



HAL
open science

Exploring Visual Signifier Characteristics to Improve the Perception of Affordances of In-Place Touch Inputs

Eva Mackamul, Géry Casiez, Sylvain Malacria

► **To cite this version:**

Eva Mackamul, Géry Casiez, Sylvain Malacria. Exploring Visual Signifier Characteristics to Improve the Perception of Affordances of In-Place Touch Inputs. ACM International Conference on Mobile Human-Computer Interaction (MobileHCI 2023), Sep 2023, Athens, Greece. 10.1145/3604257 . hal-04193358v2

HAL Id: hal-04193358

<https://hal.science/hal-04193358v2>

Submitted on 3 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Exploring Visual Signifier Characteristics to Improve the Perception of Affordances of In-Place Touch Inputs

EVA MACKAMUL, Univ. Lille, Inria, CNRS, Centrale Lille, UMR 9189 CRISTAL, France

GÉRY CASIEZ*, Univ. Lille, CNRS, Inria, Centrale Lille, UMR 9189 CRISTAL, France

SYLVAIN MALACRIA, Univ. Lille, Inria, CNRS, Centrale Lille, UMR 9189 CRISTAL, France

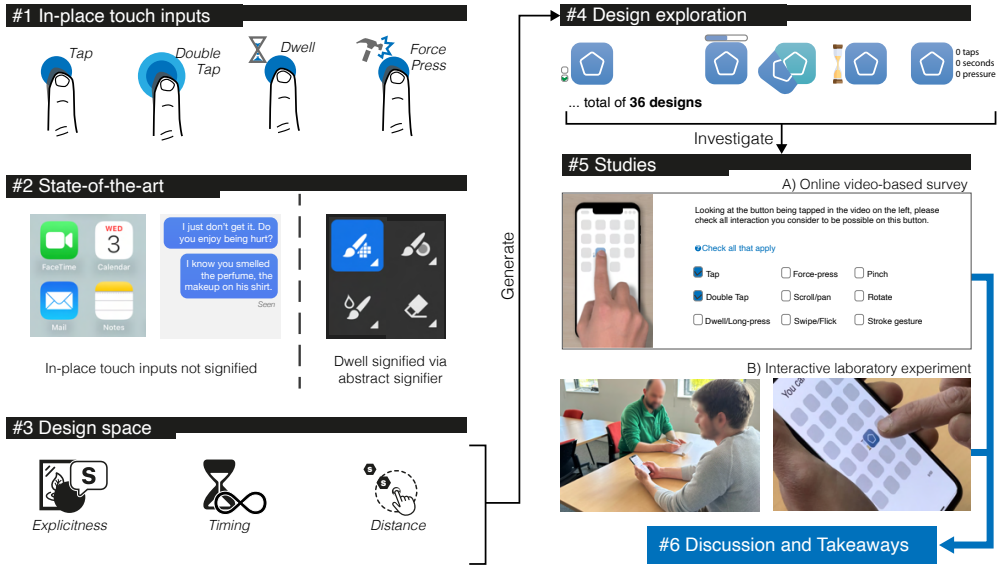


Fig. 1. 1) In-place touch inputs (tap, double-tap, long-press and force-press) are common in modern touch-based devices yet 2) these inputs are almost never signified by the interface. 3) We elaborate a design space of visual signifier characteristics that may impact perception of these inputs, 4) generate 36 designs of buttons that we contrast in 5) an online survey (N=32) and an interactive laboratory experiment (N=24), resulting in 6) takeaways on how touch-based interfaces could be designed to better communicate the availability of in-place touch inputs.

Touch screens supporting different inputs such as ‘Tap’, ‘Dwell’, ‘Double Tap’ and ‘Force Press’ are omnipresent in modern devices and yet this variety of interaction opportunities is rarely communicated to the user. Without visual signifiers, these potentially useful inputs remain unknown or underutilised. We propose a design space

* Also with Institut Universitaire de France.

Authors’ addresses: Eva Mackamul, Univ. Lille, Inria, CNRS, Centrale Lille, UMR 9189 CRISTAL, F-59000, Lille, France, eva.mackamul@inria.fr; G ery Casiez, Univ. Lille, CNRS, Inria, Centrale Lille, UMR 9189 CRISTAL, F-59000, Lille, France, gery.casiez@univ-lille.fr; Sylvain Malacria, Univ. Lille, Inria, CNRS, Centrale Lille, UMR 9189 CRISTAL, F-59000, Lille, France, sylvain.malacria@inria.fr.

Authors’ version

<https://doi.org/10.1145/3604257>

of visual signifier characteristics that may impact the perception of in-place one finger inputs. We generated 36 designs and investigated their perception in an online survey (N=32) and an interactive experiment (N=24). The results suggest that visual signifiers increase the perception of input possibilities beyond ‘Tap’, and reduce perceived mental effort for participants, who also prefer added visual signifiers over a baseline. Our work informs how future touch-based interfaces could be designed to better communicate in-place single finger input possibilities.

CCS Concepts: • **Human-centered computing** → **Interaction techniques**; **User interface design**; **Graphical user interfaces**.

Additional Key Words and Phrases: signifiers, perceived affordances, touch interfaces, in-place touch input, input methods

ACM Reference Format:

Eva Mackamul, Géry Casiez, and Sylvain Malacria. 2023. Exploring Visual Signifier Characteristics to Improve the Perception of Affordances of In-Place Touch Inputs. *Proc. ACM Hum.-Comput. Interact.* 7, MHCI, Article 210 (September 2023), 32 pages. <https://doi.org/10.1145/3604257>

1 INTRODUCTION

Touch screens supporting diverse inputs are now ubiquitous, and yet this variety of interaction opportunities remains underutilised [14, 15, 27]. A possible reason is that, even though touchscreen sensing capabilities steadily improve and the proposed interaction vocabulary accumulates, researchers and practitioners rarely address the problem of how input methods are communicated to users [32]. In the absence of visual signifiers [34, 35] (depictions signalling what can be done) aiding the perception of input opportunities, it is left to the users’ guesswork which input method is available to them at any given point. For example, although we can expect users to be familiar with the input of dwelling on a widget, they may fail to deduce in which context this interaction is meaningful. Typically, users can dwell on application buttons on home-screens or messages in messaging applications to access shortcut menus. But they can also dwell on their camera feed, gallery button, quoted tweets and Slack reactions to name a few. Knowing one of these does not automatically allow the user to extrapolate the others, especially since these UI elements also allow interaction via a simple ‘Tap’. This creates uncertainty, leaving a heavy burden on the user to accurately anticipate when and where a given input is available. Therefore, this absence of visual signifiers requires users, to allow full access to features, to first discover and then recall when an input method is available. Moreover, interfaces frequently changing and new input techniques being introduced [2, 5, 17, 42, 43] make these challenges more and more persistent.

Attempts to improve the discovery and recognition of input possibilities focus on adjusting users’ perception of affordances. A *perceived affordance* refers to any action a user perceives to be possible in a given context [35] (in contrast to *affordances* which cover what a specific user can actually do). While onboarding [26] can assist users in perceiving an affordance, it is usually presented at an inappropriate *timing*, when the interaction is not needed, and requires users to remember it for later use. Supplementary materials, such as instruction manuals on websites, require an explicit search from users and are *away* from the GUI which might explain why they are often not read [6]. Other material like online tutorials or demonstrations in keynotes [32], which only the most tech savvy users would watch, suffer from the same problems of physical and temporal distance.

In an effort to reduce these distances from the point of interaction, that is *where* and *when* the information about the possible input is needed, we investigate how visual signifiers in the interface can improve the discovery of input methods on touchscreens. Given the prevalence of touchscreens and the diverse interaction techniques they support, communicating input possibilities to users presents an important challenge. In this paper, we focus specifically on expanding the perception of *in-place touch inputs* (Figure 1), single finger contacts that do not move across the surface.

Specifically, we focus on variance in duration (*Dwell*), frequency (*Double Tap*) and Force (*Force Press*).

We propose a design space to investigate visual signifier characteristics and how they may influence the success of signifying different inputs. This design space encapsulates high-level principles that materialise in the dimensions *Explicitness*, *Timing* and *Proximity*.

Exploring the design space, we create 36 DESIGNS and examine how they influence the perception of available inputs in a two-step study. First, we conduct an online survey in which participants are presented with a video of a finger tapping the button and asked which inputs they consider possible. Second, we conduct a laboratory experiment where participants interact with a smartphone displaying the different DESIGNS. Participants are instructed to interact with the widget featuring the DESIGNS in whatever way they consider possible, while making the least errors possible.

Altogether, the results inform that different signifier designs suggest different inputs to users and that most added visual signifiers increase the perception of additional input possibilities. Additionally, participants indicated a preference for visual signifiers over a signifier-less baseline and reported DESIGNS with added visual signifiers to require less mental effort. Further, we found that real world metaphors are particularly successful when signifying less known or ambiguous *inputs*, and that timing is especially important when presenting unusual *inputs*. Finally, we found that where designs are ambiguous but strongly suggest an input, initial perceived affordances such as ‘Tap’ are discarded in favour of what is suggested, proposing a need for caution to not inadvertently remove perceived affordances when trying to add others.

In summary, we introduce a design space for visual signifiers, investigate the impact of its dimensions on user perception of possible in-place touch input methods and examine which signifier characteristics are particularly suitable, or unsuitable, to communicate additional input possibilities. In doing so, we provide takeaway considerations for future implementations of visual signifiers.

2 RELATED WORK

Signifiers are any elements that communicate appropriate behaviour to a person, i.e. *what* they can do *where* and *how*. These can take the form of any perceivable indicator such as sound or visual depiction [34, 35]. As such, signifiers are used to inform which actions (affordances [13]) the users perceive to be possible (perceived affordances [35, 36]). The distinction of *perceived affordances* from *affordances* [13] is crucial since the inclusion of signifiers does not alter affordances but aims to improve their perception. The divisive use of the term *affordance* [33] in HCI is reflected in the discourse of affordances on touchscreens. While some argue on-screen objects can *afford* certain gestures [29], others argue they merely *signify* the affordances of touchscreens by making them visible [36]. Regardless of this discourse on terminology, what is displayed guides the perception of possible input methods. Therefore, we review visual signifiers in commercial interfaces, how input possibilities can be communicated, and research on signifying interaction.

2.1 Visual Signifiers in Commercial Interfaces

Visual signifiers to communicate appropriate input in commercial interfaces may vary but are often notably absent. ‘Tap’ as the most common input might be an exception, as it is often communicated through buttons presented as outlines, icons, text, or geometric shapes featuring any of the previous [22, 23]. Meanwhile, other inputs are rarely signified. For example, the availability of ‘Swipe’ gestures is rarely signified [40], but it can be communicated indirectly, for instance with content partially displayed at the edge of the display indicating it can be brought into view [39]. Similarly, the ‘Dwell then Drag’ support of homescreen buttons on iOS and Android is not signified either. Likewise, the Gmail app allows to ‘Tap’ the profile icon to open a popup and switch between accounts but also

allows to swipe up or down on the icon to switch to the next account, which is not signified and subsequently likely to remain unknown.

Regarding in-place touch inputs specifically, they are simply not signified when buttons permit inputs beyond ‘Tap’. Examples are iOS home screen buttons that support the use of ‘Dwell’ and ‘Force’ (on iPhone models with force sensing capabilities) to open a context menu (Figure 2-a). Similarly, the availability of ‘Double Tap’, while available to ‘Like’ content on several social media apps, select text, edit messages in Slack, or zoom in on maps, images and video feeds, is never signified. An exception to the lack of signifiers, can be found in the mobile application Adobe Fresco Version 4.4.2 (2023), which permits users to dwell on menu buttons to access tool specific menus. This is signified by a small triangle pointing outward the right corner of the buttons in question (Figure 2-b), which mirrors the same functionality and signifier in their desktop applications.

Desktop interfaces alongside many of the same signifiers as mobile interfaces (such as buttons, coloured text and progress bars) feature a cursor as an important visual signifier. Since hovering over widgets is possible, the cursor can signify which input is appropriate (e.g. a caret for text input, hand for click) while widgets can indicate interactivity through animation and text pop-ups. Despite this additional visual guidance, the appropriateness of inputs such as double click is still commonly misinterpreted. As there are no visual signifiers indicating which items on screen can be double clicked, it is up to the user to determine when the input is appropriate, leading to interaction mishaps [11, 41] named “Double Dysclicksia” [41]. Once established, a misunderstanding of when ‘Double Click’ is useful and necessary persists, with users wrongly assuming the need to ‘Double Click’ to open links [20].

Both touchscreen and desktop devices, while featuring signifiers for some input, leave the discovery of interactions entirely to the users’ curiosity and supplementary materials. This is notably true for in-place touch inputs which, besides ‘Tap’, are often not signified in current commercial systems.

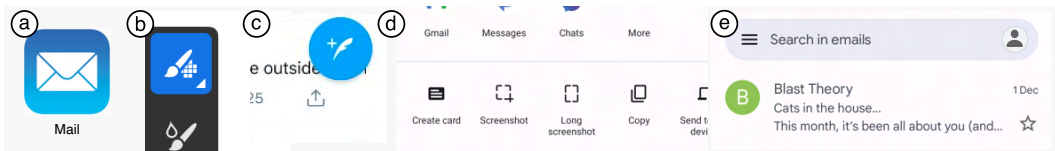


Fig. 2. Examples of visual signifiers in current mobile interfaces. a) iOS home-screen application button b) Adobe Fresco menu c) Twitter floating ‘Add Tweet’ button d) Share panel on Android OS e) Gmail mobile interface

2.2 Communicating Input Possibilities

Since core gestures such as ‘Dwell’ or ‘Swipe’ are available across devices and platforms, the assumption is that users will discover and understand them independently [8]. However, as input availability is inconsistent between contexts, there is a vast burden on the user to discover and recall interactions.

Previous research suggests that, if novel input methods are communicated, this may take the form of explicit instructions through tutorials and pop-ups, or the form of visual depictions, with users indicating a preference for depiction [32] and a dislike of supplementary instructive material [6]. Nonetheless, input opportunities on touchscreens are often not signified at all, and bet on their commercial establishment to reach the point at which they are so conventionally known that their understanding is widely implicit. Examples are often found in keynotes or advertisements, that

introduce an interaction (e.g. Steve Jobs introducing *touch-based scrolling* to great excitement of the crowd [24]).

For depictions, animation has been found efficient [1, 44] while potentially also being a distraction [32]. Additionally, layout and element sizing on touchscreens have been found to impact perceived affordances of touch gestures [28].

When examining visual depictions, different design styles such as *skeuomorphism* and *flat design* are often debated. This debate generally revolves around the assumptions that skeuomorphic design, that mimics the appearance of similar functionalities in the physical world, conveys affordances more successfully, while flat design, more minimalist and depthless, is more aesthetically pleasing [45]. However, in recent years, studies found no significant difference in performance between the two design styles [28, 38, 45]. They suggest that differences may not be attributed to design styles at all, but rather to the general quality of individual designs [38] or the level of abstraction or novelty of an object [28]. It should be noted that a design classified as skeuomorphic does not necessarily reference physical objects successfully and potentially invokes false affordances through *faulty metaphors*, one of the reason why metaphors remain debated in the context of interaction design [16, 36]. Similarly, flat designs are not necessarily detached from references to real world affordances and may employ symbolisms that aim to invoke physical qualities of real-world objects.

2.3 Signifying Interaction in HCI Research

While much research aims to “expand” [18, 43] or “extend” [7, 19] the interaction vocabulary for touch interfaces, very little is said about how it is established or communicated to users. This omission is common not only in research but also commercial products.

Nevertheless, some research expanding interactivity, includes considerations of visual signifiers, either reflecting on how to signify interaction and propose a design [7] or examining how to communicate proposed interactivity in more detail. For example, visual signifiers have been explored in the context of force-based text selection [15], undo mechanisms [3], and digital tabletops [44]. While they studied the success of the interaction method they support, they however do not include separate considerations on the visualisation and its potential impact. Nonetheless, some found the method including visualisation to be more successful than the standard it was compared to [15] while others revealed user difficulties with comprehending novel interaction solely on the basis of visualisation [44]. Interestingly, some visualisation attempted to mirror real world metaphors [3] while others employed established visual concepts such as gauges [15]. In general, users appreciated visual feedback for how their interaction was registered by the system [15, 44]. Lambert et al. explored visual characteristics of signifiers for single-handed microgestures [31], but their guidelines (e.g. show trajectory, show the actuator) are specific to non-stationary gestures and do not scale to in-place touch inputs. Visual characteristics of signifiers have also been explored in the context of directional selection gestures [47], however, while different colours and shapes were considered through qualitative feedback from participants, they only collected quantitative measures regarding the minimum size required to perceive how to interact with their UI element.

Similarly, studies have been conducted to investigate signifiers for existing touch gestures. Arleth et al. explored the effect of temporal placement of added visual signifiers for ‘Dwell’ and ‘Double Tap’ being shown before, during or after a gesture was completed and found neither temporal placement, nor the addition of the signifier in general to be significant [4]. The signifiers employed in their study, were a pulse of two rings expanding and then disappearing after one another for ‘Double Tap’ and a gradually filling ring for ‘Dwell’. Since these were the only designs investigated, and were not validated, it is unclear whether the lack of impact of temporal placement may be partially attributed to misunderstanding of the visualisations rather than the time of their appearance. In contrast, Damkjær et al. focus on different visual depictions when examining signifiers for ‘Drag’ and ‘Double

Tap' [9]. They included two designs for each input method, using a 'drag handle' and drop shadow for 'Drag' and two borderlines and a pulse animation for 'Double Tap'. Notably, both their 'Drag' signifiers mirror existing digital interface designs and are static with no feedback animation, while the 'Double Tap' signifiers are temporary and feature animations, making comparisons between inputs difficult. Similarly to Arleth et al., they found no significant difference between the signifier designs and a control. They note that 'Drag' signifiers were more efficient than 'Double Tap' overall and highlight that this may be attributed to higher familiarity with 'Drag' in general, which they do not control for.

In conclusion, some work has been done to explore individual facets of signifiers and their impact on communicating the possibility of in-place touch inputs. In this work, we propose a more detailed look at multiple signifier characteristics for diverse in-place touch inputs.

3 WHAT MAKES A SIGNIFIER SUCCESSFUL?

While signifiers have been identified as a requirement of "good design" [35] there is a lack of clarity of what constitutes good designs for signifiers themselves. Equally, their success is seen as a binary - something is successfully signified, or not. To allow for a more granular investigation, we define three performance levels for interaction signifiers:

- (1) Does not convey the intended interaction(s).
- (2) Conveys the intended interaction(s) but may convey other, non-intended ones.
- (3) Conveys only the intended interaction(s).

These can then be further separated by differentiating between users who are previously familiar with the signified interaction method and those who are not. The most successful interaction signifier would, e.g., convey specifically this interaction, even to people who were previously not aware of the existence of the interactions, *i.e.* conveys only the intended interaction to (a) individuals unfamiliar with the interaction or (b) individuals already familiar with the interaction. Furthermore it should be considered whether the added signifiers reduce the perception of established inputs therefore interfering with conventional tasks.

When testing and comparing signifiers, these can then be quantified, making the most successful signifiers those that communicate the intended interaction to the largest number of people regardless of their previous familiarity with it, while not interfering with established interactions.

4 DESIGN SPACE FOR VISUAL SIGNIFIERS

To explore how visual signifiers convey specific interaction possibilities, we propose a design space focused on abstract characteristics that can then be realised regardless of aesthetic considerations (unlike colour or style [28, 38, 45]). In doing so, we compile signifier characteristics that have previously been considered in isolation [4, 9].

As visibility is an essential part of signifiers in graphical user interfaces, we consider when (*timing*) and where (*proximity*) a signifier is displayed in relation to the interaction. In addition to these concrete concepts, we examine the level of interpretation expected from users to understand the intended meaning once they have noticed the signifier. Therefore, we include the more conceptual characteristic *explicitness* in the design space. In doing so, we do not consider visuals beyond what certain metaphors may dictate in an effort to allow abstraction and application of the design space to any desired context regardless of design requirements. Similarly, we decided against including *context* as a design space category despite its relevance. This decision was made as (1) context itself is a signifier and often times a major contributor to convey meaning to users, (2) it is not part of the widget itself and (3) contexts are way too diverse to be sufficiently explored in a first

study. As such, it would broaden the design space excessively and would complicate its exploration. Subsequently, we provide the same context through a grey grid layout in all signifier designs.

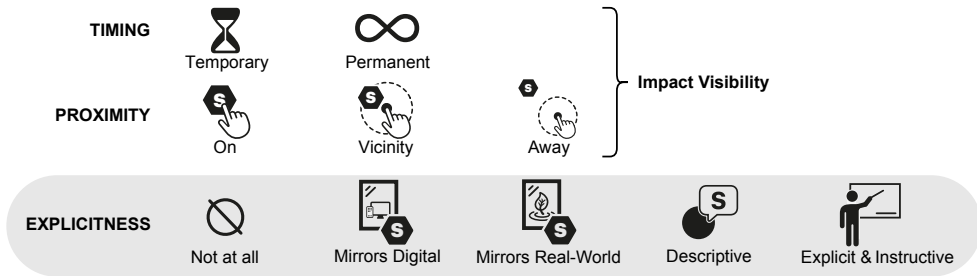


Fig. 3. Visual representation of the signifier design space featuring the characteristics *Explicitness*, *Timing* and *Proximity*. Each characteristic then includes the levels considered for it which are represented through text as well as small graphical representations. Where characteristics are considered to be influencing the visibility of a signifier is indicated through the curly bracket.

Timing. Describes when and for how long a signifier is *displayed* by the interface. The main separation considered in this respect is between signifiers that are permanently displayed and those temporarily displayed (concurrently to interaction). Notably, concurrent interaction feedback within permanently displayed signifiers may still be temporary, while the overall signifier is displayed at all times (e.g. a permanent displayed scale that fills upon interaction). As such, the permanently displayed parts of the design may be considered feedforward informing the potential interaction before it occurs. Our consideration of when the signifier is displayed is made relative to the interaction and is therefore separated into before (feedforward), during (concurrent feedback) and after (terminal feedback) interaction [10].

We consider signifiers to be *permanent* if displayed before, during, and after the interaction; and *temporary* if displayed during and immediately after the interaction. A finer consideration between different levels of temporary visibility exceeds the scope of this study but may be worth considering in future work.

Proximity. We consider proximity to the point of interaction, which can influence the perception of a signifier. It can highlight the connection of the signifier to the widget in question by being close as well as increase the likelihood of the signifier being noticed by being displayed within the area of the users focus. On the contrary, close proximity can also lead to the signifier being occluded by the touch interaction. We propose three levels of proximity. *On the widget*: the visual signifier is superimposed on the widget in question, in our case the button. *Vicinity of the widget*: the signifier is displayed in the area immediately surrounding the widget, but not the widget itself. *Away from the widget*: the signifier is visible elsewhere on the display but not connected to the widget (e.g. the edges of the screen).

Explicitness. Describes how clearly the interaction is communicated to the user and, consequently, how much interpretation, and possibly mental effort, is expected. We propose 5 levels of explicitness: *Not at all explicit*, *Mirror established digital signifiers*, *Mirror physical affordances*, *Descriptive* and *Explicit and Instructive*.

Not at all: signifiers have no link to the interaction intended to be signified. Examples are the complete absence of visual signifiers in Swhidgets [40] or force on the spacebar of iOS keyboards [15].

Mirror Established Digital Signifiers: apply commonly understood digital signifiers to communicate interactivity in a different context. Examples are progress bars, widely used to signify loading content or level of completion.

Mirror Real-World Affordances: aims at leveraging users' understanding of physical objects and processes in order to communicate appropriate interactions. Examples are controls designed to resemble physical buttons or switches, as well as slightly more abstract examples like "stacks" of images or documents that can be moved through by swiping them as if leaving through a physical collection of paper. We do not attempt to create skeuomorphic or photorealistic designs but attempt to take advantage of the users' familiarity with physical objects to improve guessability [25, 46].

Descriptive: show how the system is considering the input but do not instruct appropriate interaction. A typical example are countdowns that may show that the system is recording the progression of time but do not elaborate on how the system further utilises this information, its outcome, or which interaction it triggers.

Explicit and Instructive: unambiguously communicate interaction possibilities to the user and instruct them on how to use them, typically by using text. An example of this is mobile-app onboarding [26] which includes concrete instructions and information on how to interact with and use an application.

Note that these levels may vary in effectiveness depending on the user and their previous experience and therefore should be considered as categorical and not ordinal.

4.1 Current Signifiers in the Design Space

As the most common interaction, 'Tap' is signified most often. Especially popular for signifying 'Tap' are buttons (Figure 2) which are signifier designs *mirroring established digital signifiers* that are *permanent* and *on the interaction widget* (since they largely consist of simply the visual depiction of the interactive widget). Example Figure 2 a), b) and c) signify being tappable through shapes mirroring digital conventions while conveying the effect of the button through mirroring real-world objects. However, a) also supports 'Dwell' and 'Force' that are not visually communicated. Nonetheless, dragging is signified once a user dwells through the icons, starting to 'wriggle' in place, indicating they can be moved. The moving animation mirrors the real world, is concurrent with the interaction and on the widget.

The profile icon in example e) is a button signifying 'Tap' by mirroring established digital signifiers, however the button also supports swiping vertically to switch between accounts, which is not at all signified. Similarly, both example d) and e) allow the user to swipe, while only d) communicates this visually by showing parts of the controls accessible through a horizontal swipe at the corner of the screen. Subsequently swiping on e) represents a lack of signifiers, or a 'Not at all' explicit signifier. The categorisation of d) can be argued, as the partially shown menu option somewhat resembles sliding puzzles mirroring real world affordances but also functions to visually describe its function.

The button in example b) is a rare example of a button communicating not only 'Tap' but also 'Dwell' visually. The arrow in the bottom right corner pointing outwards is well established in desktop applications for changing tools via a right click or a long press, particularly within the Adobe Creative Suite. While this signifier is not adapted in most mobile adaptations of these applications (e.g. Photoshop, Lightroom), Adobe Fresco 4.2.1 on iOS 16.2 utilises it similarly to desktop applications where a short tap simply selects the tool while a longer press opens the

tool selection. This signifier can therefore be classified as mirroring established digital signifiers, permanent and on the widget.

4.2 Filled Design Space

We use an approach inspired by the *research through design* methodology [48] in which we produce various designs to explore and reflect on the characteristics of the design space. More specifically, we created signifier designs for the three selected interaction methods (Double Tap, Dwell, Force Press) combining different levels of explicitness, timing and proximity. The created DESIGNS are instances of a signifier, featuring characteristics explored in the design space. To limit confounding factors, we used buttons, which are widely used and can be expected to be familiar to most users. To account for the 'Not at all' explicitness level, we included a condition of a simple button with no added signifier.

Furthermore, to avoid redundancy in designs, we combined some descriptive and all instructive designs to address all the interactions explored at the same time. While explicit instruction can also take the form of audio or video content, we limited our designs to text and explored the other dimensions (Timing and Proximity) in more detail with this text.

Outside these constraints, the design space was filled semi systematically, with all combinations being considered but those presenting particular difficulties not being implemented. Nonetheless, we created a broad group of designs (36) which are presented in figure 4, covering the design space broadly and fulfilling the majority of possibilities. The only designs which have been previously implemented or tested, are 01 and 24 which are based on commercial interfaces, and 15 and 25 which are adapted from existing signifier research [4, 9].

5 RESEARCH QUESTIONS

We investigate the following questions regarding which *inputs* each DESIGN communicate.

RQ1 - Do added visual signifiers increase the perception of available *inputs*?

RQ2 - Which *inputs* are communicated by which DESIGNS, and therefore what are the perceived affordances?

RQ3 - Do the DESIGNS convey the intended *inputs*?

RQ4 - Does the success of signifiers depend on previous familiarity with the *input* in question?

RQ5 - Does Explicitness, Timing and Proximity impact the success of a DESIGN?

RQ6 - Do signifiers reduce the perceived mental effort required?

RQ7 - Do users prefer the presence of signifiers over standard widgets?

To answer these questions, we conducted 1 - an online survey followed by 2 - an in-person user study. Beginning with an online survey presents two main benefits. First, it provides relatively rapid insights regarding which *inputs* each DESIGN communicates, without biasing users' perception of *input* viability nor suffering from demand characteristics [37] resulting from interpersonal contact between researchers and participants. Second, it informs which *input* to associate as *correct* for each DESIGN for the second, in-person, user study. Then, conducting an in-person study investigates not only which *inputs* users perceive as possible, but which ones they actually perform when exposed to each DESIGN. Research questions 5 and 6 are addressed exclusively by study 2, while all others are addressed in both studies. Altogether, these two studies provide significant elements that answer the above research questions

01		A standard home screen application button for mobile phones - basis for all designs	None (Tap)	13		Two radio-button-style checkmarks displayed next to the button, that fill when the button is tapped	25		Dwell
02		A decreasing number countdown displayed on the button indicating the amount of taps left to interact meaningfully	Double Tap	14		Two empty boxes displayed next to the button, that get ticked off when the button is tapped	26		Force
03		Timer starting to count upwards "rising" from the widget and counting while touch is maintained	Dwell	15		Circles around the button indicating possible taps with the button expanding to each subsequent circle when tapped. Adapted from [4,9]	27		Dwell
04		Text counters next to the widget recording amount of taps, duration and force of the interaction	Double Tap, Dwell, Force	16		A progress bar permanently displayed at the bottom of the screen showing possible taps filling as the button is interacted with	28		Force
05		Short text on the button indicating possible interaction methods appearing when the button is interacted with	Double Tap, Dwell, Force	17		A button folding itself tighter together each time it is tapped	29		Dwell
06		Instructive text appearing next to the button when it is interacted with, listing all interaction possibilities	Double Tap, Dwell, Force	18		The button icon moving as if moving through a magazine when tapped, revealing what's below	30		Force
07		Instructive text appearing at the top of the screen when the button is interacted with, listing all interaction possibilities	Double Tap, Dwell, Force	19		While touch is maintained and force is strong the button background fills with a darker colour	31		Force
08		Instructive text permanently displayed next to the button, listing all interaction possibilities	Double Tap, Dwell, Force	20		Button background colour slowly getting darker while finger contact with the button is maintained	32		Force
09		Instructive text permanently displayed at the bottom of the screen, listing all interaction possibilities	Double Tap, Dwell, Force	21		A circumferential scale or progress bar appearing in the vicinity of the button upon interaction	33		Force
10		The button changing color each time it is tapped	Double Tap	22		A target line appearing at a distance around the widget with the button size increasing as touch is maintained	34		Force
11		A circle appearing around the button when it is tapped showing dashed grey strokes that fill with each subsequent tap	Double Tap	23		A progress bar displayed at the bottom of the screen showing time progress while the button is interacted with	35		Force
12		A progress bar displayed at the bottom of the screen showing possible taps filling as the button is interacted with	Double Tap	24		A downward pointing triangle/arrow in the bottom right corner of the widget	36		Force

Fig. 4. Table of designs filling the design space. Each design is displayed alongside the design space icons that apply to it and a text description. The vertical strip on the right displays the assumed signified interaction.

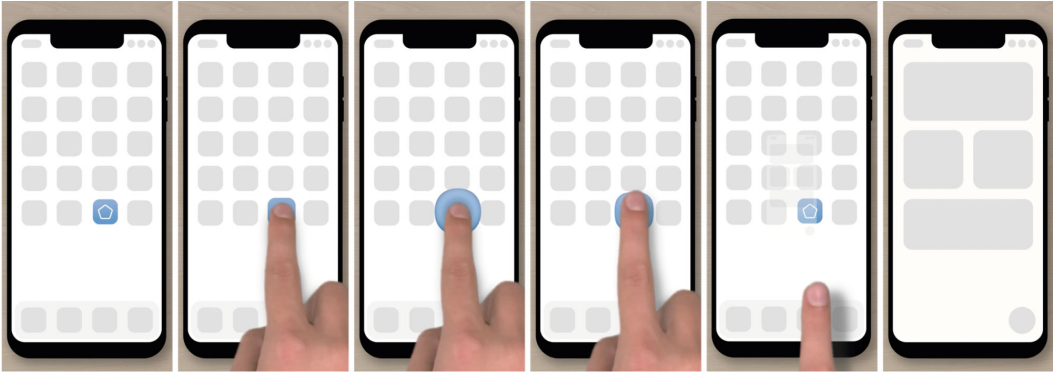


Fig. 5. Setup for presenting DESIGNS in the online study. Images show how the button and its context are displayed throughout the videos depicting them. Each design is shown by itself, before a hand moves into frame and taps the button. In response to the tap, an application opens as the hand moves out of frame.

6 STUDY 1 - ONLINE SURVEY

We first conducted a remote online survey to establish which *input* participants perceive as possible with a given DESIGN, and confirm whether the DESIGNS communicate the *input* we assumed they would when implementing them (*assumed input*), thus justifying a sufficient coverage of all in-place touch inputs.

6.1 Methodology

The survey relies on videos that present each DESIGN implemented in a button, placed on a grid similar to a smartphone home screen, and being ‘tapped’ by a finger (Figure 5). For each DESIGN, participants are invited to select from a list all *inputs* they consider to be possible with the button.

6.1.1 Presenting DESIGNS. Each video shows a finger tapping the button, which triggers a view to open. All videos are 6 seconds long and feature the same tap animation, while the position in the grid and mockup application opening varied. This variation was implemented to clearly distinguish designs from one another, after initial testing of the survey revealed that participants were confused by the changing visual characteristics of what they perceived to be the same button. To minimise variance, interaction feedback is timed to be the same length for all the design visualisations.

DESIGNS are all presented as a button in the same mobile homescreen-esque grid to suggest interaction taking place on a touchscreen. Also, while concurrent and terminal feedback of each DESIGN may vary, we provide the same ‘expected’ outcome to the tap interaction by revealing another screen (Figure 5). Finally, DESIGNS are implemented in a basic button whose shape and icon remain the same, mirroring a mobile phone homescreen application button.

6.1.2 Procedure. Participants accessed the online survey using their own computer. After being informed that the survey examined different interface designs for touch-based interactions, participants were asked to provide informed consent and to answer demographic questions. They then indicated their familiarity with Android and iOS on a 7-point Likert item ranging from 1 - *None* to 7 - *Extremely High*. Then, they were presented with a list of touch-based inputs (Tap, Double Tap, Dwell, Force-Press, Scroll/Pan, Swipe/Flick, Pinch, Rotate, Stroke Gesture) each including a text description and video illustrating the interaction, for which they indicated their familiarity with and usage of on a similar 7-point Likert item. Subsequently, the main body of the survey

began that displayed, for each DESIGN, two main parts (Figure 1, #5-A). On the left, a video-loop where a finger taps a button implementing the corresponding DESIGN, thus showcasing possible concurrent and terminal feedback one would witness if tapping it. On the right, one checkbox for each common touch screen input method (Tap, Double Tap, Dwell / Long-Press, Force-Press, Scroll / Pan, Swipe / Flick, Pinch, Rotate, Stroke Gesture). Participants were instructed to check all *inputs* they considered to be possible with the DESIGN shown in the video. When done, participants proceeded with the next button, displaying a similar page for the following DESIGN. The study ended after participants answered for all DESIGNS.

6.1.3 Participants and Apparatus. The survey was implemented in Limesurvey 5.3.12. 32 participants recruited via direct message completed the survey. They were aged 22 to 58 years ($M=31.7$, $SD=9.6$), with 30 of them indicating that they use a smartphone at least once a week. 21 of them reported a higher familiarity with Android, 7 with iOS and 2 similar familiarities. Overall, they indicated a high familiarity level with their favourite mobile OS ($Mdn=6$, $SD=1.09$).

6.1.4 Experimental Design. The survey was set up to test our main independent variable DESIGN, with the 36 instances illustrated in figure 4. To control potentially confounding variables, presentation order was randomised for each participant. Similarly, to avoid biasing participants, response options were shown either in a “logical” or “random” order¹, but stayed consistent throughout the survey for each participant. This variation was added in order to evaluate the potential influence of presentation order on which answers were selected [30]. Additional independent variables emerge by grouping DESIGNS according to the three main dimensions of design space: EXPLICITNESS, TIMING, PROXIMITY.

In summary, the experiment followed a within-subjects design: 32 participants \times 36 DESIGN = 1,152 forms filled. The survey took an average of 30 minutes per respondent to complete.

Our main dependent variable is **Selection%**: for each *input* \times DESIGN combination, it corresponds to the proportion of participants who considered this *input* to be possible with this DESIGN, regardless of whether this *input* was intended or if others *inputs* were considered to be possible. It is computed as the number of participants who selected this *input* for this DESIGN, divided by the total number of participants (compiled in a table in Appendix C).

For the additional independent variables EXPLICITNESS, TIMING, PROXIMITY, the measures for the dependent variable are first aggregated by computing the average *Selection%* for all assumed inputs for designs corresponding to a level of the design space categories per participant.

6.2 Results

First, we examined the data for outliers, by examining the total number of selected checkboxes per participant (out of the maximum 36 DESIGNS \times 9 interaction methods = 324). We found no outliers and subsequently proceeded using all participants.

Then, we examined a potential effect through examining whether the order of response options impacted the results and compared the selection of ‘Tap’ between the two response order groups. Tap selection was used for this comparison as it is the most basic and established interaction method for mobile touch interface buttons. ‘Tap’ was presented first in the logical order and 7th in the random order, but no significant difference in selection amount was found between the logical and random order (Mann-Whitney $U = 109.5$, $p > 0.05$). We subsequently analyse the results of both orders together.

¹The “logical” order starts with the most common interaction “Tap” and then groups interactions according to the fingers used as well as the similarity and complexity of the interactions. Logical interaction order: ‘Tap’, ‘Double Tap’, ‘Dwell’, ‘Force-Press’, ‘Scroll/Pan’, ‘Swipe/Flick’, ‘Pinch’, ‘Rotate’, ‘Stroke Gesture’; Random interaction order: ‘Scroll/Pan’, ‘Rotate’, ‘Stroke Gesture’, ‘Swipe/Flick’, ‘Force-Press’, ‘Double Tap’, ‘Tap’, ‘Pinch’, ‘Dwell/Long-Press’

6.2.1 *RQ1 - Do Added Visual Signifiers Increase the Perception of Available Inputs?* Visual signifiers, in general, do increase the perception of available *inputs*.

When examining which *inputs* were assumed to be available, 12/16 DESIGNS created for 'Double Tap', 19/20 DESIGNS created for 'Dwell' and 14/14 DESIGNS created for 'Force Press' increase the *Selection%* for the *input* in comparison to the baseline '01 Basic' ('Double Tap' 47%, 'Dwell' 56%, 'Force Press' 31%).

6.2.2 *RQ2 - Which Inputs Are Communicated by Which Design, and Therefore What Are the Perceived Affordances?* Examining the *Selection%* across *inputs* for each DESIGN in study 1, we found clusters of DESIGNS communicating either discrete, continuous or ambiguous *inputs*.

To examine the similarity between DESIGNS in terms of *inputs* selected, we clustered the DESIGNS using a K-mean clustering approach. Each design was considered as a vector of 9 values comprised of the *Selection%* for each *input* method for the given DESIGN. To determine the ideal number of clusters (k) we calculated silhouette scores. This resulted in 5 clusters that can be described as 'Tap' + DESIGNS that signify:

All - 'Tap', 'Double Tap', 'Dwell' and 'Force Press' (DESIGNS 04, 06, 07, 08, 09)

Continuous - 'Tap', 'Dwell' and 'Force Press' (DESIGNS 03, 05, 19, 21, 22, 23, 25, 26, 27, 29, 30, 32, 33, 36)

Discrete - 'Tap' and 'Double Tap' (DESIGNS 02, 13, 14, 15)

Ambiguous - 'Tap' with ambiguous results for 'Double Tap', 'Dwell' and 'Force Press' (DESIGNS 01, 10, 11, 12, 16, 20, 28, 31, 34, 35)

Ambiguous+ - 'Tap' with ambiguous results for 'Double Tap', 'Dwell', 'Force Press' and 'Swipe' (DESIGNS 17, 18, 24)

The **All** cluster encapsulates DESIGNS where participants selected all *inputs* that correspond to in-place touch interaction. It contains all *Explicit and Instructive* DESIGNS as well as the *Descriptive* 04 Text Counters DESIGN. This cluster is explained by the fact that text-based DESIGNS were created to include all *inputs* to avoid excessive repetition.

The **Continuous** cluster contains DESIGNS for which participants selected 'Dwell' and 'Force' alongside 'Tap'. This includes DESIGNS created to convey one or both of the continuous *input* with most of the "scale" DESIGNS falling within this cluster. This grouping shows that the perceived affordance for continuous *input* overlaps, with most DESIGNS aimed to signify 'Dwell' or 'Force Press' signifying the other as well.

The **Discrete** cluster contains DESIGNS for which participants selected 'Double Tap' as well as 'Tap'. Examples of DESIGNS in this cluster, are the checkmarks (13 (Radio) and 14 (Tick)). This clustering indicates that there is a clear difference in perception of affordance of discrete *inputs* versus continuous *inputs*.

Meanwhile, both clusters **Ambiguous** and **Ambiguous+** include DESIGNS with high 'Tap' selection and middling selection percentages for 'Double Tap', 'Dwell' and 'Force'. Cluster **Ambiguous+** contains DESIGNS for which some participants additionally selected 'Swipe'. These DESIGNS were unclear in what they conveyed and, in the case of **Ambiguous+**, conveyed a non in-place touch input. Notably, 17 *Folding* had a high selection rate for 'Dwell' (84%) but a low score for 'Force Press' as well as a relatively high selection rate for 'Swipe'

This clustering highlights the difficulty of achieving signifiers that successfully convey only the intended interaction(s) (see section 3). Since the *inputs* 'Dwell' and 'Force Press' are both perceived as continuous, the *Selection%* for them tends to rise in tandem. The highest difference between 'Dwell' and 'Force Press' *Selection%* can be seen in 3 DESIGNS. 03 *Text Timer* which is a top performing DESIGN for 'Dwell' has a *Selection%* that is 40 percentage points lower for 'Force Press'. Interestingly, the other two DESIGNS featuring the highest difference between the two

continuous *inputs* (17 *Folding* - 53pp difference, 31 *Hourglass* - 44pp difference) are both mirroring real world affordances.

6.2.3 RQ3 - Do the Designs Convey the Intended Inputs? Overall, the DESIGNS convey the intended *inputs*.

Using an arbitrary threshold of 75% to consider that a DESIGN communicates the assumed *input*, which means that above 75% of participants considered this *input* as possible to perform with the corresponding DESIGN, we found 22 DESIGNS that communicate at least one of the assumed *inputs*. Looking in detail for each *input*, we found 9/16 DESIGNS for 'Double Tap' (Mdn=75%, M=65%, SD=22%), 16/20 for 'Dwell' (Mdn=91%, M=85%, SD=13%), 9/14 for 'Force Press' (Mdn=77%, M=76%, SD=9%).

Note that looking at assumed *inputs* is not the ultimate aim of our study as, in the end, what matters is what *inputs* participants did report suggesting which affordances were perceived. Interestingly, there were 4 DESIGNS, which were not assumed to communicate 'Dwell', that had a *AssumedSelection%* over the 75% threshold (11 *Circumferential Steps*: 81%, 17 *Folding*: 84%, 32 *Scale Vertical*: 94%, 33 *Squeeze*: 84%).

Further, it should be noted, that 5 designs (17 *Folding*, 23 *Scale Distant Temporary*, 25 *Scale Circular*, 32 *Scale Vertical*, 33 *Squeeze*) fall under the 75% threshold for 'Tap' which should be signified by the button itself and its context as well as the example interaction in the videos. Notably, all of these designs do score high selection percentages for other interaction methods, suggesting that strong signifiers for one input may reduce the perception of otherwise established inputs.

6.2.4 RQ4 - Does the Success of Signifiers Depend on Previous Familiarity With the Input in Question? We did not find evidence that previous familiarity with an *input* impacts *Selection%*.

We ran a Pearson's test between both reported 'familiarity' and 'usage' and *Selection%*.

We examined both the correlations overall, and the correlations for DESIGNS where the *input* was assumed to be signified (vertical text in Figure 4). For *Familiarity*, we found a moderate linear correlation for 'Tap' ($r=0.41$) and 'Dwell' (Assumed $r=0.57$, Overall $r=0.46$), but none for 'Double Tap' (Assumed $r=-0.03$, Overall $r=0.14$) and 'Force Press' (Assumed $r=0$, Overall $r=0.02$). Regarding reported *usage* of an *input*, we found a low linear correlation for 'Tap' ($r=0.27$) and 'Dwell' (Signified $r=0.35$, Overall $r=0.26$), but none for 'Double Tap' (Signified $r=-0.15$, Overall $r=0.14$) and 'Force Press' (Signified $r=-0.07$, Overall $r=0.06$).

6.2.5 RQ5 - Does Explicitness, Timing and Proximity Impact the Success of a Design? There are significant differences in the success of DESIGNS depending on design space categories. *Explicit* and *Permanent* DESIGNS perform better while DESIGNS with signifiers *On* them generally perform worse.

EXPLICITNESS. All levels work well. *Descriptive* works best, followed by *Explicit*, *Mirrors Digital* and *Mirrors Real World*.

As each level of EXPLICITNESS is associated to a different number of designs, we first aggregated the corresponding designs for each level of EXPLICITNESS and each participant using the average *Selection%* for assumed inputs. Note that *Descriptive* is represented by only three DESIGNS (02, 03, 04). A Friedman test showed a significant effect of EXPLICITNESS on *Selection%* ($\chi^2(3)=22.6, p < 0.001$). Post-hoc comparisons using Wilcoxon signed-rank tests revealed significant difference ($p=0.01$) between the *Mirrors Real World* (Mdn=80%) and *Descriptive* (Mdn=100%) levels. Further there were significant differences ($p < 0.05$) between *Mirrors Digital* (Mdn=84%) and *Mirrors Real World* levels (see Figure 6).

The different levels of EXPLICITNESS can also be analysed for each individual assumed input. The different levels of EXPLICITNESS affect *Selection%* differently depending on the assumed input with 'Double Tap' particularly affected, 'Dwell' affected and 'Force Press' not affected.

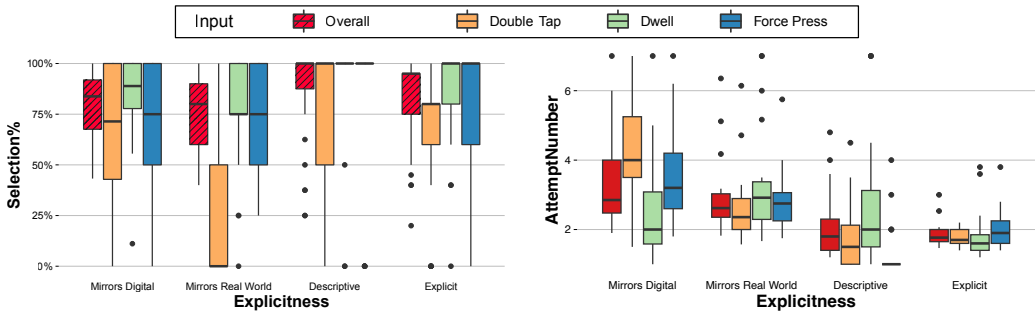


Fig. 6. Left: Average *Selection%* by EXPLICITNESS level for each *input* study 1. Right: Average *AttemptNumber* by EXPLICITNESS level for each *input* study 2.

For ‘Double Tap’, a Friedman test showed a significant effect ($\chi^2(3) = 35.8, p < 0.001$) of EXPLICITNESS on *Selection%* with post-hoc revealing significant differences ($p < 0.001$) between the *Mirrors Real World* (Mdn=0%) and all other levels (*Descriptive*: Mdn=100%, *Explicit*: Mdn=80%, *Mirrors Digital*: Mdn=71%).

For ‘Dwell’, a Friedman test showed a significant effect ($\chi^2(3) = 23.5, p < 0.001$) of EXPLICITNESS on *Selection%* with post-hoc revealing significant differences ($p < 0.01$) between the *Descriptive* (Mdn=100%) and *Mirrors Digital* (Mdn=89%) as well as *Mirrors Real World* (Mdn=75%). It further showed a significant difference ($p < 0.05$) between the *Explicit* (Mdn=100%) level and *Mirrors Real World*.

Finally, for ‘Force Press’ only, no effect of EXPLICITNESS was found on *Selection%*.

TIMING. *Permanent* is more efficient than *Temporary* for ‘Double Tap’. A Wilcoxon signed rank test found a significant effect ($V = 403.5, p < 0.001$) of TIMING on *Selection%*, showing that *Permanent* (Mdn=89%) is more efficient than *Temporary* (Mdn=82%). An investigation for each assumed input shows a significant effect ($V = 375, p < 0.001$) of TIMING only for ‘Double Tap’ (*Permanent*: Mdn=93%, *Temporary*: Mdn=56%).

PROXIMITY. *On* has lower performance compared to *Vicinity* and *Away*, especially for ‘Double Tap’ and ‘Dwell’. A Friedman test revealed a significant effect ($\chi^2(2) = 21.7, p < 0.001$) of PROXIMITY on *Selection%*. Post-hoc analysis revealed that *On* (Mdn=77%) is significantly ($p < 0.001$) less efficient than *Vicinity* (Mdn=88%) and *Away* (Mdn=92%). An investigation for each assumed input revealed a significant effect ($\chi^2(2) = 28.9, p < 0.001$) of PROXIMITY on *Selection%* for ‘Double Tap’ with post-hoc analysis showing significant differences ($p < 0.001$) between the *On* (Mdn=40%) and both other levels (*Away* (Mdn=75%), *Vicinity* (Mdn=86%)). Similarly, there is a significant effect ($\chi^2(2) = 28.1, p < 0.001$) of PROXIMITY for ‘Dwell’ with post-hoc analysis showing significant differences ($p < 0.002$) between *On* (Mdn=80%) and both other levels (*Away* (Mdn=100%), *Vicinity* (Mdn=95%)). No significant effect was found for ‘Tap’ and ‘Force Press’.

7 STUDY 2 - INTERACTIVE EXPERIMENT

We then conducted an in laboratory interactive user study.

7.1 Methodology

7.1.1 Setup. The experiment setup (Figure 1-5B) consists of a mobile application featuring a homescreen grid populated with grey squares with one DESIGN featured in the middle, mirroring the online study. In the bottom right corner a progress counter is displayed. At the top, a text is displayed which states “You can start” until the participant starts interacting with the button. If the

supported *input* is performed, the text states “GOOD!” and a green button stating “Move to next trial” appears in the middle of the screen. When any other input is performed on the button, the text displays a counter of errors. After 6 errors, a red button stating “Move to next trial” appears. This repeats for all DESIGNS with the counter at the bottom indicating progress throughout.

Presenting Designs. All DESIGNS are presented interactively, i.e. if they feature animated elements they react when interacted with. We split the DESIGNS featuring all *inputs* in the design space in Figure 4 (04, 05, 06, 07, 08, 09) into individual DESIGNS for ‘Double Tap’, ‘Dwell’ and ‘Force Press’ (e.g. 04A ‘Double Tap’, 04B ‘Dwell’, 04C ‘Force’). Similarly, we include the ‘01 Basic’ three times, once for each intended *input*. This leads to 21 DESIGNS to be included (7 times 3) in addition to the other 29 DESIGNS resulting in a total of 50 DESIGNS included in the experiment. The interactive nature of the experiment requires *inputs* to trigger an event that is either correct or incorrect. We determined a ‘correct’ *input* for each of the other 29 DESIGNS by following a ‘draft pick’ system to have all *inputs* represented equally. In the order of ‘Double Tap’, ‘Dwell’, ‘Force’ the *inputs* were assigned the DESIGN with the highest *Selection%* for the *input* from the online survey, removing the DESIGN from the pool to be picked from, until all DESIGNS were assigned to an *input*. This resulted in 10 additional DESIGNS for ‘Double Tap’ (02, 10, 11, 12, 13, 14, 15, 16, 17, 18) and ‘Dwell’ (03, 19, 22, 23, 24, 25, 27, 29, 30, 31) and 9 DESIGNS assigned to ‘Force’ (20, 21, 26, 28, 32, 33, 34, 35, 36).

7.1.2 Procedure. Mirroring the procedure in section 6.1.2, participants were asked demographic questions as well as about their familiarity with mobile operating systems and interactions, however ‘Scroll’, ‘Pinch’ and ‘Rotate’ were removed from the rating, as they were not likely to be considered according to the results of study 1. Following this, participants were given a smartphone running the application described in section 7.1.1. Participants were instructed to progress through all DESIGNS while making the least errors possible, and encouraged to comment on their thought process throughout. They were told that one of the *inputs* they just indicated familiarity with would achieve this and that it may vary between DESIGNS, but were given no further guidance on how to achieve this beyond the visual signifiers in the DESIGNS themselves. At the end of the experiment, participants were asked to rate the mental effort demanded to reach the correct *input* for all DESIGNS, as well as their personal preference on a 7-point Likert item.

7.1.3 Participants and Apparatus. The survey part of the experiment was implemented in Limesurvey 5.6.1, and the experiment in typescript, using PixiJS 7.2.2 for the rendering. It was conducted on an iPhone XS running iOS 16.4 which supports force sensing. Each design was created using Adobe After Effects and exported as an image sequence to create animations based on the input from the user. Input was classified using the the state machine presented in Appendix A.

Participants were recruited through direct messages. The experiment was completed by 24 participants (different from study 1 except for 2) aged 20 to 59 ($M=27.8$, $SD=9.8$) with all of them indicating that they use a smartphone at least once a week. 4 had a higher familiarity with Android, 12 with iOS and 8 an equal familiarity. Participants rated their overall smartphone expertise as $Mdn=5$, $M=5.33$ on the 7-point Likert scale, with the lowest being 4 - Average.

7.1.4 Experimental Design. Our main independent variable remains DESIGN. However, in cases where multiple *inputs* were combined in one single DESIGN, they were split into separate DESIGNS, leading to 50 instances, as described in section 7.1.1. We maintain the additional independent variables EXPLICITNESS, TIMING, PROXIMITY. The experiment followed a within-subject design with 24 participants \times 50 DESIGNS = 1200 DESIGNS interacted with. It took an average of 21 minutes to complete. We used a pseudo random order of presentation with the condition that the ‘correct’ input could not be the same more than twice in a row. The presentation order was inverted for half of the participants.

Our dependent variables are *Attempt%*, *AttemptNumber*, mental effort rating and personal preference rating. For each *input* × DESIGN combination, *Attempt%* corresponds to the proportion of participants who attempted to interact through this *input* with the given DESIGN, regardless of whether this *input* was correct. This mirrors *Selection%* in study 1. *AttemptNumber* for each DESIGN corresponds to the interaction attempt at which participants performed the *correct input* for the given DESIGN. If participants did not perform the *correct input* within the 6 trials, *AttemptNumber* was saved as 7, representing 'Failed'. Mental effort and personal preference ratings were measured on 7-point Likert items.

7.2 Results

First, we examined the data for outliers by examining the total trials performed per participant (out of the maximum 50 DESIGNS × 6 permitted trials = 300). We found no outliers and subsequently proceeded using all participants.

Then, we examined a potential effect of order by examining the trials required to perform the correct input for each DESIGN between the two order groups, and found a significant difference for only one DESIGN (Design 15 - $U=39$, $p=0.04$). Since the DESIGN in question is in 23rd place in the regular order and 27th in the reverse order, while the other 49 DESIGNS with larger variance in placement show no significant difference, we decide to ignore this minor effect. Therefore, we analyze the results by considering all participants as a single group.

7.2.1 RQ1 - Do Added Visual Signifiers Increase the Perception of Available Inputs? Visual signifiers, in general, do increase the perception of more diverse available *inputs* while interacting.

When considering the individual *inputs*, 13/16 DESIGNS for 'Double Tap', 12/16 DESIGNS for 'Dwell' and 13/15 DESIGNS for 'Force' increase the *Attempt%* for the *input* in comparison to 01A, 01B and 01C. However, note that the results of study 2 are less indicative than those of study 1, given that the participants were only given a chance to interact until they performed the correct *input* for a given DESIGN.

When considering which *inputs* participants attempted first in study 2, out of the 1200 total DESIGNS interacted with, the first *input* performed was 'Dwell' (400 times), 'Tap' (393 times), 'Double Tap' (173 times), 'Force Press' (155 times) and 'Swipe' (79 times), showing that participants frequently attempted different *inputs* than 'Tap'.

7.2.2 RQ2 - Which Inputs Are Communicated by Which Design, and Therefore What Are the Perceived Affordances? The *Attempt%* for the first two attempts only in study 2 showed how perceived affordances change beyond the initial perception of temporary signifiers but revealed no meaningful cluster.

This study influences the perception of available *inputs* by expecting a 'correct' *input*. Therefore, we examine only the first attempt for each DESIGN and participants rather than the full *Attempt%*.

When examining which *input* has the highest *Attempt%* in the first interaction, for each of the 50 DESIGNS, 22 are 'Tap' (DESIGNS 01A, 01C, 03, 05A, 05B, 05C, 6B, 6C, 12, 17, 18, 19, 20, 22, 23, 27, 28, 29, 30, 33, 34, 35), 7 'Double Tap' (02, 04A, 08A, 09A, 13, 14, 25), 17 'Dwell' (DESIGNS 01B, 04B, 06A, 07A, 07B, 07C, 08B, 09B, 10, 11, 15, 16, 21, 26, 31, 32, 36), 3 'Force' (DESIGNS 04C, 08C, 09C) and 1 'Swipe' (DESIGN 24).

We then examined the *Attempt%* when looking at both the first and second interaction attempt rather than just the first, to incorporate participants perception following the first exposure to temporarily displayed DESIGNS. For 11 DESIGNS, 'Tap' achieves the highest *Attempt%* after 2 attempts (01A, 01C, 05C, 10, 12, 17, 18, 19, 20, 28, 29). For another 13 DESIGNS, it is 'Double Tap' (02, 04A, 05A, 06A, 07A, 08A, 09A, 13, 14, 15, 16, 25, 35), for 21 DESIGNS it is 'Dwell' (01B, 03, 04B, 05B, 06B, 07B, 08B, 09B, 11, 15, 19, 21, 22, 23, 25, 27, 29, 30, 31, 33, 34), for 8 it is 'Force' (04C, 06C, 07C, 08C, 09C, 26, 32, 36) and for 1 it is 'Swipe' (24). Note, that 4 DESIGNS have an equally high *Attempt%* after 2 attempts for 2 *inputs*

(‘Double Tap’ & ‘Dwell’ for 15, 25; ‘Tap’ & ‘Dwell’ for 19, 29). Silhouette scores, revealed that there were no meaningful clusters, which may be due to no further interactions being performed once the *correct input* has been performed.

7.2.3 RQ3 - Do the Designs Convey the Intended Inputs? This study reveals, that, when compared to the baseline, most DESIGNS improve the perception of the intended *inputs*.

Considering the same 75% threshold from study 1 and only the *Attempt%* for the correct intended *input* for each DESIGN, only 2 DESIGNS (‘Double Tap’ 17 - Folding, ‘Dwell’ 19 - Scale Vertical Internal) fall under it. However, it should be noted, that the confirmation through interaction and the potential to arrive at the correct input through trial and error are confounding factors for these numbers.

When examining the *Attempt%* for only the first interaction performed by each user only 8 DESIGNS pass the 75% threshold (‘Double Tap’ 08A, 09A, 13, 14; ‘Dwell’ 08B, 09B, 31; ‘Force’ 04C, 08C). When extending this to the *Attempt%* for the first two interactions by each user, this grows to 21 DESIGNS (‘Double Tap’ 02, 04A, 05A, 06A, 07A, 08A, 09A, 13, 14; ‘Dwell’ 05B, 06B, 07B, 08B, 09B, 31; ‘Force’ 04C, 07C, 08C, 09C, 26, 32).

Comparison to Baseline. When comparing the *AttemptNumber* of the DESIGNS to the respective baseline for their *input* (Double Tap’ DESIGNS to 01A, ‘Dwell’ DESIGNS to 01B and ‘Force’ DESIGNS to 01C), a Wilcoxon revealed that there was a significant difference ($p < 0.05$) between the design and the baseline for 22/47 DESIGNS (11/16 ‘Double Tap’ DESIGNS, 3/16 ‘Dwell’ DESIGNS and 8/15 ‘Force’ DESIGNS). Out of those, only 2 DESIGNS (‘Double Tap’ DESIGN 17 ($p = 0.010$, $M = 4.96$), ‘Dwell’ DESIGN 24 ($p = 0.013$, $M = 3.92$)) performed significantly worse than the baseline (‘Double Tap’ $M = 3.63$, ‘Dwell’ $M = 2.71$. ‘Force’ $M = 3.71$), while all others significantly improved the perception of their respective *input*.

The measure of *AttemptNumber* illustrates (see figure 7), how the effectiveness of DESIGNS can vary widely, despite achieving similar *Attempt%* depending on how quickly participants perform a communicated *input*.

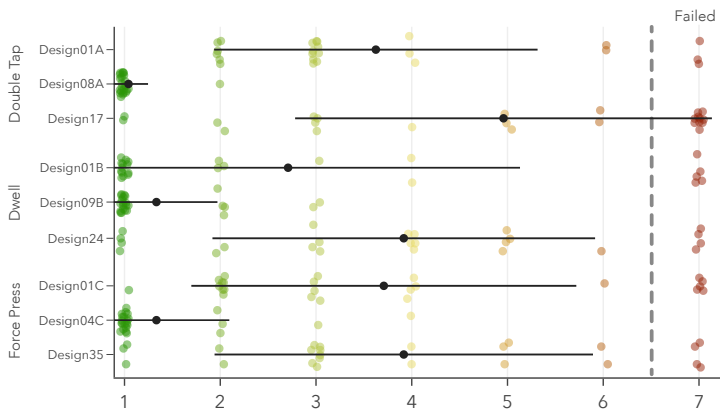


Fig. 7. Trials required to achieve the correct interaction subset showing the baseline, as well as best and worst performing DESIGN for each *input*. The full figure is shown in appendix E.

7.2.4 RQ4 - Does the Success of Signifiers Depend on Previous Familiarity With the Input in Question? We did not find evidence that previous familiarity with an *input* impacts *Attempt%*.

We ran a Pearson’s test between ‘familiarity’ and *Attempt%* for all three *inputs* and found no correlation (‘Double Tap’ $r(22) = -.07$, $p > 0.05$; ‘Dwell’ $r(22) = .06$, $p > 0.05$; ‘Force’ $r(22) = .07$, $p > 0.05$).

7.2.5 RQ5 - Does Explicitness, Timing and Proximity Impact the Success of a Design? There are significant differences in the success of DESIGNS depending on design space categories. *Explicit* and *Permanent* DESIGNS perform better while DESIGNS with signifiers *On* them generally perform worse.

The nature of this study offers confirmation for correct *inputs* and allows participants multiple trials. Therefore the average *Attempt%* is much higher than the average *Selection%* from study 1. However, since we collect *AttemptNumber*, we can compare not only whether an *input* was considered possible, but also when.

EXPLICITNESS. *Explicit* works the best, followed by *Descriptive*, *Mirrors Real World* and *Mirrors Digital*.

We aggregated the corresponding DESIGNS corresponding to each EXPLICITNESS level and each participant using *AttemptNumber*. A Friedman test showed a significant effect of EXPLICITNESS on *AttemptNumber* ($\chi^2(3) = 56.6, p < 0.001$). Post-hoc comparisons using Wilcoxon signed-rank tests revealed a significant difference ($p < 0.001$) between the *Mirrors Digital* (Mdn=2.85) and *Descriptive* (Mdn=1.8) as well as *Explicit* (Mdn=1.77) levels. There were also significant differences ($p < 0.001$) between *Mirrors Real World* (Mdn=2.62) and *Descriptive* (Mdn=1.8) as well as *Explicit* (Mdn=1.77) levels. (Figure 6).

The different levels of EXPLICITNESS can also be analysed for each individual assumed input. In contrast to study 1, the different levels of EXPLICITNESS affect *AttemptNumber* for all *inputs*.

For 'Double Tap', a Friedman test showed a significant effect ($\chi^2(3) = 48.93, p < 0.001$) of EXPLICITNESS on *AttemptNumber* with post-hoc revealing significant differences ($p < 0.001$) between all levels except for *Explicit* and *Descriptive*.

For 'Dwell' a Friedman test showed a significant effect ($\chi^2(3) = 23.94, p < 0.001$) of EXPLICITNESS on *AttemptNumber* with post-hoc revealing a difference ($p < 0.001$) only between *Explicit* (Mdn=1.6) and *Mirrors Real World* (Mdn=2.92).

For 'Force Press' a Friedman test showed a significant effect ($\chi^2(3) = 49.34, p < 0.001$) of EXPLICITNESS on *AttemptNumber* with post-hoc revealing differences ($p < 0.005$) between all levels except between *Mirrors Digital* and *Mirrors Real World*.

Given that the design space category 'Explicitness' aims to categorise how much interpretation and therefore mental effort is expected from the user, we also compared the mental effort ratings for the DESIGNS grouped according to explicitness level (Figure 8). There is a significant difference between the mental effort for 'Explicit' DESIGNS and 'Descriptive' ($U = 2396, p < 0.001$), 'Mirrors Real World' ($U = 6892, p < 0.001$), 'Mirrors Digital' ($U = 13030, p < 0.001$) and 'Not at all' ($U = 191, p < 0.001$). Additionally, there is a significant difference between 'Descriptive' and 'Mirrors Real World' ($U = 7322.5, p < 0.05$) as well as 'Not at all' ($U = 178.5, p < 0.001$). There is also a significant difference between 'Mirrors Real World' and 'Mirrors Digital' ($U = 55321.0, p < 0.005$), 'Mirrors Real World' and 'Not at all' ($U = 1130.0, p < 0.001$) as well as 'Mirrors Digital' and 'Not at all' ($U = 1220.0, p < 0.001$).

TIMING. *Permanent* is generally more efficient than *Temporary*. A Wilcoxon signed rank test found a significant effect ($V = 13, p < 0.001$) of TIMING on *AttemptNumber*, showing that *Permanent* (Mdn=1.98) is more efficient than *Temporary* (Mdn=2.56). An investigation for each assumed input shows a significant effect of TIMING for 'Double Tap' ($V = 8, p < 0.001$, *Permanent*: Mdn=1.71, *Temporary*: Mdn=2.89) as well as 'Force' ($V = 15, p < 0.001$, *Permanent*: Mdn=1.93, *Temporary*: Mdn=2.94).

PROXIMITY. *On* performs worse compared to *Vicinity* and *Away*. A Friedman test revealed a significant effect ($\chi^2(2) = 33.25, p < 0.001$) of PROXIMITY on *AttemptNumber*. Post-hoc analysis revealed that *On* (Mdn=2.92) is significantly ($p < 0.001$) less efficient than *Vicinity* (Mdn=2.09) and *Away* (Mdn=2.21).

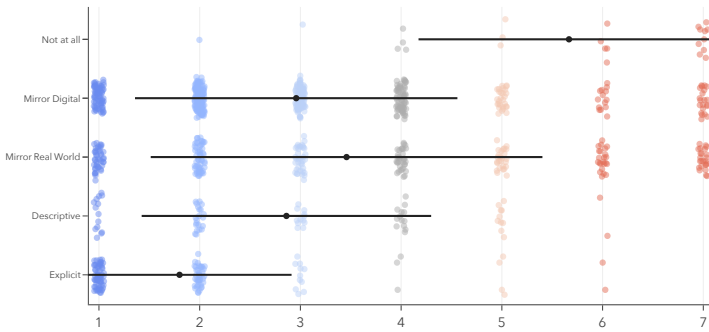


Fig. 8. Plot showing the mental effort ratings grouped by EXPLICITNESS of the corresponding DESIGNS.

An investigation for each assumed input revealed a significant effect ($\chi^2(2) = 16.695, p < 0.001$) of PROXIMITY on *AttemptNumber* for ‘Double Tap’ with post-hoc analysis showing significant differences ($p < 0.001$) between the *On* (Mdn=3) and both other levels (*Away* (Mdn=2), *Vicinity* (Mdn=1.86)). There is a significant effect ($\chi^2(2) = 11.77, p < 0.005$) of PROXIMITY for ‘Dwell’ with post-hoc analysis showing significant differences ($p < 0.05$) between *On* (Mdn=2.67) and both other levels (*Away* (Mdn=2), *Vicinity* (Mdn=2.17)). Similarly, there is a significant effect ($\chi^2(2) = 27.85, p < 0.001$) of PROXIMITY for ‘Force’ with post-hoc analysis showing significant differences ($p < 0.05$) between *On* (Mdn=3.1) and both other levels (*Away* (Mdn=2), *Vicinity* (Mdn=2.25)).

7.2.6 RQ6 - Do Signifiers Reduce the Perceived Mental Effort Required? Overall, signifiers reduce the perceived mental effort required to find an *input*.

Subjective responses in study 2 (figure 9-left) reveal that the majority of DESIGNS are perceived to require a lower mental effort. Wilcoxon tests comparing the mental effort of the individual DESIGNS to the ‘01 - Basic’ baseline, show a significant difference ($p < 0.05$) for all DESIGNS except DESIGNS 17 ($M=5.54, Mdn=6.0$) and 24 ($M=5.63, Mdn=6.0$). All other DESIGNS were perceived as requiring significantly less mental effort. Overall, DESIGNS 08 ($M=1.29$), 09 ($M=1.54$) and 14 ($M=1.67$) were perceived as the least mentally demanding (Mdn=1.0), followed by DESIGNS 06 ($M=1.58, Mdn=1.5$) and 31 ($M=2.13, Mdn=1.5$) and finally with a median of 2.0, DESIGNS 04 ($M=2.33$), 05 ($M=2.67$), 07 ($M=1.92$), 13 ($M=1.83$), 16 ($M=2.79$), 22 ($M=2.46$), 23 ($M=2.58$), 25 ($M=2.29$), 26 ($M=2.71$), 27 ($M=2.50$), 35 ($M=2.67$) and 36 ($M=2.13$).

7.2.7 RQ7 - Do Users Prefer the Presence of Signifiers Over Standard Widgets? In general, users prefer the inclusion of signifiers over a baseline without them.

While most DESIGNS received varying preference ratings (see figure 9-right), they were widely preferred over a complete absence of signifiers. Wilcoxon tests comparing the preference ratings of the individual DESIGNS to the baseline showed a significant difference ($p < 0.05$) for all except DESIGNS 17 ($M=2.17, Mdn=2.0$) and 24 ($M=1.86, Mdn=1.0$). All other DESIGNS were preferred over the baseline ($M=1.75, Mdn=1.5$), the favourite ones being DESIGNS 04 ($M=4.67, Mdn=5.0$), 06 ($M=4.92, Mdn=5.0$), 09 ($M=4.88, Mdn=5.0$), 14 ($M=4.67, Mdn=5.0$), 35 ($M=4.63, Mdn=5.0$) and 36 ($M=4.71, Mdn=5.0$).

Note that Pearson’s tests revealed significant correlations between personal preference and mental effort as well as interaction attempts for all three *inputs* (‘Double Tap’, ‘Dwell’, ‘Force’). Notably, there is a high negative correlation between personal preference and mental effort rating (‘Double Tap’ ($r(22)=-.71, p < 0.001$), ‘Dwell’ ($r(22)=-.69, p < 0.001$), ‘Force’ ($r(22)=-.66, p < 0.001$)). Meanwhile, there is only a low to moderate negative correlation between personal preference and interaction attempts (‘Double Tap’ ($r(22)=-.43, p < 0.001$), ‘Dwell’ ($r(22)=-.20, p < 0.001$), ‘Force’ ($r(22)=-.19, p < 0.001$)).

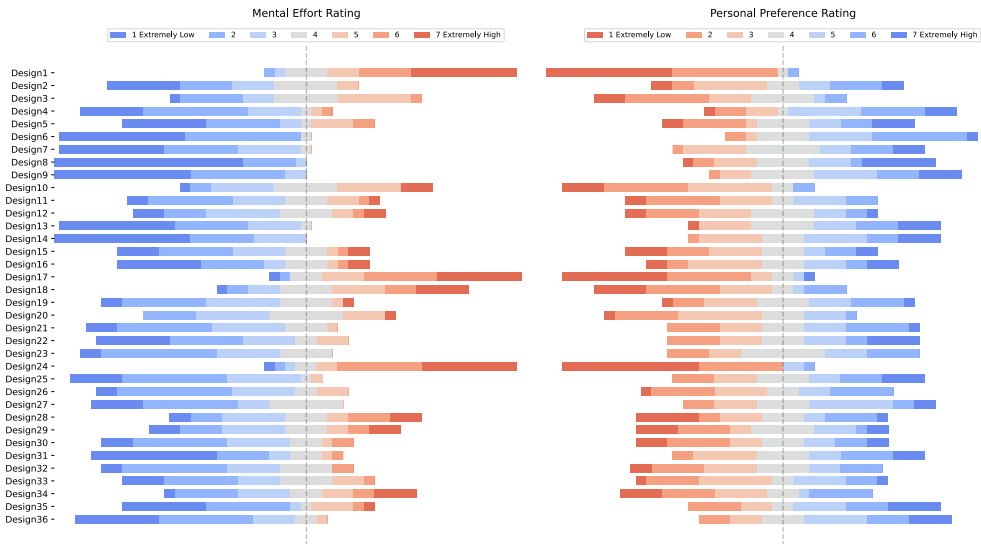


Fig. 9. Stacked bar graphs showing likert scale responses for mental effort required (left) and personal preference (right).

8 DISCUSSION

The results of both studies largely align, with study 2 confirming the results of study 1 for an interactive context. This suggests that video-based surveys may present a viable option to initially test this type of phenomenon.

In the following section, we discuss the implications for design and limitations of our studies.

8.1 Lessons Learnt From Our Studies

Takeaway 1: *Users prefer added visual signifiers.* Despite the lack of significant difference in performance for some DESIGNS and *inputs* in study 2, all but 2 DESIGNS (17, 24) were perceived as less mentally demanding. This suggests that these DESIGNS can still improve the subjective experience when these *inputs* should be communicated. This is further supported by participants indicating a significant preference for the DESIGNS with added visual signifiers over the baseline except the aforementioned 2 DESIGNS. Overall, this still suggests the usefulness of visual signifiers for improving user experience with a widget.

Takeaway 2: *Which signifier is best depends on context and target audience.* The preference ratings for the different DESIGNS did vary between participants. For instance, P21 considered the permanently displayed instructions to be “a waste of space to a certain degree”. In contrast, P13 would prefer redundancy in signifiers, including explicit ones, noting that “the best would be to first have text to tell me what to do and then a scale that tells me how I’m doing”. In addition, several participants noted that they may prefer certain DESIGNS in some contexts. For example, P12 noted for DESIGN 35, that “this is fun [...] if it’s in a game I’d like it, but if it was my alarm clock I’d hate it”.

Takeaway 3: *When inputs are established, adding signifiers may be unnecessary.* In general, the results of both studies suggest that signifiers increase the perception of available inputs and reduce mental effort. However, study 2 revealed that for ‘Dwell’ only 2 DESIGNS performed significantly better than the baseline 01B, those DESIGNS both being permanently displayed explicit text (DESIGNS 08 & 09). This can partially be attributed to the fact that ‘Dwell’ is a well established *input* and was

frequently performed as the first *input* attempt regardless of DESIGN. An example of this, is P10 stating that “[their] default [interaction] is just kind of hold it there and see if there is anything that pops up and act accordingly”. As a result, the ‘Dwell’ baseline DESIGN 01B performed significantly better than the baseline for the less common *inputs* ‘Double Tap’ and ‘Force Press’. This suggests that if an *input* is established, added signifiers do not increase the perception of the *input*’s availability overall, questioning their necessity.

Takeaway 4: When signifying ‘added’ *input* methods, proceed with caution to not inadvertently remove the perception of availability of established ones. Analysis of the Selection% in study 1 according to the design space revealed a significant difference in ‘Tap’ responses between different explicitness levels, in spite of ‘Tap’ always being shown in the animation, suggesting that the explicitness of a design may impact previously established perceived affordances. Furthermore, designs for which the response to ‘Tap’ does not surpass our 75% effectiveness threshold achieved high selection percentages for other interaction methods. This may indicate that for ambiguous design that still strongly suggest an interaction, initial perceived affordances are discarded in favour of the new signified one. Given this observation, designers need to be careful not to inadvertently remove perceived affordances when attempting to add additional ones.

Takeaway 5: The perception of signifiers varies between users, be aware that signifiers may be interpreted in unexpected ways. If signifiers are not explicit, designers need to account for users who may interpret them in unexpected ways. For example, for the DESIGNS 04A, 04B, 04C which displayed explicit counters above the icon, the majority of participants interpreted the text as intended. However, P12 interpreted the text as a warning telling them not to tap, not linger and apply zero pressure to the button. Similarly, P10, when presented with a DESIGN including a scale interpreted it “kind of a slider” and tried to swipe on the scale, while the majority of other participants considered scales as an indication of continuous *input*, leading them to either dwell or force press. These differences suggest that even widely understood visual guidance can be interpreted in various ways.

Takeaway 6: Users’ previous experiences with contexts and devices impact their perception of metaphors. When interacting with DESIGN 18, P4, once they had noticed the animation, attempted to dwell on the button and expressed their surprise when the *input* was considered incorrect, stating “on the kindle page flip happens on dwell so I expected that”. Similarly, P11 attempted to ‘Double Tap’ both DESIGNS 29 and 35 multiple times, commenting “from the [Nintendo] switch, I’m sort of programmed to tap multiple times in quick succession - that’s just my go to instinct”. These quotes highlight, that users extrapolate meaning from previous interactions. If DESIGNS mirror other usages of similar visual metaphors, they should either align with what is conveyed in those other use cases, or very clearly diverge visually.

Takeaway 7: Real world metaphors are particularly successful when signifying less known or ambiguous interactions. Explicit instructions very efficiently communicated *inputs*. However, in study 1, for both ‘Double Tap’ and ‘Dwell’ the most successful DESIGN was one mirroring established digital signifiers. In contrast, ‘Force Press’ is the only interaction method whose top results include DESIGNS that mirror real world affordances. Furthermore, in study 2 multiple participants commented on the difficulty of differentiating between continuous *inputs* but pointed to DESIGN 31 (Hourglass) as an exception, e.g. P7 stated “for this one it was clear it’s time related [in comparison to the scales]”. This may suggest that for interactions that are less established, leaning on real world metaphors is particularly helpful.

Takeaway 8: Timing matters for less established *inputs*. The comparison of *permanent* and *temporary* DESIGNS for both studies shows that *permanent* DESIGNS perform better. Notably, when considering the DESIGNS separated by *input*, there is no significant difference for ‘Dwell’ suggesting

temporarily displayed signifiers may sufficiently communicate more established *inputs* while less common *inputs* may benefit from extended visibility.

Takeaway 9: *Existing visual guidance may only work in context - consider this when adapting it to a new context.* The 24 Arrow DESIGN was not successful in conveying ‘Dwell’ (53% Selection%, 83% Attempt%, M AttemptNumber = 3.92) despite being the only commercial visual signifier we found for that *input* (figure 2b). Even worse, half of the participants in study 1 did not consider the signified interaction to be possible at all, and it was one of only two DESIGNS that (1) participants did not prefer over the baseline and (2) did not reduce the perceived mental effort. While this may be due to our participants being unfamiliar with Adobe Fresco, or Adobe products in general, it is certainly noteworthy that one of the few existing visual signifiers for mobile interfaces appears to not convey the intended meaning universally. A potential explanation may be that the context of a tool menu within an application is as much of a signifier as the visualisation itself and the visual arrow depiction only works within this context. Further work could investigate how visual depictions work in tandem with the context in which they are placed to convey meaning to users.

8.2 Clarity of Signifiers for Similar Interactions

Clustering the response patterns for different designs in study 1 revealed that, while designs tend to be most effective in communicating a specific input, similarity between the interactions increases the likelihood that multiple inputs are considered possible. Continuous inputs such as ‘Force Press’ and ‘Dwell’ could be clearly separated from the discrete input of ‘Double Tap’. This was echoed in study 2, with the overall most frequent comments being about the difficulty to differentiate between ‘Dwell’ and ‘Force’ (e.g. *“the problem between force and dwell with scales is it is hard to tell which one you are supposed to do”*-P6). Similarly, multiple participants commented on ‘Double Tap’ due to its discreet sequential interactions (e.g. P9 *“whenever there’s two of anything I immediately double tap, that’s a no brainer for me”*).

Given that, in study 1 where no ‘correct’ *input* was provided, ‘Dwell’ largely achieved a higher Selection% than ‘Force’, these responses suggest that, where less common inputs may resemble more established ones, special attention is required to ensure that the intended *input* is signified first and foremost.

8.3 Impact of Physical Interaction and Personal Differences

The laboratory study revealed that correct perception of available *inputs* is not always sufficient to achieve successful interactions since how participants interacted with the phone also impacted the interaction. All participants were handed the phone, but while some kept it in their hands, others put it down on the table in front of them. For the participants who held it, some interacted with the thumb of the hand holding it while others interacted with the index finger of the other hand. Participant 20 noted *“Hold on, my hand was covering that”* and then put the phone down on the table, while participant 17 commented, that *“it depends on how you hold your phone, since I use my thumb I block a lot of the phone, but if I would hold the phone in one hand and tap with a finger on the other, maybe I would block less of the screen”*. In addition, participant 24 had long acrylic nails which changed their way of interacting. They were the sole participant that interacted with the phone on the table moving their hand horizontally with their nails pointing towards the left edge of the phone. Subsequently, there were multiple DESIGNS that were occluded by their nails. These observations are in line with, and a possible explanation for, the finding in study 2, that signifiers on the DESIGN generally perform worse than those with some distance (*vicinity, away*). If the user physically covers the signifier, it will be harder to perceive and therefore may not improve the perception of available *inputs*.

8.4 The Impact of Familiarity

Although we found no impact of previous familiarity with *inputs* on their subsequent *Selection%* in study 1 or *Attempt%* in study 2, this may be due to the way the studies are setup. In asking participants about their familiarity with *inputs* ahead of their encounter with the *DESIGNS*, we remove the possibility of their complete lack of familiarity. In study 1 participants are additionally reminded of the existence of different *inputs* through their representation in the available checkboxes. The decision to preface the *DESIGNS* with questions regarding familiarity with different *inputs*, was made to increase the likelihood of meaningful engagement with the *DESIGNS*. Nonetheless, it should be noted that this order may impact the observable correlation between familiarity and *input* selection. Therefore, it can not be ruled out, that familiarity with *inputs* may have an impact on their perception outside of a test environment.

8.5 Future Work

In an effort to explore the root of signifier performance, these studies were limited to in-place touch interactions. However, many other inputs are available on touchscreens and rarely signified (e.g. 'Swipe' [40]). Future work should explore how these inputs may best be communicated to users.

Additionally, while our results suggest that signifiers can successfully increase the perception of *inputs* and this additional guidance reduces mental effort, these studies did not consider the outcome of these *inputs* when utilised. In real world applications it may be pertinent to not only signify that an interaction is possible, but also what may be its outcome. The combination of meaningful feedforward and interaction signifiers presents a unique challenge of communicating both the intended outcome and the means to achieve it. Future work should investigate how to signify the outcome of an action without taking precedence over signifying the availability of the input method itself. Additionally, the timing and duration of temporary signifiers and its potential impact should be explored.

Furthermore, while the studies investigated which *DESIGNS* conveyed which *inputs* to the participants, they did so within a context that introduced the *inputs* to them and increased the participants consideration of them. Therefore, to confirm whether participants would *discover* the interaction method in question thanks to the signifier in real-world implementations, a discoverability study [12, 15] needs to be performed which does not include guidance on which interaction methods may be available. Such an investigation could further examine whether the lack of correlation between familiarity and interaction selection may be attributed to the fact that participants were presented with all interaction methods beforehand or had predefined answer options that did not require them to recall any interactions. Furthermore, since early pilot studies revealed that context is a major signifier of buttons, we presented the same context for all buttons. It can be assumed that most signifier designs could be adapted to different contexts, or possibly even widgets, but this is not verified within the studies performed. Future research should examine context as a signifier and how it is linked and interacts with visualisation.

9 CONCLUSION

In this work, we investigated visual signifier characteristics and their impact on perceived affordances. We propose a design space that not only acts as a framework to create and compare designs but also provides an opportunity to reflect on the qualities considered within it. In the design space, we examine explicitness, temporality and proximity. Through these characteristics, the design space allows consideration of which metaphors and design patterns can and should be replicated.

Evaluating the perception of signifiers in a Graphical User Interfaces, and consequently the affordances user might perceive, remains uncommon and no clear protocols exist in that respect.

Motivated by the success of online studies deployed to evaluate the perception of affordances in product design [21], we decided to first deploy an online survey. Subsequently, we conducted an interactive in-person user study, to further examine the perception of input availability and the mental effort required to deduce available inputs. Overall both studies confirmed that added visual signifiers increase the perception of *input* availability. This suggests that, where interactivity extends past the norm, added visual signifiers can be a helpful tool to communicate additional interactions to users.

In conclusion, visual signifiers have the capacity to communicate inputs to users where they would otherwise not perceive them. As such, they are a valuable resource in conveying the diverse functionality of modern devices that overload controls in small interfaces. Accordingly, signifiers should be explored further.

ACKNOWLEDGMENTS

We thank Lucas Ple for his help on the implementation of the experimental platform and Shaan Chopra for her first exploration of the design space. This work was supported by the Agence Nationale de la Recherche projects Discovery (ANR-19-CE33-0006) and MIC (ANR-22-CE33-0017).

REFERENCES

- [1] Sherman R Alpert. 1991. Self-describing animated icons for human-computer interaction: a research note. *Behaviour & Information Technology* 10, 2 (1991), 149–152.
- [2] Axel Antoine, Sylvain Malacria, and G ery Casiez. 2017. ForceEdge: Controlling Autoscroll on Both Desktop and Mobile Computers Using the Force. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 3281–3292. <https://doi.org/10.1145/3025453.3025605>
- [3] Caroline Appert, Olivier Chapuis, and Emmanuel Pietriga. 2012. Dwell-and-Spring: Undo for Direct Manipulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Austin, Texas, USA) (CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1957–1966. <https://doi.org/10.1145/2207676.2208339>
- [4] Liv Arleth, Emilie Lind Damkj er, and Hendrik Knoche. 2020. “But Wait, There’s More!” a Deeper Look into Temporally Placing Touch Gesture Signifiers. In *Interactivity, Game Creation, Design, Learning, and Innovation: 8th EAI International Conference, ArtsIT 2019, and 4th EAI International Conference, DLI 2019, Aalborg, Denmark, November 6–8, 2019, Proceedings 8*. Springer, 290–308. https://doi.org/10.1007/978-3-030-53294-9_20
- [5] Jeff Avery, Sylvain Malacria, Mathieu Nancel, G ery Casiez, and Edward Lank. 2018. Introducing Transient Gestures to Improve Pan and Zoom on Touch Surfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–8. <https://doi.org/10.1145/3173574.3173599>
- [6] Alethea L Blackler, Rafael Gomez, Vesna Popovic, and M Helen Thompson. 2016. Life is too short to RTFM: how users relate to documentation and excess features in consumer products. *Interacting with Computers* 28, 1 (2016), 27–46.
- [7] David Bonnet, Caroline Appert, and Michel Beaudouin-Lafon. 2013. Extending the Vocabulary of Touch Events with ThumbRock. In *Proceedings of Graphics Interface 2013 (GI '13)*. CHCCS, Regina (Saskatchewan), Canada. <https://hal.inria.fr/hal-00839414>
- [8] Josh Clark. 2015. *Designing for touch*. A Book Apart New York.
- [9] Emilie Lind Damkj er, Liv Arleth, and Hendrik Knoche. 2019. “I Didn’t Know, You Could Do That”-Affordance Signifiers for Touch Gestures on Mobile Devices. In *Interactivity, Game Creation, Design, Learning, and Innovation: 7th EAI International Conference, ArtsIT 2018, and 3rd EAI International Conference, DLI 2018, ICTCC 2018, Braga, Portugal, October 24–26, 2018, Proceedings 7*. Springer, 206–212.
- [10] William Delamare, Thomas Janssoone, C eline Coutrix, and Laurence Nigay. 2016. Designing 3D Gesture Guidance: Visual Feedback and Feedforward Design Options. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (Bari, Italy) (AVI '16)*. Association for Computing Machinery, New York, NY, USA, 152–159. <https://doi.org/10.1145/2909132.2909260>
- [11] Anna Dickinson, Roos Eisma, and Peter Gregor. 2011. The barriers that older novices encounter to computer use. *Universal Access in the Information Society* 10, 3 (2011), 261–266.
- [12] Katherine Kennedy, Angad Srivastava, Sylvain Malacria, and Simon T. Perrault. 2022. Towards a Unified and Efficient Command Selection Mechanism for Touch-Based Devices Using Soft Keyboard Hotkeys. *ACM Trans. Comput.-Hum. Interact.* 29, 1, Article 4 (jan 2022), 39 pages. <https://doi.org/10.1145/3476510>
- [13] James J Gibson. 1979. *The ecological approach to visual perception*. Houghton Mifflin, Boston.

- [14] Vindu Goel. 2017. Apple Piles On the Features and Users Say, 'Enough!'. Retrieved December 9th, 2021 from <https://www.nytimes.com/2017/06/02/technology/apple-iphone-developer-conference.html>.
- [15] Alix Goguey, Sylvain Malacria, and Carl Gutwin. 2018. Improving Discoverability and Expert Performance in Force-Sensitive Text Selection for Touch Devices with Mode Gauges. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3173574.3174051>
- [16] Frank Halasz and Thomas P. Moran. 1982. Analogy Considered Harmful. In *Proceedings of the 1982 Conference on Human Factors in Computing Systems* (Gaithersburg, Maryland, USA) (CHI '82). Association for Computing Machinery, New York, NY, USA, 383–386. <https://doi.org/10.1145/800049.801816>
- [17] 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* (Santa Barbara, California, USA) (UIST '11). Association for Computing Machinery, New York, NY, USA, 627–636. <https://doi.org/10.1145/2047196.2047279>
- [18] Seongkook Heo, Jiseong Gu, and Geehyuk Lee. 2014. Expanding Touch Input Vocabulary by Using Consecutive Distant Taps. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) (CHI '14). Association for Computing Machinery, New York, NY, USA, 2597–2606. <https://doi.org/10.1145/2556288.2557234>
- [19] Seongkook Heo and Geehyuk Lee. 2011. Forcetap: Extending the Input Vocabulary of Mobile Touch Screens by Adding Tap Gestures. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (Stockholm, Sweden) (MobileHCI '11). Association for Computing Machinery, New York, NY, USA, 113–122. <https://doi.org/10.1145/2037373.2037393>
- [20] Jamie Holst. 2013. Users Continue to Double-Click Online. <https://baymard.com/blog/users-double-click-online>
- [21] Shih-Wen Hsiao, Chiao-Fei Hsu, and Yin-Ting Lee. 2012. An online affordance evaluation model for product design. *Design Studies* 33, 2 (2012), 126–159. <https://doi.org/10.1016/j.destud.2011.06.003>
- [22] Apple Inc. 2023. Buttons. <https://developer.apple.com/design/human-interface-guidelines/components/menus-and-actions/buttons/>
- [23] Material IO. 2023. Material Design Buttons. <https://m2.material.io/components/buttons>
- [24] Steve Jobs. 2007. Macworld San Francisco 2007 - Steve Jobs introduced the first iPhone.
- [25] Patrick W Jordan, Stephen W Draper, Kirsteen K MacFarlane, and Shirley-Anne McNulty. 1991. Guessability, learnability, and experienced user performance. In *People and Computers VI - Proceedings of the HCI '91 Conference*. Cambridge University Press.
- [26] Alita Joyce. 2020. Mobile-App Onboarding: An Analysis of Components and Techniques. Retrieved January 27th, 2023 from <https://www.nngroup.com/articles/mobile-app-onboarding/>.
- [27] Jacob Kastrenakes. 2018. Apple could remove 3D Touch from future iPhones, analyst says. Retrieved January 21st, 2023 from <https://www.theverge.com/circuitbreaker/2018/8/27/17787938/apple-new-iphone-3d-touch-pencil-support-missing>.
- [28] Sangyeon Kim and Sangwon Lee. 2020. Smash the dichotomy of Skeuomorphism and flat design: Designing an affordable interface to correspond with the human perceptuomotor process. *International Journal of Human-Computer Studies* 141 (2020), 102435.
- [29] Sangyeon Kim and Sangwon Lee. 2020. Touch Digitality: Affordance Effects of Visual Properties on Gesture Selection. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI EA '20). Association for Computing Machinery, New York, NY, USA, 1–8. <https://doi.org/10.1145/3334480.3382914>
- [30] Jon A Krosnick and Duane F Alwin. 1987. An evaluation of a cognitive theory of response-order effects in survey measurement. *Public opinion quarterly* 51, 2 (1987), 201–219.
- [31] Vincent Lambert, Adrien Chaffangeon Caillet, Alix Goguey, Sylvain Malacria, and Laurence Nigay. 2023. Studying the Visual Representation of Microgestures. *Proc. ACM Hum.-Comput. Interact.* 7, MHCI, Article 225 (2023). <https://doi.org/10.1145/3604272>
- [32] Sven Mayer, Lars Lischke, Adrian Lankswiert, Huy Viet Le, and Niels Henze. 2018. How to communicate new input techniques. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction - NordiCHI '18*. ACM Press, Oslo, Norway, 460–472. <https://doi.org/10.1145/3240167.3240176>
- [33] Joanna McGrenere and Wayne Ho. 2000. Affordances: Clarifying and evolving a concept. In *Graphics interface*, Vol. 2000. 179–186.
- [34] Donald A. Norman. 2008. THE WAY I SEE IT Signifiers, Not Affordances. *Interactions* 15, 6 (nov 2008), 18–19. <https://doi.org/10.1145/1409040.1409044>
- [35] Donald A. Norman. 2013. *The design of everyday things: Revised and expanded edition*. Basic books.
- [36] Donald A. Norman. 2018. Affordances and design. <https://ns.inria.fr/discovery/refs/norman2018affordances.pdf> retrieved January 27th, 2022 from https://jnd.org/affordances_and_design/.

- [37] Martin T Orne. 2009. Demand characteristics and the concept of quasi-controls. *Artifacts in behavioral research: Robert Rosenthal and Ralph L. Rosnow's classic books* 110 (2009), 110–137.
- [38] David Oswald. 2018. Affordances and metaphors revisited: testing flat vs. skeuomorph design with digital natives and digital immigrants. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference 32*. BCS Learning & Development Ltd., Swindon, GBR, 1–11. <https://doi.org/10.14236/ewic/HCI2018.57>
- [39] Nicole Ke Chen Pong. 2020. *Interacting with signifier-less designs—the case of swhidgets*. Ph.D. Dissertation. Université de Lille.
- [40] Nicole Ke Chen Pong and Sylvain Malacria. 2019. Awareness, Usage and Discovery of Swipe-Revealed Hidden Widgets in IOS. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces* (Daejeon, Republic of Korea) (ISS '19). Association for Computing Machinery, New York, NY, USA, 193–204. <https://doi.org/10.1145/3343055.3359713>
- [41] Jef Raskin. 2000. *The humane interface: new directions for designing interactive systems*. Addison-Wesley Professional.
- [42] Christian Rendl, Patrick Greindl, Kathrin Probst, Martin Behrens, and Michael Haller. 2014. Presstures: Exploring Pressure-Sensitive Multi-Touch Gestures on Trackpads. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) (CHI '14). Association for Computing Machinery, New York, NY, USA, 431–434. <https://doi.org/10.1145/2556288.2557146>
- [43] Anne Roudaut, Eric Lecolinet, and Yves Guiard. 2009. MicroRolls: Expanding Touch-Screen Input Vocabulary by Distinguishing Rolls vs. Slides of the Thumb. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA). Association for Computing Machinery, New York, NY, USA, 927–936. <https://doi.org/10.1145/1518701.1518843>
- [44] Mindy Seto, Stacey Scott, and Mark Hancock. 2012. Investigating Menu Discoverability on a Digital Tabletop in a Public Setting. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces* (Cambridge, Massachusetts, USA) (ITS '12). Association for Computing Machinery, New York, NY, USA, 71–80. <https://doi.org/10.1145/2396636.2396647>
- [45] Konstantinos Spiliotopoulos, Maria Rigou, and Spiros Sirmakessis. 2018. A comparative study of skeuomorphic and flat design from a UX perspective. *Multimodal Technologies and Interaction* 2, 2 (2018), 31.
- [46] Jacob O. Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A. Myers. 2005. Maximizing the Guessability of Symbolic Input. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (Portland, OR, USA) (CHI EA '05). Association for Computing Machinery, New York, NY, USA, 1869–1872. <https://doi.org/10.1145/1056808.1057043>
- [47] Koji Yatani, Kurt Partridge, Marshall Bern, and Mark W. Newman. 2008. Escape: A Target Selection Technique Using Visually-Cued Gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) (CHI '08). Association for Computing Machinery, New York, NY, USA, 285–294. <https://doi.org/10.1145/1357054.1357104>
- [48] J. Zimmerman. 2005. Video sketches: exploring pervasive computing interaction designs. *IEEE Pervasive Computing* 4, 4 (2005), 91–94. <https://doi.org/10.1109/MPRV.2005.91>

A FINITE STATE MACHINE

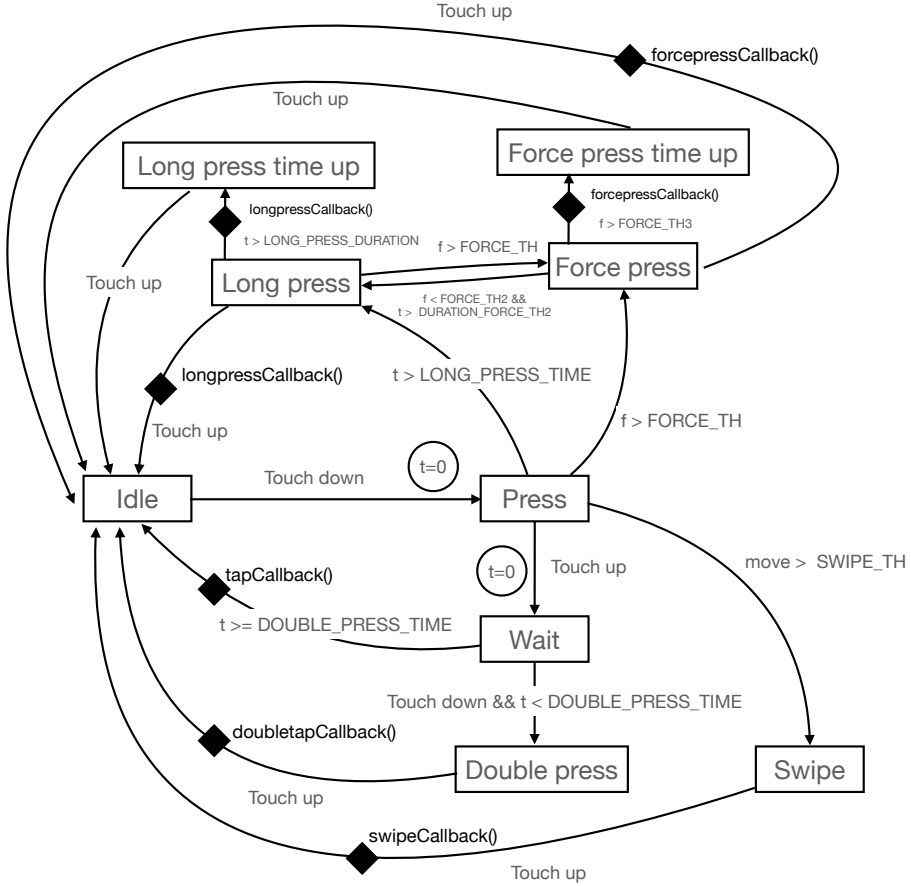


Fig. 10. Finite State Machine used to classify input states in real time during the laboratory experiment. We used the thresholds 0.3s for DOUBLE_PRESS_TIME and 0.75s for LONG_PRESS_DURATION, as defined by default on iOS; 10 pixels for SWIPE_TH; 0.9 for FORCE_TH, 0.7 for FORCE_TH2 and 0.1 s for DURATION_FORCE_TH2. The thresholds for force have arbitrary units but range from 0 to 1 as provided by Pressure.js 2.2.0 and were set using informal tests. Each design is associated with a callback that is considered as correct. Any call to another callback results as an error.

B TOUCH INTERACTION CATEGORISATION

<i>Touch Gesture</i>	<i>Contact Points</i>	<i>Position</i>	<i>Speed</i>	<i>Motion Pattern</i>	<i>Duration</i>	<i>Frequency</i>	<i>Pressure</i>
Tap	1	static	-	-	short	1	regular
Double Tap	1	static	-	-	short	2	regular
Long Press	1	static	-	-	extended	1	regular
Strong Tap	1	static	-	-	short	1	high
Scroll	1	dynamic	medium	Move from contact point position in straight line to other point	varied	1	regular
Swipe / Flick	1	dynamic	high	Move from contact point position in straight line to other point	short	1	regular
Pinch	2	dynamic	varied	Move the 2 contact points straight towards one another	varied	1	regular
Spread	2	dynamic	varied	Move the 2 contact points straight away from one another	varied	1	regular
Rotate	2	dynamic	varied	Move the 2 contact points in clockwise or counter-clockwise curve	varied	1	regular
Stroke Gesture	1	dynamic	varied	Move in predefined pattern	varied	1	regular

C TABLE VISUALISING STUDY 1 RESPONSES

Table showing the *Selection%* for each DESIGN and *input*. Cells represented with a green background represent inputs that were assumed to be signified for the design. Others appear with a blue background. The intensity of the background colour is proportional to *Selection%*.

	Tap	Double Tap	Dwell	Force Press	Scroll	Swipe	Pinch	Rotate	Pattern
01 Basic	94	47	56	31	3	6			16
02 Text Countdown	81	81	44	38					
03 Text Timer	84	38	94	50		3			6
04 Text Counters	91	81	91	84					3
05 Text Small	94	25	88	84					
06 Text Temporary Attached	94	78	88	75			3		
07 Text Temporary Removed	97	81	84	78		6	3		3
08 Text Permanent Attached	94	84	94	75					
09 Text Permanent Removed	91	75	91	81		3			
10 Color Change (Steps)	94	56	47	38	3	6			3
11 Circumferential Steps	88	47	81	50	3	9		16	22
12 Scale (Check) Distant Temp	88	56	66	53	3	6			9
13 Checkmarks (Radio)	91	88	41	25			3		3
14 Checkmarks (Tick)	94	91	34	25	6				
15 Circles Around	88	75	53	53		3	3		
16 Scale (Check) Distant	97	59	69	44	3	12			6
17 Folding	72	28	84	31	12	38		3	3
18 Leaf Through	78	34	53	31	9	44			9
19 Scale Vertical Internal	78	25	94	78	3				
20 Color Change (Gradual)	97	25	66	53		3			6
21 Scale Circular Distant	81	22	88	62	3	3	3	9	3
22 Scale Radial	81	19	94	62	3	9	3		9
23 Scale Distant Temporary	72	28	94	66		6			3
24 Arrow	84	28	53	28	3	41	3		31
25 Scale Circular	72	25	97	66				3	
26 Scale Horizontal	75	25	91	72		3			
27 Scale Distant Permanent	84	22	94	69		3			
28 Sinking	81	28	59	62	12	12	6	3	12
29 Filling Balloon	75	28	84	59	9	16			9
30 Filling Water	91	25	91	62	9	6			6
31 Hourglass	91	41	72	28		3		6	
32 Scale Vertical	69	28	94	72	3	3			3
33 Squeeze	69	25	84	84	3	6			9
34 Ripples	91	41	53	66	6	9	3		9
35 Material (Concrete)	91	56	59	72	6				6
36 Highstrike	75	28	72	84	9	3			3

D TABLE VISUALISING STUDY 2 RESPONSES

Table showing the *Attempt%* for each DESIGN and *input*. Cells represented with a green background represent inputs that correct for the design. Others appear with a red background. The intensity of the background colour is proportional to *Attempt%*.

	Tap	Double Tap	Dwell	Force	Swipe		Tap	Double Tap	Dwell	Force	Swipe
Design01A	75	88	42	29	38	Design21	33	17	79	96	21
Design01B	42	21	79	25	21	Design22	46	21	83	38	4
Design01C	79	58	50	79	29	Design23	54	21	83	21	12
Design02	25	100	0	0	12	Design24	42	17	83	38	79
Design03	54	33	88	25	21	Design25	25	50	88	50	17
Design04A	46	100	4	4	4	Design26	8	0	58	100	8
Design04B	29	21	79	29	12	Design27	50	29	92	42	8
Design04C	8	8	8	100	4	Design28	71	38	50	79	46
Design05A	67	100	21	17	4	Design29	54	38	75	38	17
Design05B	46	4	100	8	17	Design30	54	21	79	38	17
Design05C	75	4	29	100	8	Design31	17	4	96	4	8
Design06A	42	100	62	0	4	Design32	4	12	71	96	8
Design06B	62	8	92	21	12	Design33	50	50	58	75	33
Design06C	54	17	54	96	0	Design34	58	21	71	83	29
Design07A	46	100	42	8	4	Design35	50	71	38	83	17
Design07B	46	4	100	8	4	Design36	17	12	54	96	33
Design07C	33	8	62	96	4						
Design08A	4	100	0	0	0						
Design08B	4	8	92	12	8						
Design08C	8	4	17	100	0						
Design09A	12	100	17	0	4						
Design09B	17	4	100	4	8						
Design09C	4	0	42	100	4						
Design10	67	92	54	46	21						
Design11	54	92	75	21	21						
Design12	71	92	62	17	8						
Design13	17	96	21	0	8						
Design14	8	100	8	0	4						
Design15	17	96	62	42	8						
Design16	42	88	54	21	25						
Design17	67	58	50	62	67						
Design18	58	79	54	38	38						
Design19	62	21	71	46	12						
Design20	67	46	58	92	29						

E FIGURE VISUALISING THE INPUT ATTEMPTS IN THE INTERACTIVE STUDY

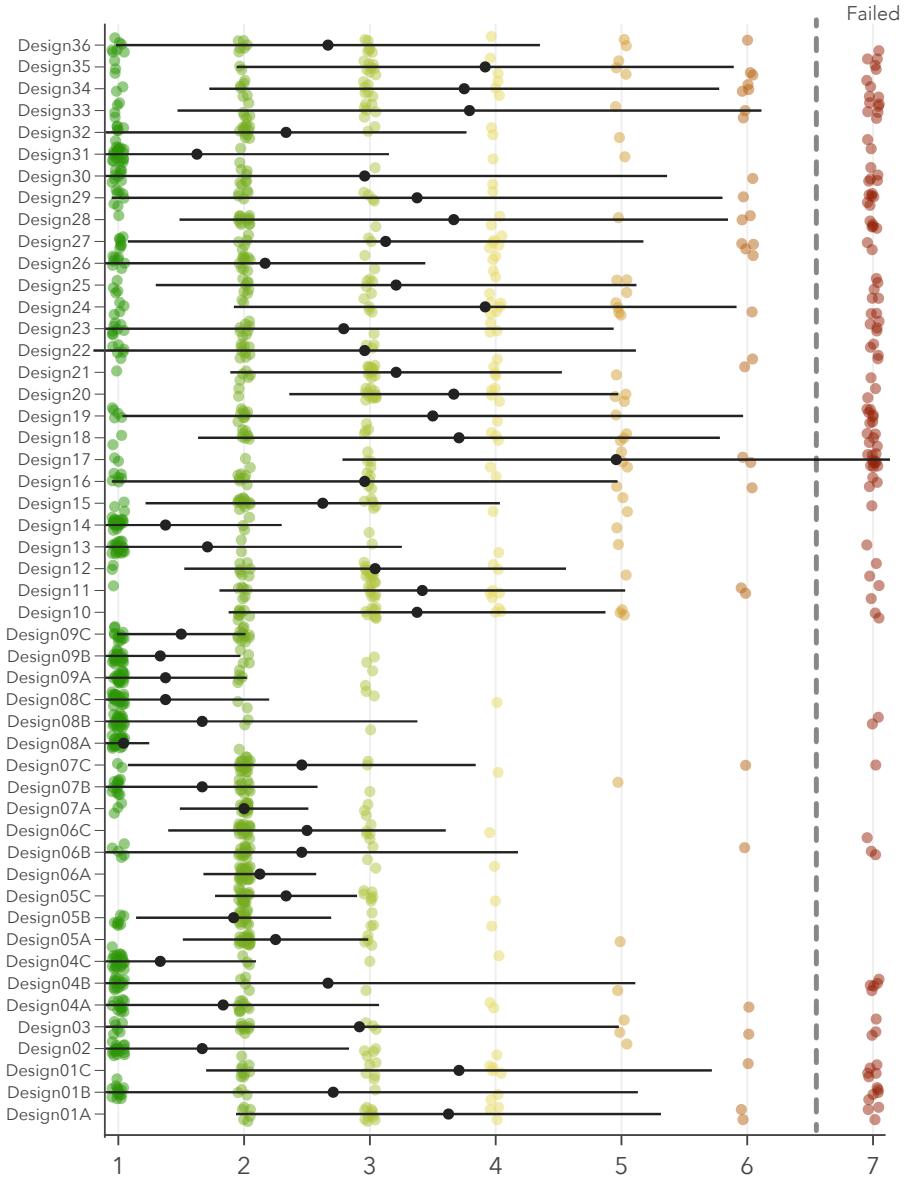


Fig. 11. Trials required to achieve the correct interaction.

Received January 2023; revised May 2023; accepted June 2023