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

Emilie Lind Damkjær
Aalborg University
Aalborg, Denmark
edamkj14@student.aau.dk

Liv Arleth
Aalborg University
Aalborg, Denmark
larlet14@student.aau.dk

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 then 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 signifiers for said gesture can appear. This is useful when overloading controls on an interface. We conducted an experiment to determine whether the temporal placement of a signifier made any difference for the discoverability of an affordance, in this case double tap and long tap. To test this, we developed an application which we exposed to 64 test participants, and logged their every gesture. A questionnaire was also administered after each trial concerning the participants' own experience with the app. Kruskal-Wallis tests were performed on the data with no significant results. No clear tendencies were found regarding whether the temporal placement of the signifier affected the discoverability of the chosen affordances. However, we believe the concept of nanointeractions to be a valuable contribution to the field of interaction design in and of itself.

Author Keywords
HCI; interaction; usability design; microinteractions; touch interactions; signifiers; feedforward; affordance; gestures; nanointeractions;

INTRODUCTION
Verplank [17] says that we as interaction designers must answer three questions: 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 (such as a sound signifying success, or the sensation of a button being pressed), which will in turn inform the user how to proceed.

In our previous paper, we have discussed the concept of microinteractions, and how it is significant to consider 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 [4]. However, even this approach is simplifying things, as each gesture is a process of various small actions, and the user may process information or change their course mid-gesture. This paper focuses on the nature of these types of interactions.

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 separate the word *interaction* into several specialised definitions. *Macrointeractions* are 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* is the small interaction with a contextual purpose limited to a single gesture, e.g. clicking the shutter button to take a photo [10].

However, even gestures as simple as a touch screen button press is comprised of several smaller actions: approaching the button with your finger, touching the button, and letting go. These are the type of action we will be calling *nanointeractions*.

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

Signifiers

In this paper, a *signifier* refers to a specific design aspect telling the user something about the affordances of the application. These can for example be visual, haptic, or auditory. In our previous work [4], different visual signifiers (a shadow on a button, a drag handle among others) were tested to see which one(s) best conveyed the affordance. In smartphone application design, signifiers are often in the form of text or based on design guidelines.

Affordance, Feedforward, and Feedback

The three terms *affordance*, *feedforward* and *feedback* are commonly used in the field of interaction design, but does not always indicate the same thing. We partially rely on Vermuelen et al.'s synthesis and definitions of these three terms [16]. However, even in their synthesis, there is some ambiguity as to the definition of *affordance*, and therefore we define *affordance* as the possible actions in relation to an object, e.g. a button can be pressed. Affordance is not always clear and can in fact be *hidden*. Therefore we further make a distinction between *perceived affordance* and *hidden affordance*, which are subcategories of affordance. The perceived affordance is what can easily be gathered from signifiers on an object, and hidden affordance is when an object has affordances that are not visible to the user. A long tap on a button is often a hidden affordance. It should be noted that Norman has previously used the term perceived affordance to mean what he now calls a signifier [12], and therefore several papers use these terms interchangeably. Our use of the term should not be confused with the term signifier, as they are entirely different concepts.

Feedforward tells the user something about what would happen if they perform the action allowed by the affordance, e.g. pressing a green button or a button with a text label saying "ok" on it is a confirmation action.

Feedback is what happens during or after you perform the action allowed by the affordance, e.g. the confirm action is performed and a toast appears.

Gestures

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

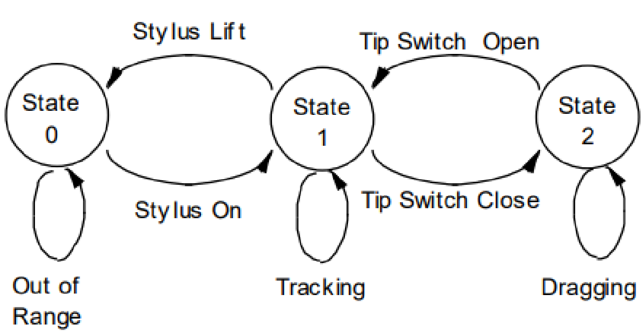


Figure 1: Buxton's three state model [3].

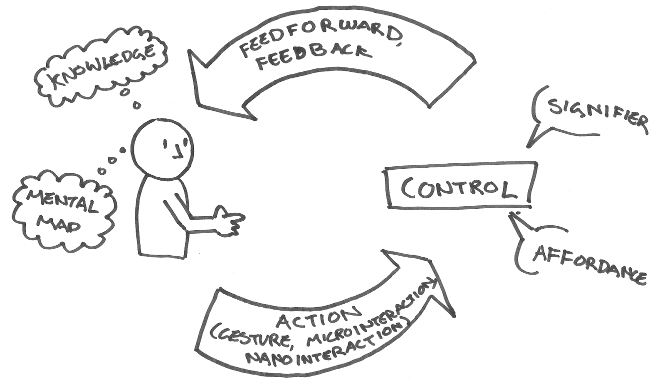


Figure 2: The cycle of a user interacting with a control, modified from Verplank's sketch [17].

Another way to describe gestures besides using words, is utilising models. In this paper we use Buxton's three state model [3] to break down gestures. As seen in Figure 1, 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 2.

BREAKING DOWN GESTURES

When Verplank speaks of interactions, he uses the example of flipping a switch and seeing the light come on [17]. However, as we have previously stated, even a 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 previously hidden affordance of fading the light gradually. Breaking gestures into nanointeractions in this manner rather than one instantaneous action is useful when designing for a touch device.

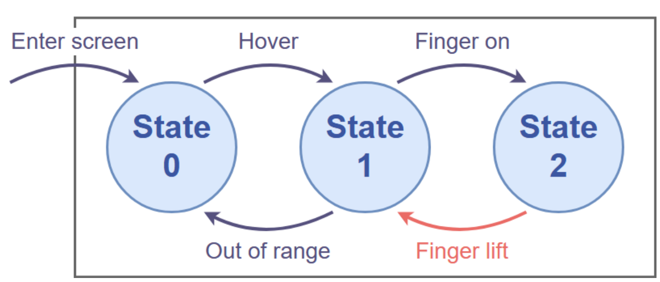


Figure 3: The *tap* gesture, broken into nanointeractions.

To show how gestures are broken into nanointeractions, we provide some examples visualised through diagrams inspired by Larsen [8], using Buxton's three state model [3].

The simplest touch interaction, the *tap*, is illustrated in Figure 3. 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 doesn't happen until the action is completed, i.e. when the user lifts their finger.

Figure 5 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 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 6. 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.

In many cases 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. As can be seen in Figure 4 it adds the drag gesture to the long tap model by introducing 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.

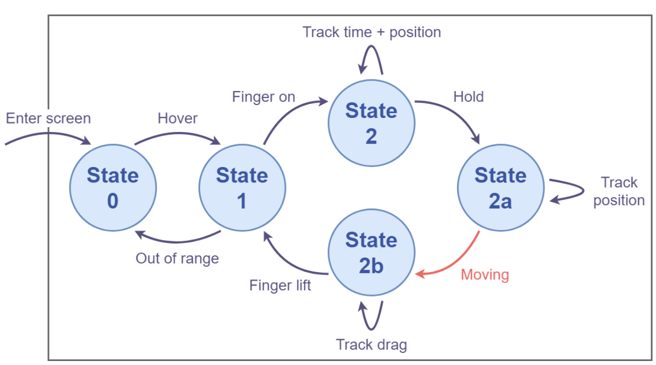


Figure 4: The *long tap* gesture continued with the *drag* gesture, broken into nanointeractions.

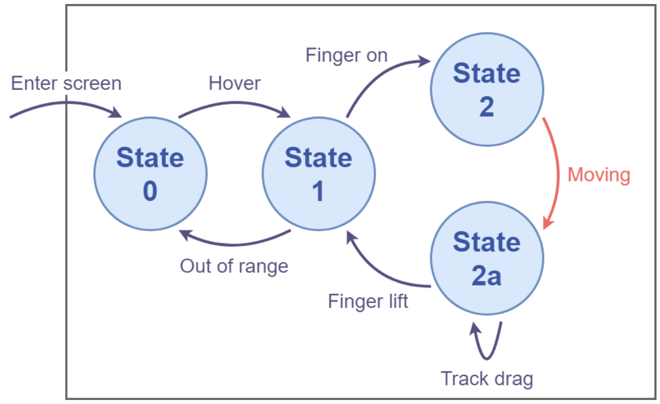


Figure 5: The *drag* gesture, broken into nanointeractions.

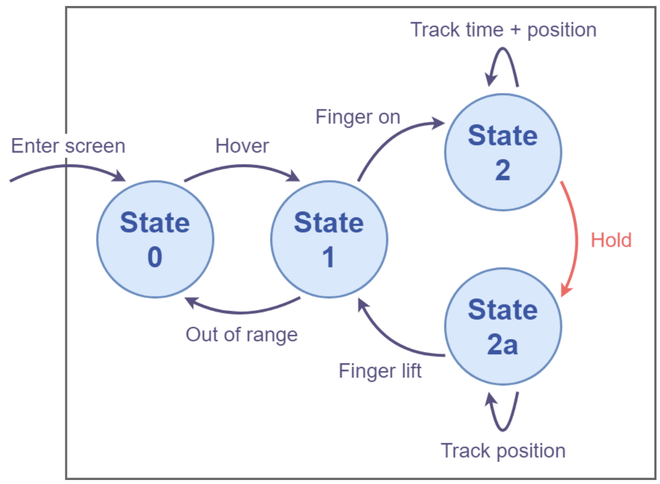


Figure 6: The *long tap* gesture, broken into nanointeractions.

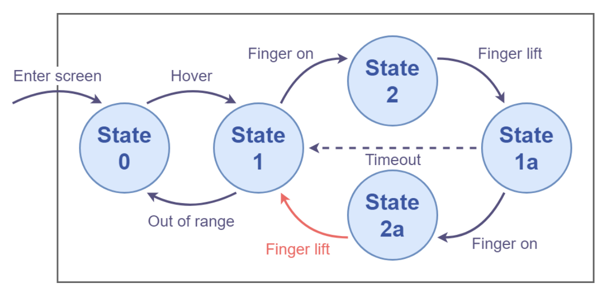


Figure 7: The *double tap* gesture, broken into nanointeractions.

The gesture *double tap* illustrated in Figure 7 doesn't 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 7 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.

Figures 3-7 illustrate the nanointeractions of hypothetical controls with an affordance of only one gesture each. However, it is useful to overload controls so that the user may perform several gestures on one control. To illustrate how to break down a system with numerous gesture affordances, Figure 8 depicts both the *tap*, *drag*, *long tap*, *long tap+drag* and *double tap* affordances, broken into nanointeractions.

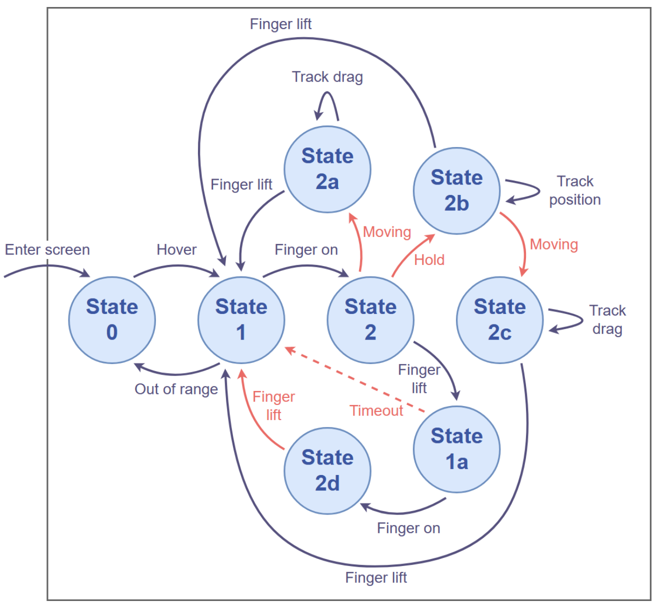


Figure 8: The combination of the *tap*, *drag*, *long tap*, *long tap+drag* and *double tap* gestures, broken into nanointeractions.

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 longtap+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 8, 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.

RELATED RESEARCH

In this section we discuss design guidelines for touch interfaces, examine current research in exploratory affordance design and re-examine our previous research based on our model for gestures broken down into nanointeractions.

Design Guidelines

To ease the user's interaction with a touch screen, the designer must consider the perceived affordances based on the feedforward given to the user. It may be beneficial to follow design guidelines, as users will have certain expectations as to how an interface will look depending on which devices they are accustomed to.

Many systems, such as Apple, provide developer guidelines to designers in order to keep signifiers consistent and easily comprehensible to users. Their guidelines include adding a border or background to buttons when necessary to make them appear tappable [1]. They also recommend displaying a row of dots [1] at the bottom of the screen to show the user their position in a flat list of pages, as seen in Figure 9.

Figure 9: Dots at the bottom of the screen to indicate the existence of other pages, as well as the user's position within the list of pages [1].

If the user is to pick one option out of a list, the designer may implement a *picker*, seen in Figure 10, as recommended by the guidelines. Note that Apple's standard picker design includes signifiers for the affordance of dragging; other options than the currently active one are dimly visible and increasingly warped the further they are from the middle. This emulates a haptic cylindrical object the user may roll with their finger.

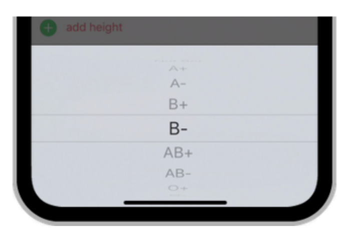


Figure 10: A picker, allowing the user to select a single item in a list of many [1].

Another form of control is the *slider* [1] as seen in Figure 11, which allows the user to slide a value between its minimum and maximum value (such as screen brightness). The finite track and a thumb with a drop shadow signifies that the user may drag the thumb to a position on the track.

Figure 11: A slider [1].

It should be noted that for some recommended controls, Apple provides design guidelines, but no explanation as to what will signify to the user which gesture is needed. One example is the edit menu [1]. The guidelines state that by default, the user may long tap or double tap an element in a text field in order to bring out the edit menu with options to cut, copy, etc. However, they provide no explanation as to how the user would know to either long tap or double tap. Another example is refreshing a page. The affordance to drag the screen downwards is often available [1], with no signifiers to let the user know.

According to Google's design guidelines for touch interfaces, signifiers on the interface should indicate which gestures are afforded [5]. Many examples of gestures are provided, but the guidelines fail to provide examples of signifiers to let users perceive the affordances before a gesture is executed.

Note that while Android and iOS users may have different habits stemming from the design patterns they are used to [15], neither Apple nor Google provide sufficient guidelines as to how to design signifiers for non-standard gestures (for touch screens, any gesture other than tapping). In our previous research, we found this to be a challenge as well [4].

Affordance Design

One way to signify certain affordances is by using metaphors to tap into users' existing knowledge. Oakley et al. utilises this idea for their smartwatch prototype as a way to introduce new affordances [13]. Examples of this can be seen in Figure 12.

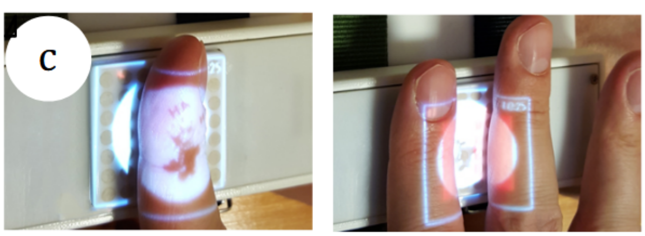


Figure 12: How the mute and pause/play functions are toggled, respectively [13].

A finger placed vertically across the center of the watch was implemented as a way to activating 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 which is two vertical 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 is signified only through feedback — no feedforward is provided to the user, thus Oakley et al. relied on explaining these affordances to test users, and we do not know whether previous knowledge is sufficient to discover the affordances in this case.

It should be noted that they never quantitatively test these affordances, but rather rely on participant reports for their evaluation of the understandability. 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 lack a metaphor or signifier and were thus completely hidden affordances, such as the ability to copy an existing appointment in a calendar application by picking up a finger from flat against the screen to just a fingertip.

An alternative way to signify affordances is to guide the user through nudges, rather than design the object itself with feedforward. This concept was explored by Lopes et al. [9]. Their product, Affordance++, can stimulate the user's arms as they approach the object in question to nudge them towards the proper gesture. An example of this can be seen in Figure 13.

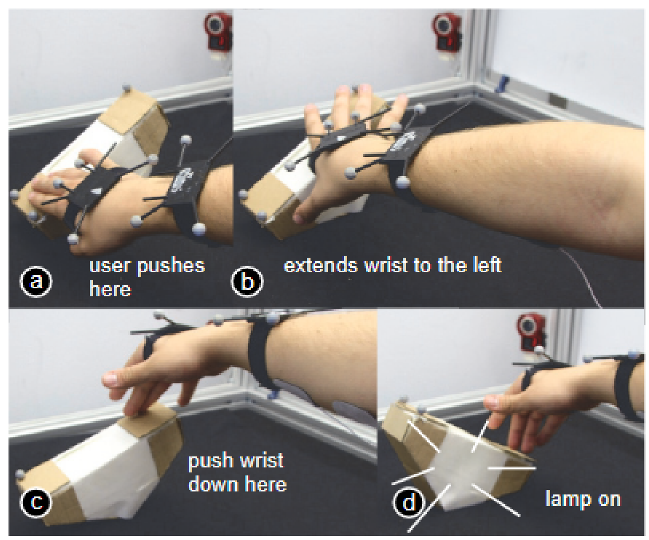


Figure 13: Affordance++ helps the user interact with an unfamiliar lamp [9].

They found that this is a useful way to communicate especially dynamic interactions to the user, e.g. shaking a spray can before spraying. However, it is unrealistic for real-world application, as it is unlikely users will want to wear an arm-mounted device at all times.

What is interesting about the Affordance++ is that the feedforward is given not by the control itself (the lamp), but rather by an external device. Furthermore, the nudges continuously provide feedforward to the user based on which state they are currently in, directly nudging them towards the next state. This is shown in Figure 14, where we have illustrated the interaction using the previously described model.

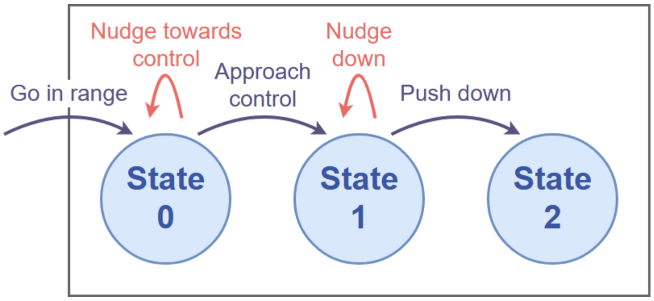


Figure 14: Turning on the lamp, guided by Affordance++.

It should be noted 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.

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. Here the user is provided with all the affordances by getting feedforward in the form of what is usually a list of options available.

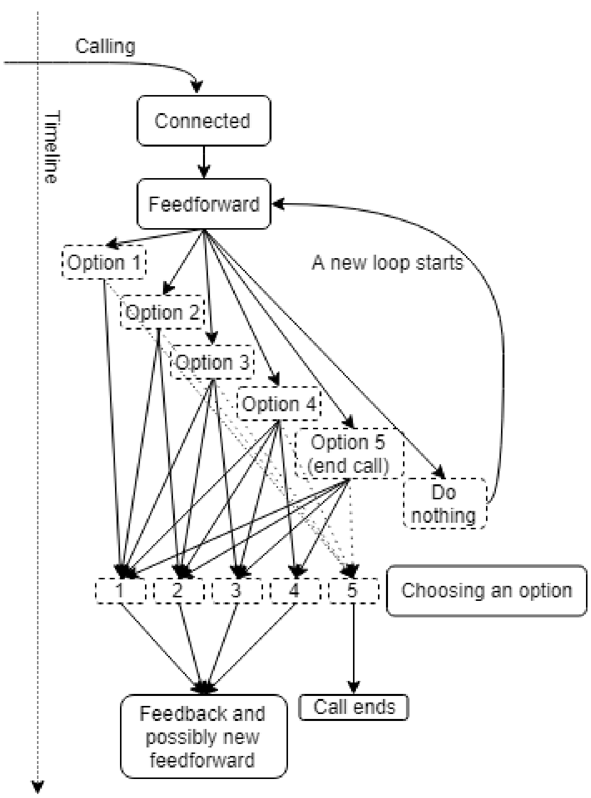


Figure 15: A chronological illustration of a hypothetical answering machine.

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 doesn't know what they are searching for and needs to listen through all the options more than once. As illustrated by Figure 15 the user's options are limited by lack of knowledge of said options, until a certain amount of time has passed. Often this can lead to an overwhelming amount of information, and listening to the options more than once. Overloading in this situation is not a possible solution, since it is limited to one modality, audio.

Harrison et al. acknowledge the need for overloading, and achieve a level of finger overloading by differentiating between different types of finger input - the tip, the pad, the nail, and the knuckle [6]. They found that it is possible to build a touch surface which can accurately differentiate between these distinct inputs, but performed no tests concerning how to signify these affordances.

Pedersen and Hornbæk [14] propose touch force as a modality for overloading. They found that users can accurately control two different levels of pressure, although test users stated that the gesture took some getting used to. Moreover, users expressed fatigue after having touched the screen with increased pressure for a while, indicating that this modality is inferior as an interaction and should be used to a limited degree.

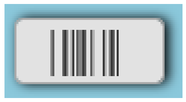
Aslan et al. propose the *gazeover* as a way to implement something similar to the mouseover on a mouse-and-keyboard setup, but potentially available for touch interfaces [2]. However, they do not test user's ability to perceive this affordance on a touch interface.

Breakdown of Previous Design

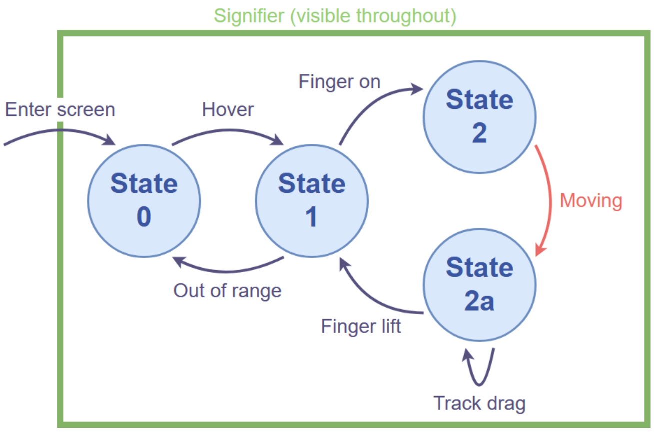
In our previous study on signifiers [4] we designed four signifiers for the same control, two for dragging and two for double tapping. For the dragging control, the two signifiers were a drag handle and a drop shadow, as illustrated in Figures 16(a) and 16(b). Figure 16(c) shows when the signifier is visible in green, and when feedback for the gesture is provided in red. For both dragging signifiers the affordance was present at all times, illustrated by the green border on the box, and the feedback was limited to the regular drag feedback of moving the control. No real differences existed between the two signifiers on a functional level, only the visuals were different.

This was not the case for the two signifiers for the double tapping control. Both rely on temporal information, but the presentation and execution of them are different. In Figure 17(a) the border signifier is presented, and Figure 17(b) shows when the affordance signifier is present and when feedback is given. When the control is single-tapped, the inner border lit up in green and faded away if it wasn't pressed again before the interaction entered timeout. If the control was double tapped, the outer border would also become green. This means that the affordance signifier was present the entire time, but feedback was given three distinct places in the interaction process: at the conclusion of a single tap; in the event of timeout; and at the conclusion of a double tap.

(a)

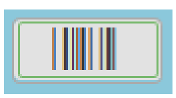


(b)

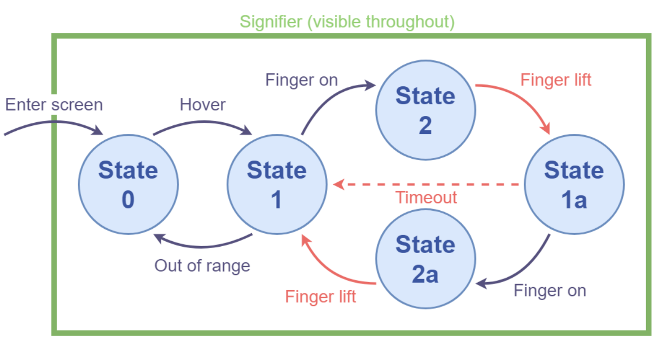


(c)

Figure 16: (a) The drag control with a drag handle signifier. (b) The drag control with a drop shadow signifier. (c) The three-state diagram for the two dragging controls from our previous study [4].



(a)

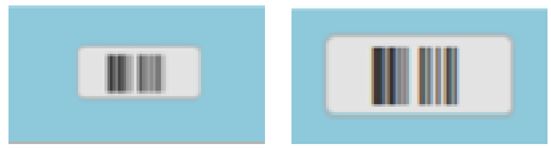


(b)

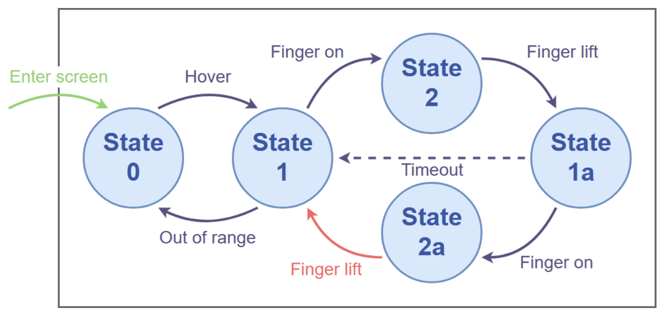
Figure 17: (a) The double tap control with a temporal feedback border signifier. (b) The three-state diagram for the double tap border signifier from our previous study [4].

The other double tap control signifier was a pulsing animation, as can be seen in Figure 18(a), initiated when entering the screen, and then never repeated. This signifier is illustrated in Figure 18(b) as a green arrow when entering State 0. The only feedback given is when the double tap is completed.

When looking over these control signifiers with the three-state model in mind, it becomes clear why our double tap signifiers had a poor success rate. They both relied on temporal signifiers to convey the affordance of the control. In the case of the pulse signifier, it only showed up upon entering the screen, and then there was no way to repeat it, making it easy to miss for the user. The temporal aspect of the border signifier was slightly better in this regard, as the user was able to make the signifier repeat itself. However, the intention of it conveying that the outer border would light up if pressed again was not clear based on only the temporal aspect. So even though there is feedback, the feedback is not particularly informative. The dragging signifiers performed as we expected and their three-state diagram is also exactly like a standard dragging control with the addition of a signifier present at all times.



(a)



(b)

Figure 18: (a) The double tap control with a temporal pulse signifier. (b) The three-state diagram for the double tap pulse signifier from our previous study [4].

A lot of creative solutions exist using different modalities to communicate affordance or implement overloading. However, many lack signifiers for the affordance overload, causing hidden affordances. This is unintuitive if the user is given no instruction, damaging the usability of the application in question. This is a research gap we intend to explore by investigating how, on touch screen devices, we can turn hidden affordances, such as double tap and long tap, into perceived affordances, thereby unlocking intuitive overloading. We will investigate this by focusing on when, during the nanointeractions of a gesture, signifiers appear.

EXPERIMENT

In this experiment we tested whether the temporal placement of a signifier for double tap and a signifier for long tap affected participants' ability to discover the relevant affordance.

Based on the identified research gap we set up two null hypotheses, with the dependent variable being the discoverability of the relevant gesture, and the independent variable being the signifier and the temporal placement thereof:

1. The addition of a signifier does not matter for the discoverability of the affordance of the relevant gesture.
2. The temporal placement of the signifier does not matter for the discoverability of the affordance of the relevant gesture.

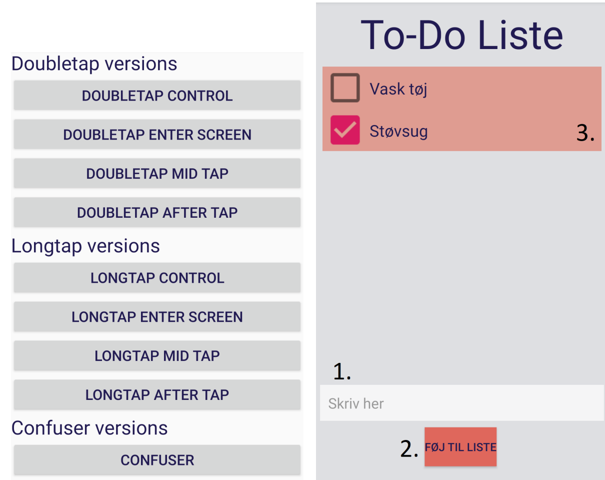
The Prototype

To test the hypotheses, we created a prototype app in Android Studio. This prototype was designed for the purpose of exposing test participants to nine different stimuli as controlled by the test facilitator. These stimuli are as follows:

- Four versions of the app with a double tap affordance:
  - A control version with no visible signifier for the double tap affordance.
  - A version in which a signifier appears upon entering the screen (State 0 in Figure 7). This signifier reappears on a five second loop, repeatedly.
  - A version in which a signifier appears when the user touches the screen (State 2 in Figure 7).
  - A version in which a signifier appears after the user has lifted their finger from the screen, thus completing a single tap (returning from State 1a to 1 on Figure 7).
- Four versions of the app with a long tap affordance:
  - A control version with no visible signifier for the long tap affordance.
  - A version in which a signifier appears upon entering the screen (State 0 in Figure 6). This signifier reappears on a five second loop, repeatedly.
  - A version in which a signifier appears when the user touches the screen (State 2 in Figure 6).
  - A version in which a signifier appears after the user has lifted their finger from the screen, thus completing a single tap (returning to State 1 on Figure 6).
- A confuser version implemented in order to account for the learning factor. In this version, instead of a double tap or long tap affordance, there is a swipe affordance with no signifier indicating this.

Upon launching the app, the screen shows a menu of these nine versions. This menu is not to be seen by test participants, but is a tool for the test facilitators to control which version is applied. This menu can be seen in Figure 19(a).

Once a version has been chosen, the application redirects to the screen seen in Figure 19(b). As can be seen, it is a simple to-do list application which allows users to add items to a list of chores, mark the ones they have completed, and delete items from the list completely. 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 is empty upon button press, no item is added to the list. As can be seen in Figure 19(b), each list item contains a box which may be checked upon tapping it. This box is also checked if the list item is single tapped at all, whether it is on the text or in the empty space to the right.



(a) (b)

Figure 19: (a) The version menu on the prototype application. (b) The main screen of the to-do list app.

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

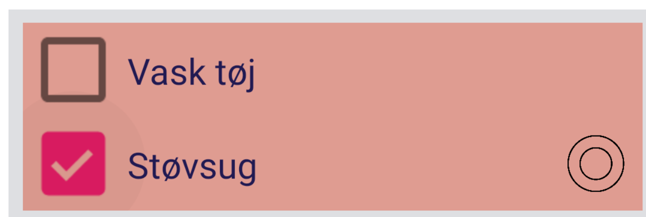


Figure 20: The signifier for the double tap affordance.

As described previously, a signifier revealing the given affordance may appear at various times depending on which version is currently chosen. For the double tap versions (except for the control version), the signifier is a pulse of two rings which expand one after the other, then disappear. This is to emulate the idea of a double tap. A screenshot of this can be seen Figure 20.

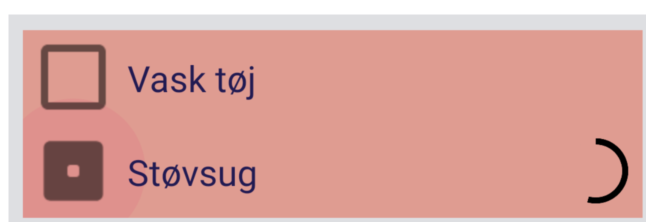


Figure 21: The signifier for the long tap affordance.

For the long tap versions (except for the control version), the signifier is a wheel which gradually fills out over time, indicating that the user may hold their finger on the screen for an extended period of time. A screenshot of this can be seen in Figure 21.

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

Experiment Design

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

The independent variable is the temporal placement of the signifier. For each gesture, four conditions were tested, plus a confuser to slightly alleviate the learning curve of participants:

- A control version with no signifier (Long tap - LC, Double tap - DC).
- An enter screen version, where the signifier appeared after the screen was opened and every five seconds after this on a loop (Long tap - LE, Double tap- DE).
- A version with the signifier appearing during the middle of a tap (Long tap - LM, Double tap - DM).
- A version with the signifier appearing after a tap (Long tap - LA, Double tap - DA).
- A version with no signifier and the required gesture being swipe (C).

The dependent variable is the discoverability of the affordance of the relevant gesture. This is measured by looking at the success rate, the time until success and the amount of different gestures tried before the correct one was performed. This was collected by logging this information within the prototype. Another measure we used was the participants' answers to a series of questions inspired by the NASA Task Load Index (TLX) [11]. We tested the raw TLX method (as described by Hart [7]) in a pilot trial, but since the test participant found the original scale confusing, we tweaked the questions to instead include a smaller scale of 0 to 10. The questions were as follows:

- How mentally demanding did you find the task on a scale of 0-10, where 0 is not at all mentally demanding and 10 is extremely mentally demanding?
- How much did you feel you had to rush when performing the task on a scale of 0-10, where 0 is no time pressure and 10 is extreme time pressure?
- How much success did you have in accomplishing the task, on a scale of 0-10, where 0 is no success and 10 is complete success?
- How much effort did you have to put in when accomplishing the task, where 0 is no effort at all and 10 is extreme effort?
- How irritated, stressed, annoyed, or frustrated did you feel when performing the task, where 0 is none at all and 10 is extreme frustration?

Due to the experiment becoming too long to be able to recruit people of off the street, each participant only tried five conditions: two of each gesture and the confuser 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.

The only requirement for the participants was that they not have a background in interface design. For this test 64 random people in the the age group of 14-77 were recruited off of the streets of Aalborg and or amongst employees at Regionshuset Nordjylland. The experiments were conducted 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. The age distribution is illustrated in Figure 22.

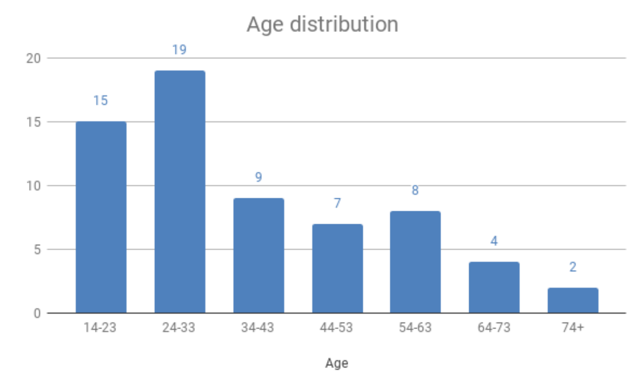
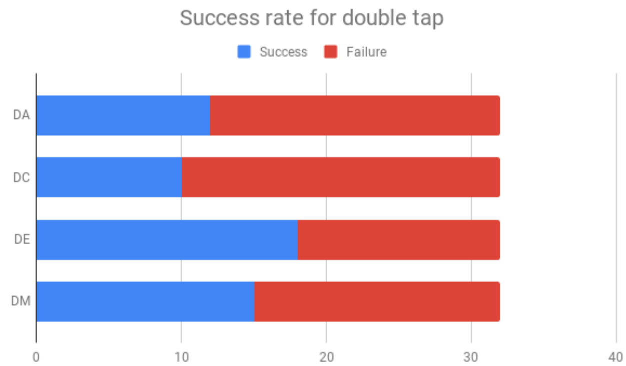


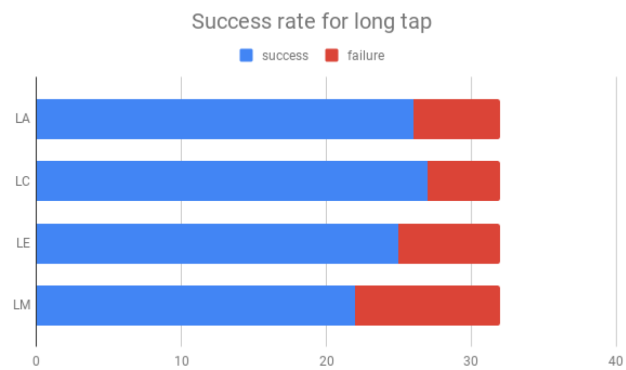
Figure 22: The age distribution of the participants, in intervals of 10 years.

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. When the participant either succeeded in the last task or gave up, they were asked some follow up questions regarding their actions while using the version. They were then asked the TLX-inspired questions. This process repeated with the next four versions, with the third version always being the confuser condition. After the participant finished with the fifth version, the log was copied



(a)



(b)

Figure 23: The distribution of successes and failures of the double tap (a) and long tap (b) versions.

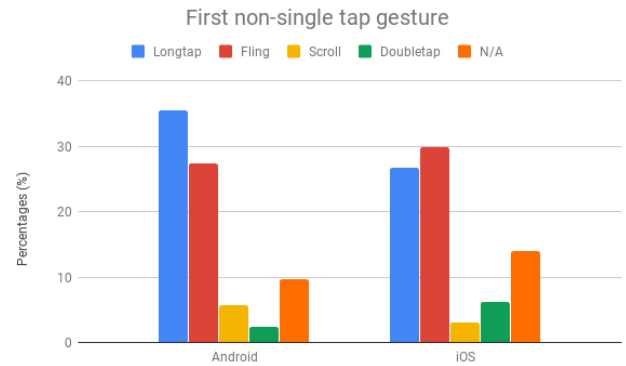
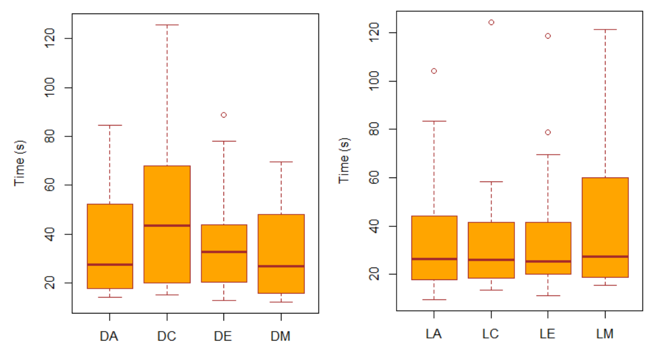


Figure 24: The first gestures performed by participants in each trial, presented in percentages. N/A represents the trials where participants never tried a non-single tap gesture.



(a) (b)

Figure 25: Box plots of the time taken until success of (a) the double tap versions and (b) the long tap versions

from Android Studio to a text file. The entire procedure took approximately 15 minutes from start to finish.

Results

In this section we will present our results from the test. Note that we do not compare results from the different gestures, as it is trivial that users should attempt the long tap gesture more frequently than the double tap gesture.

The Logging Data

The distribution of successes and failures of each version excluding the confuser is illustrated by Figures 23(a) and 23(b). A trial was considered a success if an item was deleted. A quick look on this graph shows a tendency for participants having an easier time discovering the affordance of long tapping than the affordance of double tapping.

An interesting tendency for the first non-single tap gesture performed in each trial when pairing it with the OS the participants are used to, is that Android users are more likely to try a longtap first and iOS users are more likely to try a swipe first. This is illustrated in Figure 24.

Due to only being able to use the successful trials for statistical tests, the sample sizes for each condition reduces from 32 for each to 10 (DC), 18 (DE), 15 (DM), 12 (DA), 27 (LC), 25 (LE), 22 (LM) and 26 (LA). These are all <30 meaning that even though the data gathered from the logging is parametric, we should not use parametric tests. In a case of a sample of this size, outliers also affect the mean quite a bit, so in our case the median is a better representation of the central tendency of the data, which is another reason to choose a nonparametric test. Parametric tests also assume a normal distribution, which all the samples are checked for. When the distribution is not normal we use a Kruskal-Wallis test instead of a one-way ANOVA when comparing the four double tap versions and when comparing the four long tap versions. A Kruskal-Wallis test works by assigning ranks to all the data–from smallest to largest–then dividing it back into the samples, calculating the sum of ranks for each sample and using the mean of the sum of ranks to calculate the test statistic, H. It calculates whether one of the samples differ significantly from one of the others, but does not tell us which one. To discover this, post-hoc tests are needed. Comparing the double tap versions' success rate to each other resulted in ($H(3) = 4.65$, $p\text{-}value = 0.1993$), meaning there is no significant difference between any of the four versions. The same procedure yielded ($H(3) =$

2.54, *p-value* = 0.4681) for the long tap versions, also resulting in no significant difference between the four versions.

Looking at completion time for the successful trials, normality checks were performed and assumptions were not met for either double tap or long tap, therefore Kruskal-Wallis tests were performed. We got ($H(3) = 2.6$, *p-value* = 0.4489) when comparing the four double tap versions, and ($H(3) = 2.65$, *p-value* = 0.8572) when comparing the four long tap versions. The distribution of the data is illustrated in Figure 25. No significant difference was found for either gesture regarding completion time.

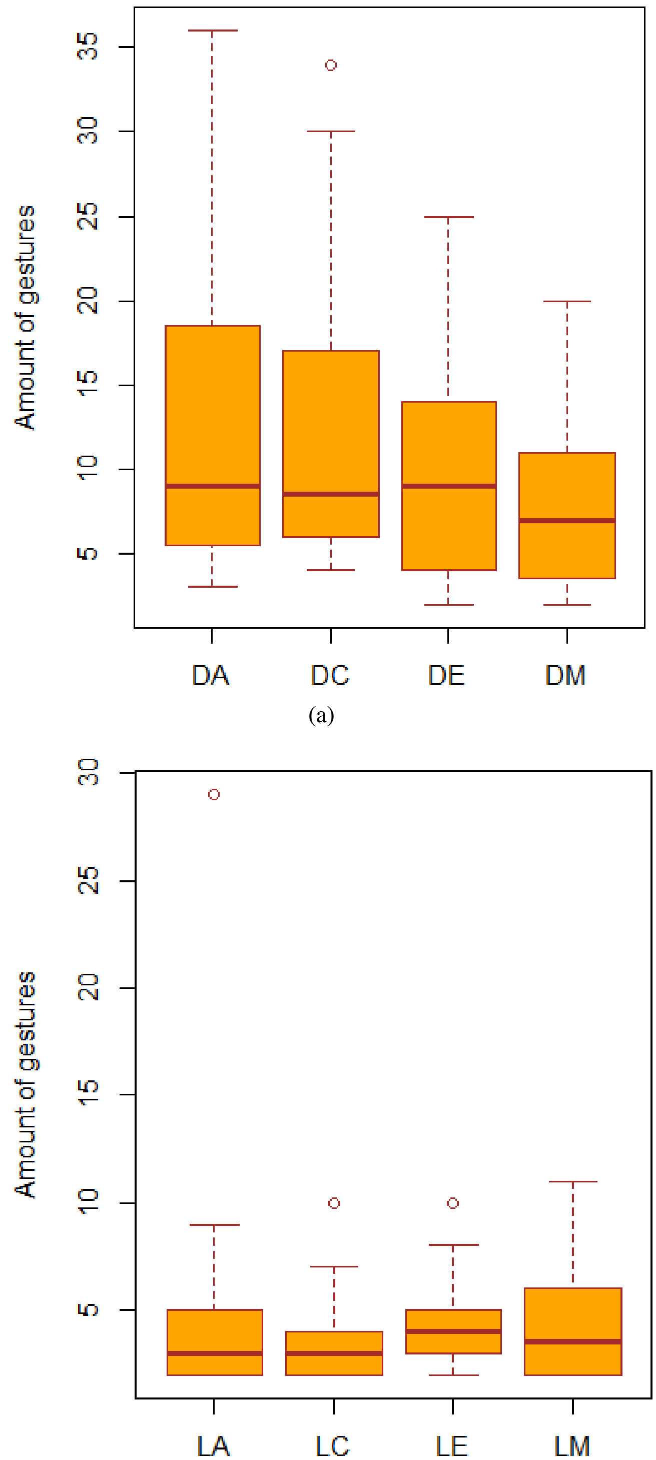
Comparing the amount of gestures performed before success, normality checks were performed and assumptions were not met for double tap or long tap, therefore Kruskal-Wallis tests were performed. This resulted in ($H(3) = 2.62$, *p-value* = 0.4535) for the four double tap versions, and ($H(3) = 3.79$, *p-value* = 0.2849) for the four long tap versions. Neither are significantly different. The distribution of the amount of gestures are illustrated in Figure 26.

Cutting out the single taps from the equation, normality checks were performed and assumptions were not met for double tap or long tap, therefore Kruskal-Wallis tests were performed, yielding ($H(3) = 3.67$, *p-value* = 0.2999) and ($H(3) = 2.61$, *p-value* = 0.4565) for the double tap versions and long tap versions respectively. The distribution is illustrated in Figure 27. No significant difference was found.

Seeing no significant difference in the amount of gestures performed, we look to the variety of gestures performed before success, excluding the single tap. Normality checks were performed and assumptions were not met for double tap or long tap, therefore Kruskal-Wallis tests were performed. Comparing the four versions of double tap yields ($H(3) = 2.63$, *p-value* = 0.4516), and comparing the four long tap versions yields ($H(3) = 2.82$, *p-value* = 0.4199). The distribution is illustrated in Figure 28. No significant difference was found.

The Questionnaire Data

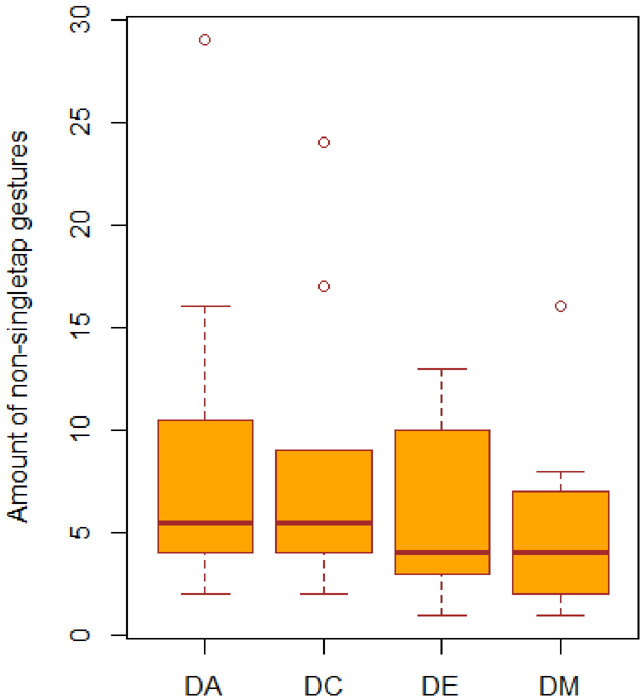
The questionnaire data is analysed based on the five questions focusing on the mental demand, temporal demand, performance, effort and frustration experienced during the trials. The results are compared between the four different versions of long tap and double tap respectively. The overall distribution of answers is visualised in Figures 29 and 30. 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 elected not to answer the questions for two of the trials, which were C and DC. Since we are uninterested in statistically analysing the confuser, the only impact this had was the sample size of DC being reduced to 31. The data gathered from the questionnaire is based on a ranking scale, and thus ordinal and non-parametric, and is analysed by performing a Kruskal-Wallis test, to determine whether any significant differences exist.



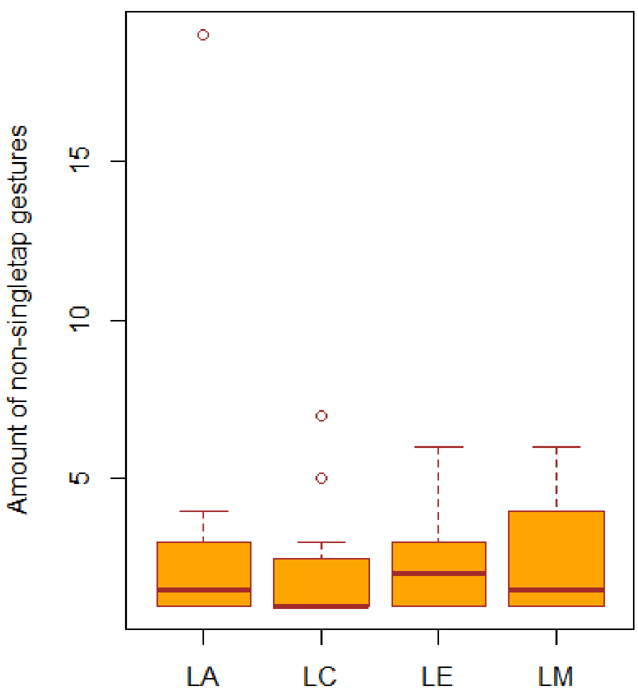
(a)

(b)

Figure 26: Box plots of the amount of gestures performed before success for (a) double tap and (b) long tap.

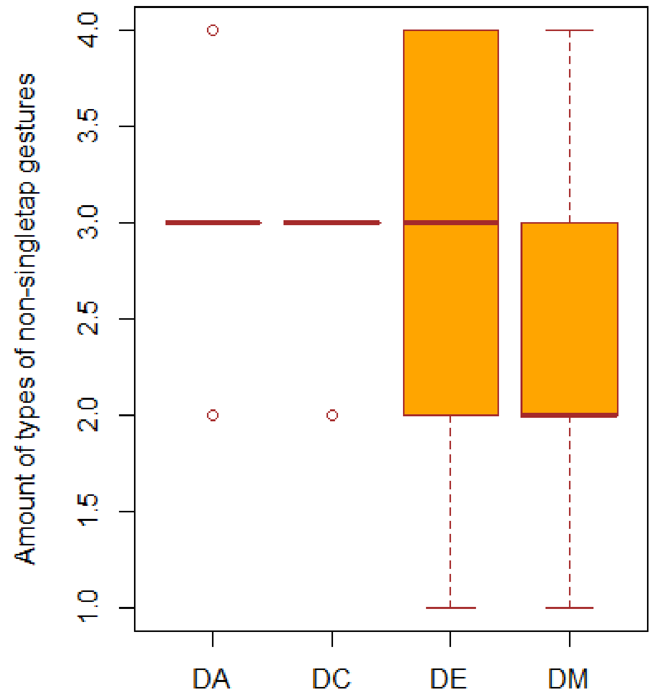


(a)

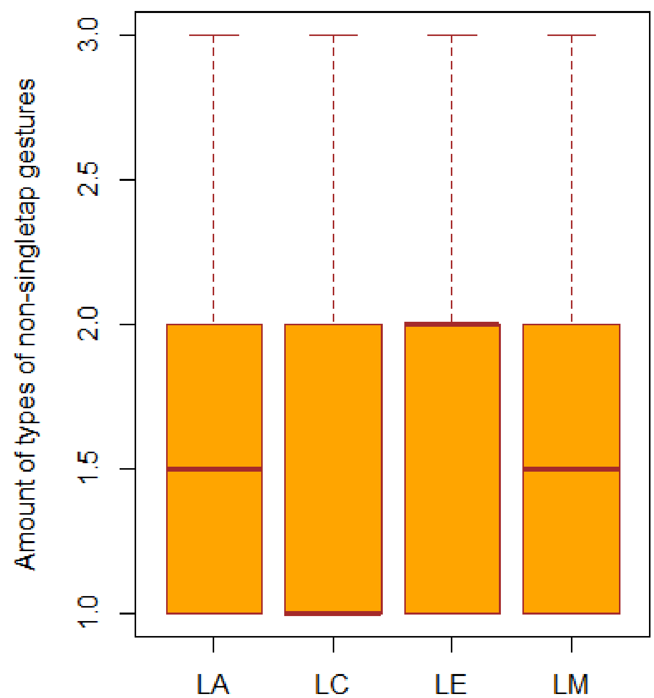


(b)

Figure 27: Box plots of the amount of non-single tap gestures performed before success for (a) double tap and (b) long tap.



(a)



(b)

Figure 28: Box plots of the amount of different types of gestures performed before success, excluding the single tap. (a) shows the results for the double tap versions, and (b) shows the results for the long tap versions.

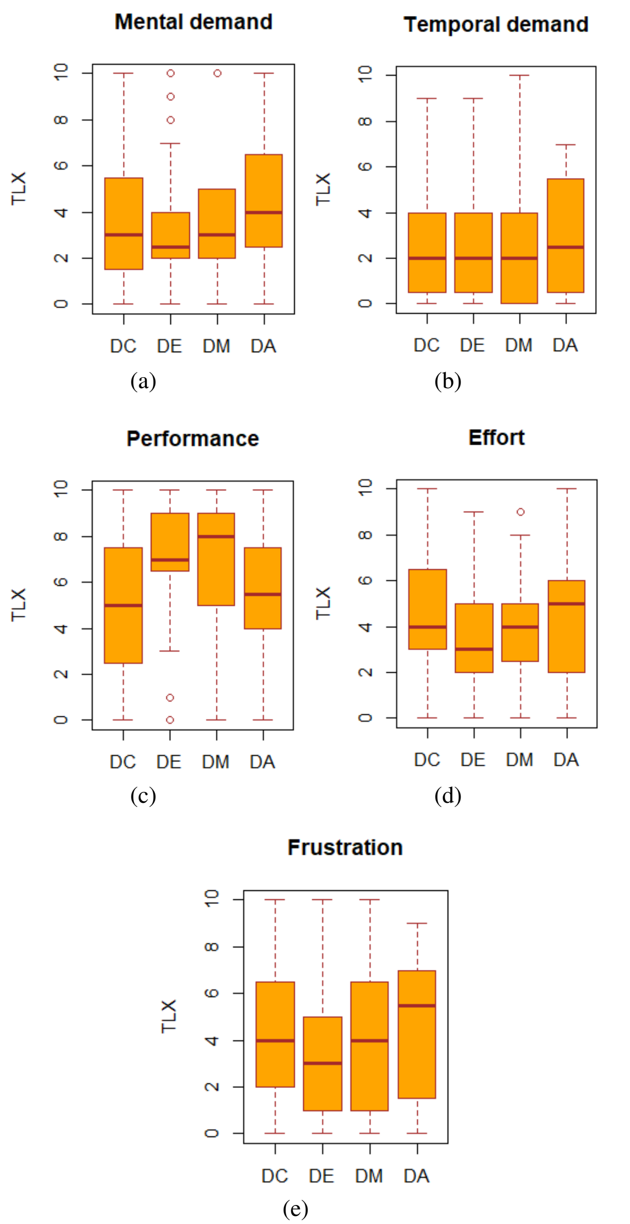


Figure 29: Box plots of TLX data from the double tap versions. (a) mental demand, (b) temporal demand, (c) performance, (d) effort, and (e) frustration.

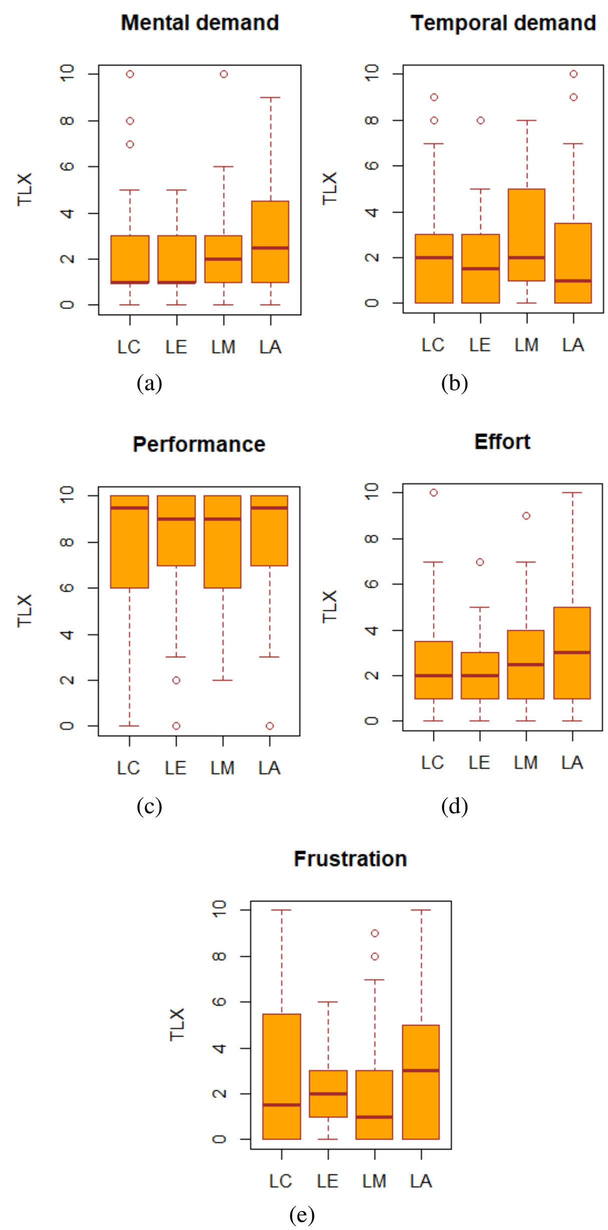


Figure 30: Box plots of TLX data from the long tap versions. (a) mental demand, (b) temporal demand, (c) performance, (d) effort, and (e) frustration.

Comparing the answers for mental demand for the double tapping versions yielded ($H(3) = 3.36$, *p-value* = 0.3388) meaning no significant difference between the four versions was found. The long tap versions yielded ($H(3) = 3.45$, *p-value* = 0.327) and no significant difference.

Temporal demand yielded ($H(3) = 1.35$, *p-value* = 0.717) for double tapping and ($H(3) = 1.63$, *p-value* = 0.6524) for long tapping, and no significant differences.

The answers for performance resulted in ($H(3) = 6.41$, *p-value* = 0.09345) for double tapping and ($H(3) = 0.54$, *p-value* = 0.9097) for long tapping and no significant differences.

The results for effort were ($H(3) = 3$, *p-value* = 0.3914) for double tapping and ($H(3) = 4.66$, *p-value* = 0.1982) for long tapping, and no significant differences.

The frustration data yielded ($H(3) = 2.4$, *p-value* = 0.4931) for double tapping and ($H(3) = 1.34$, *p-value* = 0.719) for long tapping, and no significant differences.

DISCUSSION

The results of our experiment showed no significant difference between the different versions. This means that we can neither disprove that the addition of a signifier does not 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. Even though we couldn't reject our null hypotheses we can still gather information from the data. Based on the logging data and what participants said during the experiment, long tap is the most common gesture when no signifiers are present to indicate other affordances are possible. During the test it was also made clear that a swipe is also a common gesture when wanting to delete an item in the type of application we made, especially for iOS users. Our results show that in this experiment our signifiers did not effectively communicate the affordances of long tap or double tap.

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) heightens the validity of our results, and the fact that the results correspond quite well with the log data also strengthens the validity of our results.

Reliability is based on whether the results can be replicated if the experiment is repeated under the same circumstances. This is affected negatively by the fact that we changed 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 reliability of our results in that aspect. Overall, both the validity and the reliability of our experiment and results are good.

During the test, we discovered some issues which may have impacted our results in some way.

We realised during the test that if the user swipes a list item 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 swiped.

We also realised during the first few trials that the way we phrased the task of deleting an item was misleadingly vague and caused some test participants to 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.

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 was accounted for when analysing the log. 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.

The concept of nanointeractions opens up several alleys of potential future research which may be interesting. 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 (e.g. performing a long tap, but then "cancelling" by moving one's finger) and how designers can take advantage of this.

It may also be interesting to explore whether the order in which the different versions were experienced has an effect on the data. It is possible that, if a gesture (e.g. a long tap) is not possible in the first version a participant experienced, the participant may never attempt that gesture again in later trials where the gesture *is* possible. On the other hand, if the first version has the affordance of the most obvious non-tap gesture to that participant (often long tap), the participant may be more willing to try other gestures, as they have already seen that deleting an item is possible.

We also find that it would be interesting to study the relationship between the position on the screen of a visual signifier and the position on which the user performs a gesture. 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 an issue for signifiers which are always visible.

In this report, 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 if we were to do further research, we would most definitely study the nature of other gestures as well.

We also chose to focus this report entirely on visual signifiers, but other types of feedback may have completely different effects on the user. Future research may reveal an interesting relationship between audio or haptic feedback and nanointeraction stages.

CONCLUSION

While the experiment in itself did not show any significant results, we argue that this project is still highly valuable for future research. Interaction design as a field suffers from differing use of terminology, as researchers have so many different backgrounds. In this report, we contribute to a more unified language when talking about human-computer interaction and the mechanics thereof. We further cement previously determined language, and define terms for levels of interaction not previously discussed in detail, while also analysing current research into different affordance design angles with different modalities with our new 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 open the door to several new considerations. In this report, we focus on the discoverability of gesture affordances depending on whether a signifier is made visible before any gesture is attempted, during a gesture, or after a gesture has been completed. 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. 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 to design affordances for.

ACKNOWLEDGEMENTS

In closing, the authors of this paper would like to give our thanks to everyone who volunteered to participate in our investigation. We thank Aalborg Hovedbibliotek and Region Nordjylland Regionshuset for allowing us to use their facilities for testing. We would also like to thank our supervisor, Hendrik Knoche, without whom this project would not have been possible.

REFERENCES

1. Apple, inc. 2019. iOS Human Interface Guidelines - Buttons. `https://developer.apple.com/design/human-interface-guidelines/ios/controls/`. (2019). Accessed on: 06-03-2019.
2. Ilhan Aslan, Michael Dietz, and Elisabeth André. 2018. Gazeover – Exploring the UX of Gaze-triggered Affordance Communication for GUI Elements. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18)*. ACM, New York, NY, USA, 253–257. DOI: http://dx.doi.org/10.1145/3242969.3242987
3. William Buxton. 1990. A three-state model of graphical input. In *Human-computer interaction-INTERACT*, Vol. 90. 449–456.
4. Emilie Lind Damkjær, Liv Arleth, and Hendrik Knoche. 2018. "I Didn't Know, You Could Do That"-Affordance Signifiers for Touch Gestures on Mobile Devices. In *Interactivity, Game Creation, Design, Learning, and Innovation*. Springer, 206–212.
5. Google. 2019. Material Design - Gestures. `https://material.io/design/interaction/gestures.html`. (2019). Accessed on: 06-03-2019.
6. Chris Harrison, Julia Schwarz, and Scott E. Hudson. 2011. TapSense: Enhancing Finger Interaction on Touch Surfaces. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*. ACM, New York, NY, USA, 627–636. DOI: http://dx.doi.org/10.1145/2047196.2047279
7. Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. DOI: http://dx.doi.org/10.1177/154193120605000909
8. Jeppe Veirum Larsen and Hendrik Knoche. 2017. States and Sound: Modelling User Interactions with Musical Interfaces. In *New Interfaces for Musical Expression 2017New Interfaces for Musical Expression*. NIME.
9. Pedro Lopes, Patrik Jonell, and Patrick Baudisch. 2015. Affordance++: Allowing Objects to Communicate Dynamic Use. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2515–2524. DOI: http://dx.doi.org/10.1145/2702123.2702128
10. Rudy McDaniel. 2015. Understanding microinteractions as applied research opportunities for information

designers. *Communication Design Quarterly Review* 3, 2 (2015), 55–62.

11. NASA. 2019. NASA TLX: Task Load Index. https://humansystems.arc.nasa.gov/groups/TLX/. (2019). Accessed on: 30-07-2019.

12. Donald A Norman. 2008. The way I see IT signifiers, not affordances. *interactions* 15, 6 (2008), 18–19.

13. Ian Oakley, Carina Lindahl, Khahn Le, and Islam MD Rasel Lee, DoYoung. 2016. The Flat Finger: Exploring Area Touches on Smartwatches. *CHI 2016* (2016).

14. Esben Warming Pedersen and Kasper Hornbæk. 2014. Expressive Touch: Studying Tapping Force on Tabletops. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 421–430. DOI: http://dx.doi.org/10.1145/2556288.2557019

15. Ved. 2017. Interaction Design patterns : iOS vs Android. https://medium.com/@vedantha/interaction-design-patterns-ios-vs-android-111055f8a9b7. (2017). Accessed on: 06-03-2019.

16. Jo Vermeulen, Kris Luyten, Elise van den Hoven, and Karin Coninx. 2013. 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*. ACM, 1931–1940.

17. Bill Verplank. 2009. Interaction Design Sketchbook. (2009).