triggered by a single tap but could narrate its associated function (feedforward) on a long-press [3]. While the question of whether and how to signify affordances of touch gestures on controls is relevant to all touch screen surfaces, we deem the importance of overloading controls particularly valuable in assistive interfaces for users with special needs. We turn our attention to illiterate users and focus on how to use overloading to make text editors on mobile phones accessible to them. While illiterate people can master interactions on smart phones through e.g. rote learning [4] composing text messages represents a desirable skill and challenge [1]. Audio feedback can be used to help illiterate users understand what word(s) a text object in the UI represents [1]. By editor we denote an environment in which users can enter and manipulate some form of data, represented e.g. as text or images.

In a mobile banking application for illiterate users [5] money was represented by pictures of actual bills rather than numbers. Users could drag the desired number of bills into an editor that showed the amount of bills picked so far. Nothing signified the affordance of dragging the bills. Users had to discover this affordance through other means such as introductions or feedback. The editor mainly used swiping and dragging as gestures.

The mobile texting editor for illiterate users by Friscira et al. allowed for listening to complete messages or individual words by tapping on them and provided different options for entering words that did not require typing [1]. The interface overloaded these controls (the word buttons) with two affordances - single and double tap. Displaying words inside a white box on a blue message bubble background, making them visually resemble flat design buttons signified the single tap affordance. The affordance of double tapping words in older messages to add them to the editor had no signifier.

3 Study on Affordance Signifiers

We set up an experiment to investigate how to best signify touch gesture affordances to the user without using text or introductory instructions.

Inspired by Friscira et al.'s text editor [1], we created an SMS (short message service) editor prototype (see Fig. 1) containing a bank of words (rectangles with barcodes in the light blue area) and a grey editor area in which the words were supposed to be assembled to form a message. Single-tapping word elements played out the word as audio. We created a task with a limited set of words to choose which the user had to add to the editor to create the message. Each of the light grey boxes contained one word and the barcode labels on each meant to simulate that participants could not read the words but had to rely on the audio to form the correct message. The editor had no cursor or means for rearranging the words. All words had to be added in the right order. We implemented six different versions of this interface to test a total of four signifiers and two affordances. Two control versions without explicit signifiers served as a baseline.

They were visually identical to Figs. 1A and B, afforded populating the word editor by means of double tapping on or dragging words from the word bank.

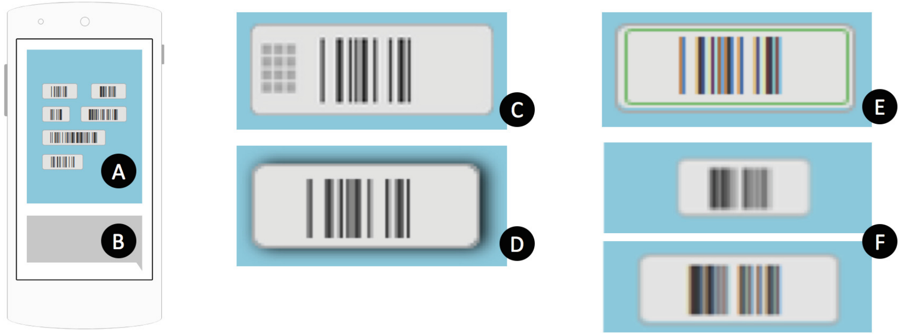

Fig. 1. The study prototype (left) with word bank (A) and editor (B), and magnified the signifiers for dragging (middle) - drag handle (C), and drop shadow (D), and for double tap - border (E) and pulse animation (F).

One signifier for dragging used a common design pattern known as a drag-handle - a cluster of dots - on the word element intended to invoke a textured, "grabbable" surface (see Fig. 1C). The second signifier employed a drop shadow on the word elements (see Fig. 1D) to make them appear to be floating above the blue background.

One signifier for double tapping employed feedback using two borders around the words. On the first tap, the innermost border briefly turned green (see Fig. 1E). On the second tap the outer border became green, too. The second signifier had the words pulsate from small (see Fig. 1F top) to large (bottom) twice, immediately after opening the page mimicking the word buttons being pressed down twice.

3.1 Procedure

The study evaluated the effectiveness and efficiency of the four signifiers - two for each of dragging and double tapping through a balanced between-subjects usability test that measured the participants' completion rates in composing a message. We relied on a proxy population of students, who, however, could not read the symbols, which encoded the words in the editor. While this might raise concerns for the validity of the study, the object of the study was not related to the textual content as such but the understanding of the signifiers of the word objects. Signifiers of affordances are rarely communicated through text in UIs. This approach ensured that the single tap - the primary gesture on mobile devices - provided the audio feedback and the participants had to find other gestures to populate the editor with word objects. We drew on interaction analysis to further the design of UIs to benefit textually illiterate users. In a similar vein, see e.g.

Huenerfauth's approach drawing on task analysis and structured design [2]. The discussion section returns to this topic.

Thirty-six students participated in the test as volunteers, none of whom followed a design oriented education. The test followed a mixed design in which each participant experienced two affordance signifier versions of a smartphone text editor prototype - one for each afforded gesture (dragging and double tapping). The order of gestures was counterbalanced between participants. This resulted in a total of 72 trials from 12 trials for each of the six versions.

After initial explanations, the participant had to compose the phrase "I am very happy" in each version of the prototype. After succeeding or giving up each task, they reported their thought process and reflections about their actions while solving the task. A video camera captured a close-up of the phone screen for subsequent video analysis. We scored task completion for each trial. A subsequent video interaction analysis counted the number of gestures and types (e.g. tap, drag, triple-tap etc.) the participants tried before finding the correct gesture, for cases in which they were successful.

3.2 Results

First, we verified with a Fisher's Exact test that the attempt number (first/second try) did not affect the participants' success rates (p(two-tailed) $= 0.4735$). In terms of success rates, participants were better at discovering dragging (see Fig. 2) than double tapping (see Fig. 3). Each figure shows the different signifier conditions. The drag handle was the most effective at signifying its affordance. The least effective signifier was the pulse animation for double tapping. Notably, both double tap signifiers resulted in lower success rates than their no signifier control version.

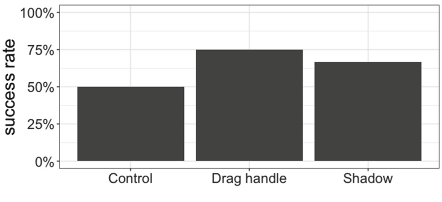

Fig. 2. Success rates for dragging

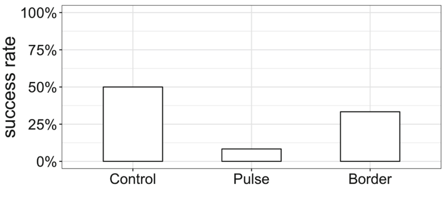

Fig. 3. Success rates for double tapping

Four two-tailed Fisher's Exact tests showed that, compared to their corresponding control, none of the signifiers - *drag handle* ($p = 0.40$), *shadow* ($p = 0.68$), *border* ($p = 0.68$), and *pulse animation* ($p = 0.07$) made the participant perform significantly different. However, one-tailed Fisher's Exact tests of the double tap signifiers showed that the lower success rates of the *pulse animation* was significantly worse (p(one-tailed) $= 0.034$) than its control version. The pulse animation made finding the right gesture more difficult than when not receiving extra information other than the shape of the word objects.

The video interaction analysis counted the number of gestures the participants tried before finding the correct gesture. We further classified these gestures by type (e.g. tap, drag, and triple-tap). Figures 4 and 5 depict the number of gestures before finding the correct one by signifier for those participants who managed to finish the task successfully. The statistical analysis on the signifiers' efficiency omitted the *pulse* animation, as only one participant found this affordance. Two-tailed T-tests comparing the number of gestures and the number of different types of gestures tried before identifying the correct one against the control conditions found no significant differences ($p > 0.05$). However, the one-tailed T-test comparing *border* to its control was close to significance ($t(4) = 2.1$, $p = 0.051$). On average it took participants twice as many gestures with the border signifier to successfully complete the task than the participants in the control condition. In summary, none of the drag signifiers were more or less efficient than its control in helping the participants to find the correct afforded gesture. But the *border* signifier was less efficient at signalling the double tap affordance than the plain word button objects.

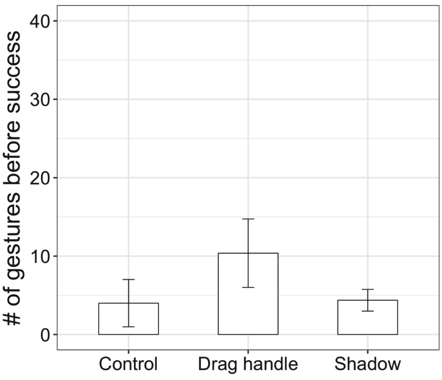

Fig. 4. Dragging signifier efficiencies with 0.95 confidence interval error bars

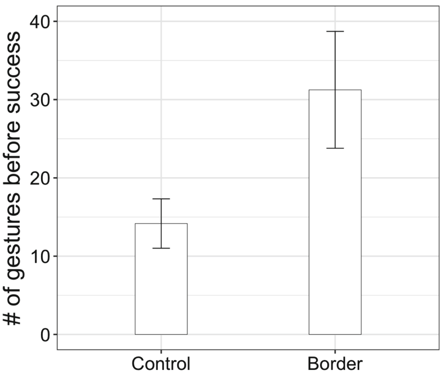

Fig. 5. Double tapping signifier efficiencies with 0.95 confidence interval error bars

4 Discussion

The signifier with the highest success rate was the drag handle, however, it was not significantly better than the control and required additional space in the UI for each word object more than the similarly performing drop shadow version. A few participants misunderstood our drag handle as a symbol for a number pad or a menu. The pulse animation signifier for double tapping made users find the affordance less often. The most likely interpretation seems that the participants perceived this signifier as a gimmick or artefact of rendering the page and not associated with the individual controls. The green *border* signifier also seemed to lead to confusion as participants required more interactions before successfully completing the task. One possible explanation for the poor signifying performance of *border* and *pulse* animation could be the low temporal availability of the cues. For *border* it was only briefly available after the first tap and could have been partially occluded by the interacting finger.

We used textually literate students, none of whom were experts in user experience design or touch interactions, but who had ample experience with smartphones and other UIs. However, we did not control for their familiarity with double tap or drag gestures on mobile touch device screens. Arguably, dragging gestures are more common in current mobile UIs (Android and iOS) than double tapping but the success rates of their respective control conditions were identical. Technical literacy in terms of exposure to touch gestures might be important as a variable to control for in future studies.

Out of the 36 test participants 24 tried identifying the words based on their barcode length. Illiterate people in non-logographic languages might be to some degree gauge the lengths of the word objects from the duration of a spoken a word to help in such task. While this does not call into question the internal validity of the results on the signifiers it might be helpful for the design of future studies in this area, especially when involving proxy populations.

5 Conclusion

We evaluated signifiers for dragging and double tapping affordances for movable word objects in a touch text editor. The temporally available signifiers for double tapping were either less effective (pulse animation on initial page render) or less efficient (green border after first tap) than a control condition that provided participants with less information. Such signifiers with limited temporal availability (e.g. on page opening) may lead to confusion. Drag handles yielded slightly better performance but required substantial space in the UI.

References

1. Friscira, E., Knoche, H., Huang, J.: Getting in touch with text: designing a mobile phone application for illiterate users to harness SMS. In: Proceedings of the 2nd ACM Symposium on Computing for Development, p. 5. ACM (2012)
2. Huenerfauth, M.: Developing design recommendations for computer interfaces accessible to illiterate users. Master's thesis, University College Dublin (2002)
3. Knoche, H., Rao, P.S., Jamadagni, H.S., Huang, J.: Actions and advice in coli: a mobile social network to support agricultural peer learning. In: Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, pp. 1191–1198. ACM (2015)
4. Knoche, H., Huang, J.: Text is not the enemy - how illiterates use their mobile phones. In: NUIs for New Worlds: New Interaction Forms and Interfaces for Mobile Applications in Developing Countries - CHI 2012 Workshop, Austin (2012)

5. Mesfin, W., Grønli, T.M., Ghinea, G., Younas, M.: A mobile money solution for illiterate users. In: 2015 IEEE International Conference on Mobile Services (MS), pp. 328–335. IEEE (2015)
6. Norman, D.: The Design of Everyday Things: Revised and Expanded Edition. Constellation (2013)