

"But Wait, There's More!" A Deeper Look into Gestures on Touch Interfaces

Arleth, Liv; Damkjær, Emilie Lind; Knoche, Hendrik

Published in:

ArtsIT 2019, DLI 2019: Interactivity, Game Creation, Design, Learning, and Innovation

DOI (link to publication from Publisher):

[10.1007/978-3-030-53294-9_20](https://doi.org/10.1007/978-3-030-53294-9_20)

Publication date:

2020

Document Version

Accepted author manuscript, peer reviewed version

[Link to publication from Aalborg University](#)

Citation for published version (APA):

Arleth, L., Damkjær, E. L., & Knoche, H. (2020). "But Wait, There's More!" A Deeper Look into Gestures on Touch Interfaces. In *ArtsIT 2019, DLI 2019: Interactivity, Game Creation, Design, Learning, and Innovation* (pp. 290-308). Springer. https://doi.org/10.1007/978-3-030-53294-9_20

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

”But Wait, There’s More!” A Deeper Look into Temporally Placing Touch Gesture Signifiers

Liv Arleth, Emilie Lind Damkjær, and Hendrik Knoche^[0000–0003–3950–8453]

Aalborg University, Rendsburggade 14, 9000 Aalborg, Denmark

Abstract. The language used in interaction design is affected by the wide array of academic backgrounds of interaction designers. Therefore one word may have several meanings, which can be confusing when conducting research in this field. In this paper, we define three levels of interaction: macro-, micro- and nanointeractions, the latter of which is the focus of this study. We use Buxton’s three state model to break down common gestures on touch interfaces into nanointeractions, thereby identifying where in the process of a gesture its signifiers can appear. This is useful when overloading controls in small interfaces. We conducted an experiment to determine whether the temporal placement of a signifier before, during, or after a gesture made any difference for the discoverability of a double and long tap affordance. No clear tendencies were found regarding the temporal placement of the signifier. However, the concept of nanointeractions can be a valuable tool for interaction design.

Keywords: HCI · Interaction · Usability design · Microinteractions · Touch interactions · Gestures · Nanointeractions

1 Introduction

Verplank [15] stressed the importance of three questions that interaction designers must answer: How does one *do*? How does one *feel*? How does one *know*? The user has some form of knowledge (*know*) from previous applications with which they have interacted, and perhaps a mental map of how they imagine the current application to work. When presented with some form of control (such as a button), the interface may provide the user with feedforward (*feel*), revealing some information as to what will happen if certain gestures are performed on the control. The user will process this information based on their previous knowledge and expectations, and perform some action (*do*) on the control. Based on which action is performed, the control may provide some form of feedback (*feel*) such as a sound signifying success, or the sensation of a button being pressed, which enables the user to update their knowledge and provides support on how to proceed.

Damkjær et al. [3] discussed the concept of microinteractions and their importance to not only the overall process of interacting with an application, but also how the user chooses each gesture based on what they know and feel. However, even this approach was simplifying things, as each gesture consisted of a series

of smaller actions, and the user may process information or change their course mid-gesture. This paper focuses on the nature of these types of nanointeractions.

2 Terminology

The terminology in the field of interaction design and user interfaces is not always consistent due to interaction designers, UX designers, etc. coming from many different scientific backgrounds, each with their own language. Therefore, we describe our terminology thoroughly in the hopes it will help streamline the language in the field of interaction and UX design.

Interactions The word interaction is used to describe many different things. Sometimes it refers to the overarching task like using an application to take a photo. Other times it refers to the action of tapping the shutter button in a camera application. And sometimes it refers to the action of tapping on the screen of a smartphone. To make it clear to which we refer, we distinguish between different levels of *interaction*. *Macrointeractions* refer to the overarching tasks, the process itself, e.g. taking a photo. These usually benefit from the user having a good mental map of the system in question. A *microinteraction* indicates the small interaction with a contextual purpose limited to a single gesture, e.g. clicking the shutter button to take a photo [9]. However, even gestures as simple as a touch screen button press are comprised of several smaller actions: approaching the button with your finger, touching the button, and letting go. These are the types of (inter-)actions we refer to as *nanointeractions*.

Signifiers In line with Norman’s [11] definition, in this paper, a *signifier* refers to a design aspect communicating to the user the existence of a specific action possibility (affordance) on some control in the user interface. The modalities to communicate can for example be visual, haptic, or auditory. In smartphone application design, signifiers commonly take the shape of text or interface control elements (such as buttons, sliders, etc.) that are based on design conventions and guidelines. Often affordances lack explicit signifiers.

Affordance, Feedforward, and Feedback The three terms *affordance*, *feedforward* and *feedback* are commonly used in the field of interaction design, but do not always denote the same concept. We drew inspiration from Vermuelen et al.’s synthesis and insights of these three terms [14]. But for the scope of this paper, we rely on Norman’s definition of *affordance*. An *affordance* is a relationship between an actor and the properties of an object, e.g. a button can be pressed. Affordances are not always clearly signified and can in fact be *hidden*. Therefore we distinguish between *perceivable* and *hidden* affordances. Perceivable affordances are supported by and can be understood from signifiers on an object while hidden affordances lack visible or otherwise perceivable signifiers. For example, the affordance of a long tap on a button is often hidden in touch

interfaces. It should be noted that Norman has previously used the term perceived affordance to mean what he now calls a signifier [10], and therefore several papers use these terms interchangeably.

Feedforward communicates to the user something about what would happen (the function that gets triggered) if they performed the afforded action. For example, pressing a green button or a button with a text label saying "ok" on it having the effect of confirming something. *Feedback* refers to information becoming available during or after performing the action allowed by the affordance. Staying with the previous example, pressing the above "ok" button might yield feedback in the form of a toast message ("Thank you!") confirming that the button press has been registered.

Gestures A touch *gesture*, e.g. a tap, is the physical action performed on a touch interface. This should not be confused with a microinteraction. A microinteraction involves a gesture but must also include a purpose such as scrolling by using a fling. In this paper, we look at the gestures tap, double tap, long tap, and drag.

Another way to describe gestures besides using words, is utilising models. In this paper we use Buxton's three state model [2] to break down gestures. As seen in Figure 1(b), the model consist of three states: state 0 is the state where you are out of range of the gesture; state 1 represents when you are in range; and state 2 is when the intended gesture is carried out.

To show how all these terms play together in terms of Verplank's interaction model, we have modified his sketch according to the terms established in this section, shown in Figure 1(a).

3 Breaking Down Gestures

When Verplank speaks of interactions, he uses the example of flipping a switch and seeing the light come on [15]. However, as we have previously stated, even a

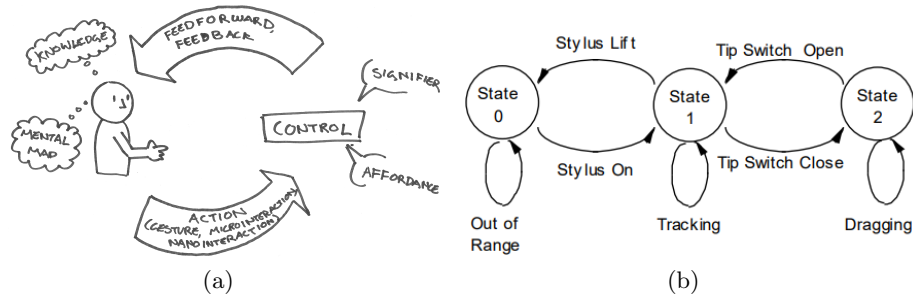


Fig. 1. (a) The cycle of a user interacting with a control, modified from Verplank's sketch [15] and (b) Buxton's three state model [2].

gesture as simple as flipping a switch may be broken down into a series of nanointeractions. The user may discover new information during a nanointeraction, before they have completed the intended gesture.

As an example, placing one’s finger on the switch is a nanointeraction in and of itself. At this state the user may feel that it is impossible to flip the switch one way, but possible to flip it the other. The nanointeraction of applying pressure while flipping the switch gives the user some information on how much resistance the switch gives, and thus how much pressure they must apply. They may even discover a sequential affordance [4] of fading the light gradually that was previously hidden. Breaking gestures into nanointeractions in this manner rather than one instantaneous action is useful when designing user experiences for a touch device. To show how gestures are broken into nanointeractions, we provide some examples visualised through diagrams inspired by Larsen’s work [7], using Buxton’s three state model [2].

The simplest touch interaction, the *tap*, is illustrated in Figure 2(a). In our model state 0 represents when the user’s finger is out of range of the control on the touch screen, state 1 when the user’s finger is in range of the control, and state 2 when the user interacts with the control. For the tap gesture this is when the finger touches the control. The arrow marked in red represents the moment that the system delivers feedback regarding the affordance of the control. In the case of a tap this does not happen until the action is completed, i.e. when the user lifts their finger.

Figure 2(b) illustrates the *drag* gesture. Here, users may transition to state 2a by performing the nanointeraction of moving their finger once placed on the screen. Moving the finger on the screen is actually several nanointeractions, but for the sake of simplicity it is illustrated as only one in our model. The

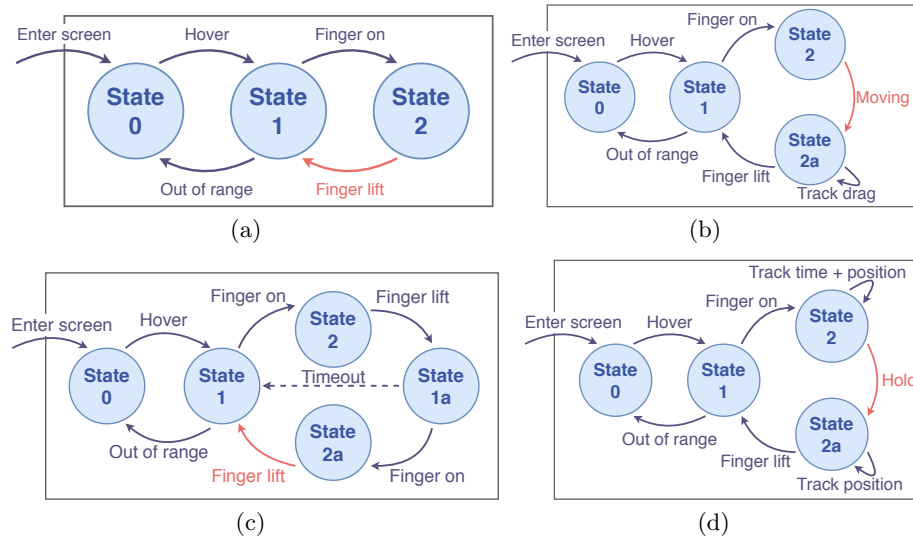


Fig. 2. Gestures broken down into nanointeraction state diagrams: (a) *tap*, (b) *drag*, (c) *double tap*, and (d) *long tap*.

gesture is completed when the user lifts their finger, thus returning to state 1. The affordance of dragging is revealed by feedback the instant movement is detected, thereby appearing during the gesture. This differs from a tap in that rather than after the gesture is complete, the feedback is provided mid-gesture, allowing users to discover the affordance before completing the gesture. However, this takes place during a very short amount of time, so it is not always possible to react to the feedback.

The *long tap* gesture is depicted in Figure 2(d). To get to state 2a the user must hold their finger's position for a certain amount of time, and when this threshold is reached the hidden affordance is revealed and the gesture is completed when the finger is lifted.

Often, a *long tap* is combined with a *drag* when implementing them on touch screen. This is similar to selecting an object by clicking and holding when using a computer mouse. Figure 3 illustrates how adding the drag gesture to the long tap model introduces an additional state, state 2b. Compared to a normal long tap the signifier revealing the affordance is delayed to correspond to the signifier revealing the drag affordance instead, i.e. giving feedback when the user's finger moves after the long tap.

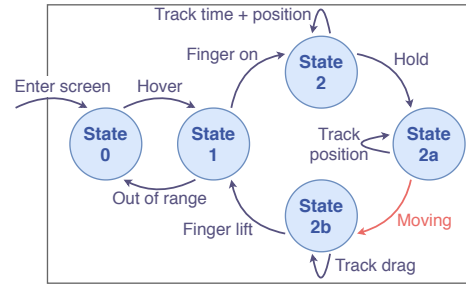


Fig. 3. The *long tap* gesture continued with the *drag* gesture, broken into nanointeractions.

The gesture *double tap* illustrated in Figure 2(c) does not have a linear state sequence like the three previously described gestures. As the name implies it is the act of tapping twice. Like the long tap, this gesture has a temporal aspect to the nanointeractions, i.e. the time between *finger lift* and *finger on* determines whether you perform a double tap or just two separate taps. This is illustrated in Figure 2(c) where state 1a is present. This represents the temporal nature of the gesture by working as a timer state, where if you timeout you go back to state 1 and have to start the gesture over again. Unless specifically designed for it, the affordance is not revealed until the user's finger is lifted after the second tap, making it very hard to discover.

To illustrate how to break down a system with numerous gesture affordances, Figure 4 depicts both the *tap*, *drag*, *long tap*, *long tap+drag* and *double tap* affordances, broken into nanointeractions. All five gestures require the user to first approach the control, and then place their finger on it, thus entering state 2 on touch interfaces. Moving one's finger while in this state reveals the affordance of dragging, which initiates state 2a. The user may complete the dragging gesture by lifting their finger off the control. If, instead, the user rests their finger on the control while in state 2, they enter state 2b, initiating a long tap. From here, they may either complete this gesture by lifting their finger, or initiate a long tap+drag by moving their finger. If the user lifts their finger while in state 2, they have performed a regular tap. However, as both the single- and double tap

are afforded by this system, the single tap is not complete until a certain timer runs out, ensuring that the user did not perform a double tap.

This timeout will usually be quite short, so the user does not sense a delay upon tapping a control. If, instead, they place their finger back on the control and lift it again, a double tap is performed. In Figure 4, the red arrows signify places in the process where feedback is typically provided.

Illustrating interactions in this manner and thinking of gestures as a system of nanointeractions rather than something that happens instantaneously may prove useful when considering how to design an application with meaningful signifiers, both in terms of feedforward and feedback.

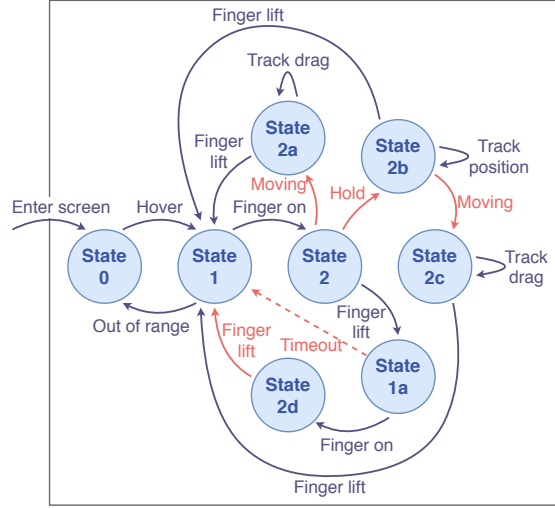


Fig. 4. The combination of the *tap*, *drag*, *long tap*, *long tap+drag* and *double tap* gestures, broken into nanointeractions.

4 Related Research

One way to signify certain affordances is by using metaphors to tap into users' existing knowledge. Oakley et al. utilised this idea for their smartwatch prototype as a way to introduce new affordances [12]. A finger placed vertically across the center of the watch was implemented as a way to activate the mute function on a media player application, as the action resembles placing a finger across lips. Two fingers across the watch toggle between pause and play, as the two fingers resemble the traditional 'pause' icon (two vertical parallel lines). Placing the finger horizontally along the bottom of the screen emulates the shape subtitles take on a screen, and thus enables the subtitles. This was discoverable only through feedback—no signifiers were provided to the user. Oakley et al. relied on explaining these affordances to test users, and did not report on whether previous knowledge of the used metaphors was sufficient to discover the affordances in this case. It should be noted that they never quantitatively tested these affordances, but rather relied on participant self-reports in how far they understood the gestures. Users supported the notion of metaphors and found the pause and mute functions "intuitive", but there was no formal testing of error rates etc. Furthermore, some of the implemented gestures lacked a metaphor or signifier and were thus hidden affordances.

An alternative way to signify affordances was by guiding the user through nudges, rather than designing the object itself with signifiers. This concept was explored by Lopes et al. [8]. Their system, Affordance++, stimulated the user’s arms as they approached the object in question to nudge them towards using the proper gesture. They found that this was a useful way to communicate especially dynamic interactions to the user, e.g. shaking a spray can before spraying. However, it is limited to real-world contexts in which users are willing to wear an arm-mounted device at all times. Interestingly, this nudging signifier was not provided by the control itself (the lamp), but rather by an external device. Furthermore, the nudges continuously provided signification to the user based on which state they are currently in, directly nudging them towards the next state. This is shown in Figure 5, where we have applied our model to the interaction. Note that overloading would create a problem for this solution, as the possibility of moving to several different states from the current one means that there is little point in being nudged towards just one.

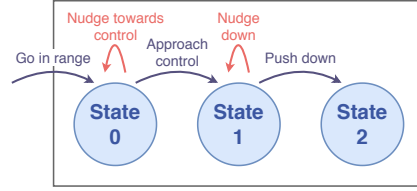


Fig. 5. Turning on the lamp, guided by Affordance++.

Another way to provide users with feedforward without visual signifiers, is to rely on audio, as is the case with e.g. answering machines and automated phone call systems (see Figure 6). Here the user is provided with audio signifiers of all the affordances and feedforward in the form of a list of available options. This system provides both feedforward and feedback. However, sometimes it provides too much of it, or gives the feedforward in a problematic order. It is also very time-consuming, especially if the user does not know what they are searching for and needs to listen through all the options more than once. As illustrated by Figure 6 the user’s options are limited by lack of knowledge of said options, until a certain amount of time has passed. Overloading in this situation is not a possible solution, since it is limited to one modality, audio.

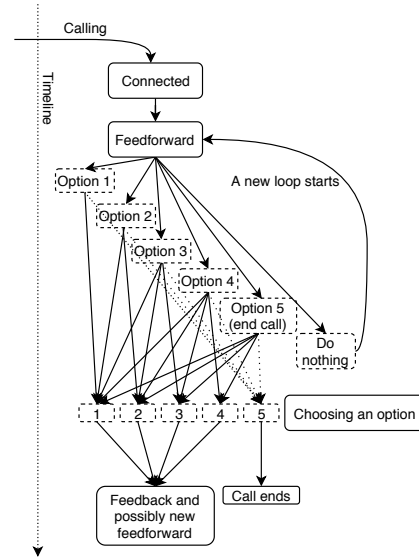


Fig. 6. A chronological illustration of a hypothetical answering machine.

Harrison et al. acknowledge the need for overloading in touch interfaces, and achieved it by differentiating inputs from different parts of the finger - the tip, the pad, the nail, and the knuckle [5]. But they did not performed any tests on how to signify these affordances. Pedersen

and Hornbæk [13] proposed touch force as a modality for overloading. Their users could accurately control two different levels of pressure, although this took some getting used to. Moreover, users expressed fatigue after having touched the screen with increased pressure for a while, indicating that this modality was inferior as an interaction and should be used to a limited degree. Aslan et al. proposed the *gazeover* as a way to implement something similar to the mouseover on a mouse-and-keyboard setup, but potentially available for touch interfaces [1]. However, they did not test user’s ability to perceive this affordance on a touch interface. Damkjær et al. [3] tested different visual signifiers (a shadow on a button, a drag handle among others) to see which one(s) best conveyed the affordances of dragging and double tapping. Signifiers with a temporal element performed worse than signifiers with no temporal element.

A lot of creative solutions have been proposed using different modalities to communicate affordances or implement overloading. However, many lack signifiers for the affordances, resulting in hidden affordances impairing usability. We intend to explore this research gap by investigating how we can explicitly signify and turn hidden affordances of touch screen gestures such as double tap and long tap, into perceived affordances. This knowledge can then inform the design of overloading controls. The following study investigated when, during the nanointeractions of a gesture, signifiers should appear.

5 Experiment

In this experiment we tested whether the temporal placement of signifiers for double tap and long tap affected the discoverability of these affordances. Based on the identified research gap we set up two hypotheses, with the dependent variable being the discoverability of the relevant gesture, and the independent variable being the signifier and its temporal placement:

1. A signifier improves the discoverability of the affordance of a gesture.
2. Early placement of the signifier improves discoverability of the affordance.

5.1 The Prototype

To test the hypotheses, we created a prototype app in Android Studio. Four variations were created for each affordance:

- Ctrl - A control version with no visual signifier.
- Enter - The signifier appeared on entering state 0, and then repeats in a five second loop (the green spot in Figure 7).
- Middle - A signifier appeared when the user touched the screen, thus entering state 2 (the purple spot in Figure 7).
- After - A signifier appeared after a completed single tap (the yellow spot in Figure 7).

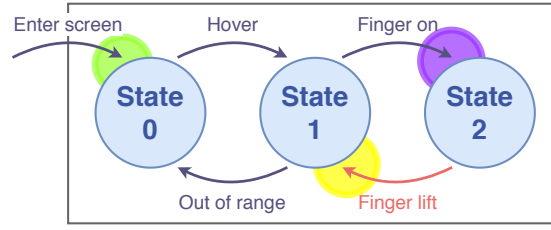


Fig. 7. Signifiers can appear on *entering* the current UI screen (green), on touching the screen (*middle*, purple), and *after* a single tap (yellow).

To counter potential learning effects we added a distractor version. In this version, instead of a double tap or long tap affordance, there was a fling affordance with no signifier indicating this.

Upon launching the app, the screen shows a menu of these nine versions. This menu was not to be seen by test participants, but was a tool for the test facilitators to control which version was applied. This menu can be seen in Figure 8(a).

Once a version had been chosen, the application redirected to the screen seen in Figure 8(b). As can be seen, it was a simple to-do list application, which allowed users to add items to a list of chores, mark the ones they have completed, and delete items from the list. Users may write the name of a chore in the text field (1), then press the button (2) to then add that item to the list (3). If the text field was empty upon button press, no item was added to the list. As can be seen in Figure 8(b), each list item contains a box which may be checked upon tapping it. This box was also checked if the list item was single tapped at all, either on the text or in the empty space to the right.

For each version of this app, there was a hidden affordance to delete items added to the list. For four of these versions, the trigger was to double tap, for four others it was to long tap, and for the last one (the distractor) the trigger was to fling. The gesture in question must be performed on the item the user wished to delete, but it did not matter whether the gesture was performed on the text, the box, or the empty space.

As described previously, a signifier revealing the given affordance may appear at various times depending on the chosen version. For the double tap versions (except for the control version), the signifier was a pulse of two rings which

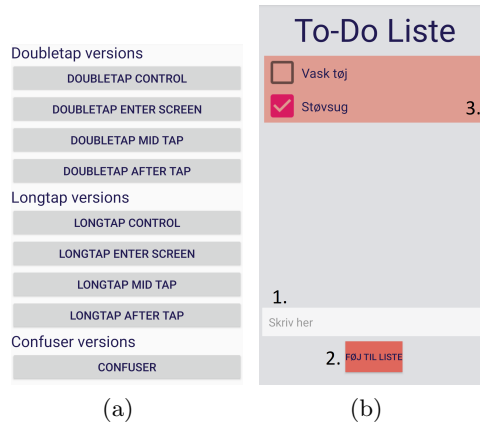


Fig. 8. (a) The version menu on the prototype application. (b) The main screen of the to-do list app.

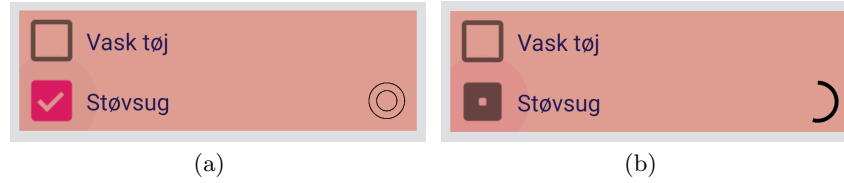


Fig. 9. The affordance signifiers for the (a) double tap and (b) long tap

expanded one after the other, then disappeared. This was to emulate the idea of a double tap. A screenshot of this can be seen Figure 9(a).

For the long tap versions (except for the control version), the signifier was a wheel which gradually filled out over time, indicating that the user may hold their finger on the screen for an extended period of time (see Figure 9(b)).

Throughout the application, every touch gesture performed, as well as every successfully added, marked, or deleted item, was logged for analysis purposes.

5.2 Experiment Design

The evaluation consisted of two experiments, one for the double tap and the other for the long tap gesture. The hypotheses apply to both experiments.

The independent variable was the temporal placement of the signifier. For each gesture, the four conditions were tested, plus the distractor to slightly counter the learning curve of participants.

The dependent variable was the discoverability of the affordance of the relevant gesture measured by the success rate, the time until successful deletion, and the number of different gestures tried before finding the correct one. This was collected by logging this information within the prototype. Another measure we used was a series of questions inspired by the NASA Task Load Index (TLX) [6]. We tested the raw TLX method in a pilot trial, but since the test participant found the original scale confusing, we changed the scale to a range from zero to ten, with zero being the least and ten the greatest amount possible. The questions were as follows:

- How mentally demanding did you find the task?
- How much did you feel you had to rush when performing the task?
- How much success did you have in accomplishing the task?
- How much effort did you have to put in when accomplishing the task?
- How irritated, stressed, annoyed, or frustrated did you feel during the task?

Due to the experiment becoming too long to be able to recruit people off of the street each participant only tried five conditions: two of each gesture and the distractor as the middle trial, making the experiment a between subjects design. To alleviate the learning curve of the participants somewhat, the order of the conditions was determined by using a Latin square design.

5.3 Participants

The only requirement for the participants was that they not have a background in interface design. For this test 64 random people in the age group of 14 to 77 years old were recruited off of the streets in Aalborg and amongst employees at Regionshuset Nordjylland. The experiments were conducted at three different places due to recruitment issues: a space at Aalborg University, an office at Regionshuset Nordjylland and the Main Public Library in Aalborg. Out of the 64 participants, we had 40 female and 24 male participants, 39 used iOS and 25 used Android on their smartphones, and 52 were right-handed.

5.4 Procedure

The apparatus used for this experiment consisted of a Sony Xperia XZ2 Compact smartphone, a laptop for running our prototype and saving the log, a laptop for notetaking and questionnaire answers, and a video camera on a tripod to film the participant's hands interacting with the smartphone.

The procedure for the experiment was as follows: first the participant signed a consent form and was explained the procedure by one of the two facilitators. They then filled out a demographics questionnaire. The video camera was turned on when the participant received the first version of the prototype. They were told to first add an item to the list, then mark an item as completed and finally to delete an item. A trial was considered a success if an item was deleted. When the participant either succeeded in deleting or gave up, they answered some follow up questions regarding their actions with that version followed by the TLX-inspired scales. This process repeated with the next four versions, with the third trial always being the distractor condition. After the participant finished with the fifth trial, the log was copied from Android Studio to a text file. The entire procedure took approximately 15 minutes from start to finish.

5.5 Results

When the data was not normally distributed we used a Kruskal-Wallis test instead of one-way ANOVAs when comparing the dependent variables of the temporal signifier placements each in the double tap and long tap versions.

We did not compare results between the two gestures, as users should attempt the long tap more frequently than the double tap, because the former is more common in touch screens in general. Figures 10(a) and 10(b) depict the distribution of successes and failures of each temporal placement of signifiers excluding the distractor. A quick comparison of these figures confirms that participants were much more successful at discovering the long tap than the double tap affordance. However, CHI-square tests found no significant differences of the success rates between the temporal placements of the signifiers neither for long¹ nor for double tap². We found a trend for the first gesture people used in each

¹ $\chi^2(3, N = 32) = 2.56, p = 0.46$

² $\chi^2(3, N = 32) = 4.69, p = 0.2$

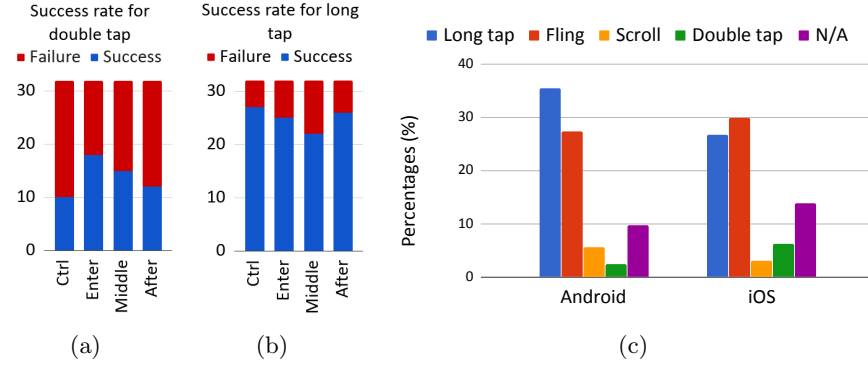


Fig. 10. The distribution of successes and failures of the double tap (a) and long tap (b) versions. (c) The percentage of first gestures performed in each trial other than single taps. N/A represents trials with no gesture attempts but single taps.

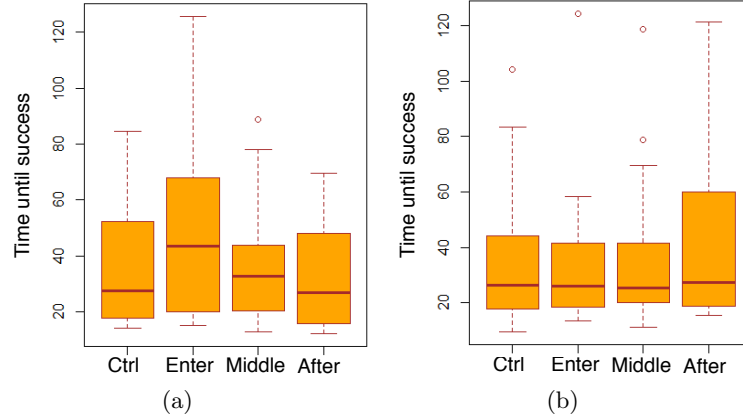


Fig. 11. Box plots of the time until successful deletion for (a) double tap and (b) long tap.

trial other than single taps, which depended on the mobile operating systems the participants were used to. Android users tried long taps first and iOS users were more likely to try a fling first (see Figure 10(c)). Much less frequently people tried scrolls (slowly swiping) and double taps. In some cases the participants did not try anything but single taps (denoted as N/A in the figure).

Kruskal-Wallis tests of the completion times found no significant differences between the temporal placements for the double ($H(3)=2.6$, $p=0.45$) and long tap versions ($H(3)=2.65$, $p=0.86$). See Figure 11 for an overview of the data.

Similarly when comparing the number of gestures performed before success, Kruskal-Wallis tests found no significant differences for the temporal placement of signifiers for the double ($H(3)=2.62$, $p=0.45$), and long tap gesture ($H(3)=3.79$, $p=0.28$). The distribution of the number of gestures are illustrated in Figure 12. We tested whether removing single taps from the data changed these results but Kruskal-Wallis tests found no significant differences for the

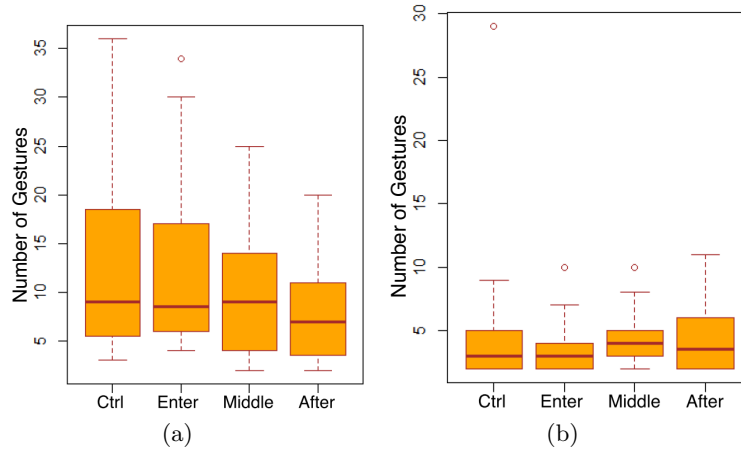


Fig. 12. Box plots of the number of gestures performed before success for (a) double tap and (b) long tap.

double ($H(3)=3.67$, $p=0.3$) and long tap version ($H(3)=2.61$, $p=0.46$) in that subset of the data either. We then looked at whether the temporal placement influenced the variety of gestures the participants tried before success and excluded the single taps from this analysis. Again Kruskal-Wallis tests found no significant differences for neither double ($H(3)=2.63$, $p=0.45$) nor long tap ($H(3)=2.82$, $p=0.42$) gestures.

The questionnaire data analysis focused on the mental demand, temporal demand, performance, effort and frustration experienced during the trials. We compared between the four different signifier placements of long tap and double tap respectively. Figure 13 visualizes the overall distribution of answers. No answers were discarded in this analysis since the questionnaire results' ability to be statistically analysed were not affected by whether a participant succeeded or not. One participant chose to not answer the questions for two of the trials, which were the distractor trial and the Ctrl for double tap. This reduced the sample size of Ctrl for double tap to 31. The data gathered from the questionnaire was based on a ranking scale, and thus ordinal and non-parametric, and analysed through Kruskal-Wallis tests. Comparing the answers from the questionnaire yielded no statistically significant results (see Table 1 for all the statistical reporting details). A consistent trend in the data across both gestures was that the temporal placement *after* the single tap required the most effort and yielded the highest frustration and mental demand.

6 Discussion

The results showed no significant differences between control condition and 1) the addition of signifiers and 2) the different temporal placements of signifiers. This means that we can neither disprove that the addition of a signifier does not

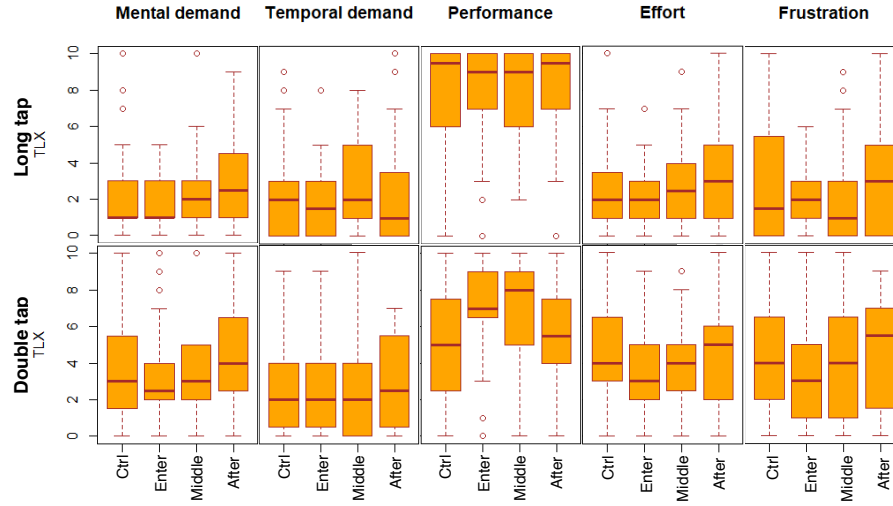


Fig. 13. Box plots of TLX sub-scale for both the long tap results and the double tap results

	Double tap	Long tap
Mental demand	$H(3)=3.36, p=0.34$	$H(3)=3.45, p=0.33$
Temporal demand	$H(3)=1.35, p=0.72$	$H(3)=1.63, p=0.65$
Performance	$H(3)=6.41, p=0.09$	$H(3)=0.54, p=0.91$
Effort	$H(3)=3, p=0.39$	$H(3)=4.66, p=0.2$
Frustration	$H(3)=2.4, p=0.49$	$H(3)=1.34, p=0.72$

Table 1. The H test statistics and p -values from the Kruskal-Wallis tests comparing the signifier placements by gesture

matter for the discoverability of the affordance of the relevant gesture, nor that the temporal placement of the signifier does not matter for the discoverability of the affordance of the relevant gesture. However, we can still gather relevant and helpful information from the data. Based on the logging data, long tap was the most common gesture when no signifiers were present to indicate other affordances were possible. During the test it became made clear that a fling was a common gesture for deleting an item in the type of application we made, especially for iOS users.

Our results showed that in this experiment the design of the signifiers for long and double tap were not better at communicating the affordances than the condition without a signifier. This was particularly problematic for the less common double tap affordance. However, the double tap signifier used in this study (the two expanding rings to the right of the control, see Figure 9(a)) communicated this affordance much better (56% success in the 'Enter' placement) than pulsing the control itself with comparable temporal placement in our pre-

vious study [3]. Only 8% of people found the double tap affordance through that approach. This large difference can be partially explained by the increased opportunity to perceive the affordance in this study through looping the animation every five seconds in comparison to showing it only once at entry. But there were further differences in task design, the controls that were signified, and the visual design of the signifier itself, which might have affected discoverability.

Regarding the validity of our results a few things should be considered. The order in which participants tried the different versions was balanced using a Latin square design having all combinations an equal amount of times during the experiment. Our participants came from a wide variety of backgrounds and age groups. We even had a few different nationalities. These two factors strengthen the validity of our results. Our primary data was gathered from the log of the application. When looking through the logs the program did not always assign the correct gesture name to what the application clearly read, e.g. in a double tap version a double tap not resulting in a delete action, therefore having been read as two single taps, or in a long tap version a scroll being read as a long tap and success being reached that way. This was all corrected by having a person read through the log and note the results, but it does hurt the validity of our log data. Regarding the data from the questionnaire, self-reporting is always hard to validate especially when the experiment is a between-groups setup. Having the questions based on a standard in the field (NASA TLX) increases the validity of our results as does the fact that the results corresponded quite well with the log data.

There was a number of limitations to our findings. The reliability of our results was potentially affected by changing location three times with one of these locations being in a public space, albeit a somewhat isolated corner of this space. The wide age range and background of our participants should have given us a reliable sample of the population, strengthening the external validity of our results. During the first few trials the way we phrased the task of deleting an item was misleadingly vague and making some test participants believe that checking a box on the screen was sufficient. When we rephrased the assignment to specify that the item had to disappear altogether, participants understood the task better. During the test, we discovered that when the user executed a fling slowly, it sometimes registered as a long tap, regardless of finger movement. This means that while the log data may show that some users discovered the long tap affordance, they may have actually have attempted a fling. This problem of the temporal unfolding of gestures leading to misinterpretation represents a common problem with overloading controls with different gestures that interaction designers will need to address in the future.

Several times throughout the experiment, in their search for a button to delete an item, test participants would accidentally return to the secret menu screen or to the phone's home screen. This may have impacted the user in two ways: The confusion may have caused the participants to feel more insecure and less inclined to try different approaches to the task. On the other hand, if a

participant caught a peek of the text on the buttons, it may have revealed to them an affordance that they did not previously perceive.

Further analysis could explore whether the order in which the different versions were experienced had an effect on the discoverability. It is possible that, if a gesture (e.g. a long tap) was not possible in the first version a participant experienced, the participant might never attempt that gesture again in later trials when the gesture *was* possible. On the other hand, if the first version had the affordance of the most obvious non-tap gesture to that participant (often long tap), the participant might be more willing to try other gestures, as they had already seen that deleting an item was possible. Studying the relationship between the position on the screen of a visual signifier and the position at which the user performs a gesture represents another avenue for future work. On touch screens, the user's finger may block the visual signifier from their view if it appears upon touch, which may cause the user to never see the signifier, thus hindering discoverability. This is less of a problem for signifiers, which are always visible. But they occupy space in the interface.

The concept of nanointeractions opens up several alleys for future research. With more time and resources, we would have performed a large-scale within-subjects experiment with a more easily understandable signifier in order to better determine the viability of revealing signifiers to the user while they are at a certain nanointeraction stage in a gesture. Furthermore, it may be valuable to explore the nature of changing gestures and how designers can take advantage of this.

In this paper, we focused our research on two gestures: long tap and double tap. This was to keep the scope manageable. There are many other touch gestures which can be broken down into nanointeractions, and the complexity of some of them make them especially interesting. An example is the drag gesture requiring the user to change the position of their finger. While we have mapped a drag gesture into nanointeractions, the user may in theory change their course many times *while dragging*, which could be considered nanointeractions in and of themselves. Every touch gesture is different and future research should focus on how these can be combined best when overloading controls with several gestures. Furthermore, if one thinks of touch gestures as a series of nanointeractions, one may also explore the nature of changing gestures. For example, if the user has initiated a long tap by placing their finger on a control and holding it there, but then moves the finger away from the control before lifting, they have changed their course "in the middle of" a gesture, which designers and future researchers may take into account, as it allows for new combinations of gestures for which to design affordances and signifiers. This paper focused entirely on visual signifiers, but other types of feedback may affect users differently. We believe future research investigating the relationship between audio or haptic feedback and nanointeraction stages could be particularly fruitful.

7 Conclusion

In this paper, we focused on the discoverability of gesture affordances depending on whether a signifier was made visible before any gesture was attempted, during a gesture, or after a gesture has been completed. While the experiment showed no significant differences between the temporal placement of signifiers, we argue that this study still has great value for future research. The possibility of attempting to let the user perceive a previously hidden affordance as they are currently "in the middle of" a gesture has not previously been explored, and we hope that future researchers will further investigate. We analyzed current research into different affordance design angles with different modalities with this terminology. Our main contribution to the field is the concept of nanointeractions. Most research thinks of touch screen gestures as one single interaction, without breaking it down. However, if we as interaction designers instead consider the elements which make up a gesture — the nanointeractions — it will reveal opportunities for novel gesture designs and overloading controls by combining these in interfaces with limited space.

References

1. Aslan, I., Dietz, M., André, E.: Gazeover – Exploring the UX of Gaze-triggered Affordance Communication for GUI Elements. In: Proceedings of the 20th ACM International Conference on Multimodal Interaction. pp. 253–257. ICMI '18, ACM, New York, NY, USA (2018). <https://doi.org/10.1145/3242969.3242987>
2. Buxton, W.: A three-state model of graphical input. In: Human-computer interaction-INTERACT. vol. 90, pp. 449–456 (1990)
3. Damkjær, E.L., Arleth, L., Knoche, H.: "I Didn't Know, You Could Do That"-Affordance Signifiers for Touch Gestures on Mobile Devices. In: Interactivity, Game Creation, Design, Learning, and Innovation. pp. 206–212. ArtsIT 2018, DLI 2018, Springer (2018). https://doi.org/10.1007/978-3-030-06134-0_23
4. Gaver, W.W.: Technology affordances. In: Proc. of CHI'19. pp. 79–84. ACM (1991)
5. Harrison, C., Schwarz, J., Hudson, S.E.: TapSense: Enhancing Finger Interaction on Touch Surfaces. In: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology. pp. 627–636. UIST '11, ACM, New York, NY, USA (2011). <https://doi.org/10.1145/2047196.2047279>
6. Hart, S.G.: Nasa-Task Load Index (NASA-TLX); 20 Years Later. Proceedings of the Human Factors and Ergonomics Society Annual Meeting **50**(9), 904–908 (2006). <https://doi.org/10.1177/154193120605000909>
7. Larsen, J.V., Knoche, H.: States and Sound: Modelling User Interactions with Musical Interfaces. In: New Interfaces for Musical Expression 2017 New Interfaces for Musical Expression. NIME (2017)
8. Lopes, P., Jonell, P., Baudisch, P.: Affordance++: Allowing Objects to Communicate Dynamic Use. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. pp. 2515–2524. CHI '15, ACM, New York, NY, USA (2015). <https://doi.org/10.1145/2702123.2702128>
9. McDaniel, R.: Understanding microinteractions as applied research opportunities for information designers. Communication Design Quarterly Review **3**(2), 55–62 (2015)

10. Norman, D.A.: The way I see IT signifiers, not affordances. *interactions* **15**(6), 18–19 (2008)
11. Norman, D.A.: *Design of Everyday Things: Revised and Expanded*. Basic Books, New York (2013)
12. Oakley, I., Lindahl, C., Le, K., Lee, DoYoung, I.M.R.: The Flat Finger: Exploring Area Touches on Smartwatches. *CHI 2016* (2016)
13. Pedersen, E.W., Hornbæk, K.: Expressive Touch: Studying Tapping Force on Tabletops. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. pp. 421–430. *CHI '14*, ACM, New York, NY, USA (2014). <https://doi.org/10.1145/2556288.2557019>
14. Vermeulen, J., Luyten, K., van den Hoven, E., Coninx, K.: Crossing the bridge over Norman’s Gulf of Execution: revealing feedforward’s true identity. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. pp. 1931–1940. ACM (2013)
15. Verplank, B.: *Interaction Design Sketchbook* (2009), <http://billverplank.com/CiiD>