Self-Previewing Gestures and the Gesture-and-Effect Model: Experimentation with Responsive Visual Feedback for New and Unlearned Interactions

Jacques Chueke	George Buchanan	Stephanie Wilson	Luis Anunciação
Centre for HCI Design	Centre for HCI Design	Centre for HCI Design	PUC-Rio
jacques.chueke.1 @ city.ac.uk	*george.buchanan.1 @ city.ac.uk*	*stephanie.wilson.1 @ city.ac.uk*	*luisfca @uerj.br*

City, University of London - Centre for HCI Design, Northampton Square, London, UK, EC1V 0HB
PUC-Rio - Dept. de Artes & Design, Rua Marquês de São Vicente, 225 Gávea, 22451-900 – Rio de Janeiro – RJ – Brasil

Multi-touch gestures embedded in touch-based interfaces and devices present challenges. It can be difficult for users to discover different gestures and understand their effects. The study reported in this paper hypothesises that presenting automatic visual prompts, termed self-previewing gestures (SPGs) in this research, that depict touch and preview gesture execution will mitigate the problems that users encounter with unfamiliar gestural interfaces. A within-subjects experiment (n=45) is reported in which an iPad application with two alternative gestural designs, and five alternative user interface versions (one industry, two research baselines, and two SPGs) was created with the purpose of making the available gestures evident. A rating system that adapts Norman's Theory of Action to touch-based interactions by making use of known principles within interaction design (perceptible affordances, feedforward and feedback) is proposed. The system was used to assess participants' perceptions of and interactions with the SPGs, and the results revealed positive and negative aspects of designs and UI versions.

Perceptible Affordances, Feedforward, Feedback, Touch-based interaction, Gestures, Empirical Study, NUI.

1. INTRODUCTION

Previous work (e.g. Norman, 2012; Vermeulen et al., 2013; Wigdor & Wixton, 2011) has observed issues with gestural interfaces in which users are unaware of and/or fail to discover hidden gestures and UI tools, resulting in both inabilities to perform gestures and errors in executing them. Norman and Nielsen (2010) commented: '…How is anyone to know, first, that this magical gesture exists, and second, in which settings it operates?' Norman (2012) reiterates: 'One of the powers of modern computers is discoverability, you can explore, but with gesture systems it's a pain. It's amazing how many things people don't know about the computers they use and there's no way to find out.'

These issues challenge the notions of perceptible affordances, feedforward and feedback within the domain of Human-Computer Interaction (Baerentsen & Trettvik, 2002; Gaver, 1991; Kaptelinin & Nardi 2012; McGrenere, 2000; Norman, 1988, 1999; St. Amant, 1999; Hartson, 2003; Turner, 2005), which scrutinise how UI objects (e.g. controls, menus, toolbars, modes, domain objects, etc.) should be represented and respond to user action or inaction. Researchers are re-thinking these theories (Djajadiningrat et al., 2004; Vermeulen et al., 2013; Wigdor & Wixton, 2011) in order to adapt graphical user interfaces (GUIs) to systems that afford novel input methods and, as a result, facilitate users' adoption of them. The resulting systems are generally called natural user interfaces (NUIs) (Norman, 2010; Sorensen, 2010; Wigdor, 2010). In many cases, NUIs lack traditional WIMP-GUI (windows, icons, menus, pointer) controls and embed post-WIMP (Beaudouin-Lafon, 2000; Dam, 1997) user interfaces adapted to touch and other forms of 'natural' input.

In addition, the relevant literature assessing the conceptual implications of novel forms of UI to gestural interactions is, at best, sparse. To bridge this gap, this paper reports a study exploring participants' assessments of automatic visual prompts that depict touch and gestures on the screen of a tablet. To support the study and enable a detailed analysis of the data, we developed a 'gesture-and-effect' model of touch-based

interactions and a corresponding rating system. The rating system allowed us to assess users' success or failure to identify potential gestures and correctly predict their outcomes.

2. BACKGROUND AND RELATED WORK

Existing studies have provided some initial evidence for the potential of feedforward[1] (Djajadiningrat et al., 2002, 2004; Freitag, 2012; Vermeulen et al., 2013) to support the learning of new touch gestures (e.g. OctoPocus, see Bau & Mackay, 2008) and more recent self-revealing gestures (Hofmeester, 2012; Wigdor & Wixton, 2011, pp. 145–156). The current prevalence of self-revealing (e.g. Ripples, see Wigdor et al., 2009) annotations, animations, gesture completion paths (ShadowGuides, see Freeman et al., 2009; SimpleFlow, see Bennett et al., 2011) and automatic tips on hand-held devices (FastTap, see Gutwin et al., 2014) during interactions suggests that designers are aware of this problem, and endeavour to support their users in the learning process.

The common factor in most findings from extant research is lower error rates (around 20%) in execution (e.g. Paperlens, see Spindler et al., 2009; Ripples, see Wigdor et al., 2010; ShadowGuides, Freeman et al., 2009; SimpleFlow, see Bennett et al., 2011) for gestural interfaces that implement feedforward prompts in response to participants starting interactions, compared to higher rates (up to 50%) for basic interfaces in which participants were left to discover gestures with no visual aid. As observed in these studies, users are shown the continuation (see terminology used in Wu, 2005), which is the effect of gestural interactions, but not how to start them. This approach is extended here by examining the benefits of feedforward before touch occurs, following Vermeulen's (2013, p. 1938) reframing of the feedforward technique. We hypothesised that moving this feature to earlier in the interaction sequence could improve the discovery of unfamiliar gestures and result in fewer errors in execution. We have termed this approach "Self-Previewing Gestures" (SPG).

3. DESIGNING THE INTERACTIONS

In this study, two alternative visual designs were deployed with SPG incorporated into the designs. These designs present contrasting approaches to communicating with users. The study employed an unfamiliar gesture vocabulary in order to challenge both expert (e.g. familiar with iOS gestures) and novice users to touch devices. The most important

---

[1] Feedforward (Vermeulen et al., 2013) is an interaction technique that 'provides a "preview" of how to proceed during interaction and an indication of the actual outcome.'

visual aid on which participants relied to identify available gestures was the SPG, rather than previous knowledge. The gestures did not match the expected effects observed in current tablets. The two designs are: **1. 'Iconic':** The iconic version displays a pictorial representation of a hand touching the screen with the appropriate number of fingers to initiate registration of the gesture. It also used text labels to indicate the action required (i.e. 'open') or the UI object to be triggered (i.e. 'menu'). **2. 'Symbolic':** The second design 'Symbolic' uses simple geometric forms to depict touch points over the screen. The touch points were designed in an uneven fashion to simulate the human touch over the screen. Noticeably, this design style does not use textual support. It is more *abstract* and provides a less direct visual metaphor. Both designs used arrows to demonstrate direction.

Three interactions were selected as being unfamiliar in touch devices and, therefore, supposedly more challenging for users. For each of the three interactions, an image depicting the touch points required (the SPG), the movement and the type of action (such as a tap or pinch) was presented onscreen, together with an animation of the system's response to the gesture. The user was then asked to replicate the gesture. The system response was kept as a constant factor, meaning that the feedforward created was identical in both designs. The preview of the 'effect' of the gesture (or system response) comprised a 'ghost' (a term used in this study to indicate a translucent clone of an object) that was animated along with the gestural affordance to preview the correct position for activation. Insights for the interaction techniques were drawn from Wigdor et al.'s (2009) Ripples technique, Wigdor and Wixton's (2011) self-revealing gestures 'chrome' layer for MS Surface and Hofmmester's (2012) prototype work for Windows 8 touch. The three interactions are shown in Figure 1, Figure 2 and Figure 3. The gesture depictions are shown in a magnified view at the bottom of each picture, and the effect of each gesture is shown in a sequence of frames. The screen size is 2048x1536px at 264 ppi (Apple iPad models, 2013), and the interactions are displayed over a fictitious booklet application.

1. Open application: In the prototype screen, a multi-touch gesture opens an application, which is traditionally opened by a single tap. The 'open application' interaction proceeds as follows: The gestural affordance appears over the button (Figure 1-a). The animation demonstrates the button moving down along with the gesture (b) before returning to its place. The user is expected to move the button in the appropriate fashion to open the application (c).

2. Pull hidden menu: This gesture requires the participant to swipe horizontally from an 'invisible' activation area to reveal a hidden menu. Neither the optimum touch range nor the UI component is visible. The interaction proceeds as follows:

The gestural affordance appears on the left hand side of the screen (shown in zoom in Figure 2-a). The animation demonstrates the menu moving sideways along with the gesture (b) and then returning to its place. The user is expected to swipe
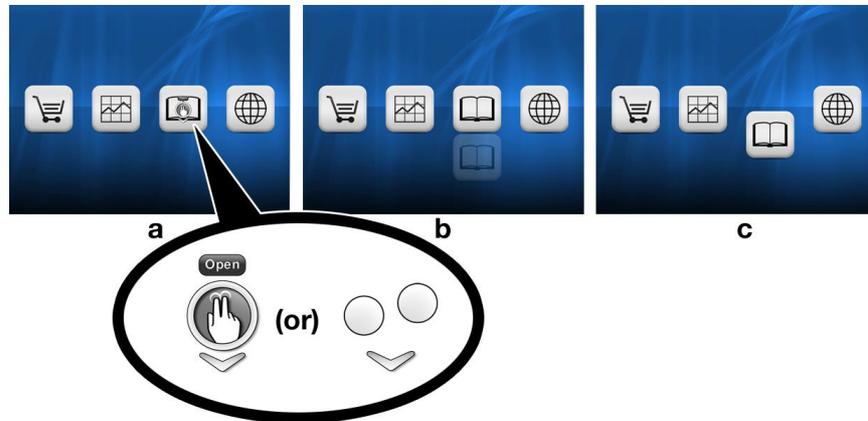
Figure 1: *Interaction 1 'Open application'. The SPG appears over the button (a). The animation demonstrates the button moving down along with the gesture (b) before returning to its place. The user is expected to move the button in the appropriate fashion to open the application (c). Either Design 1 'Iconic' or Design '2' Symbolic are shown in a randomised sequence.*
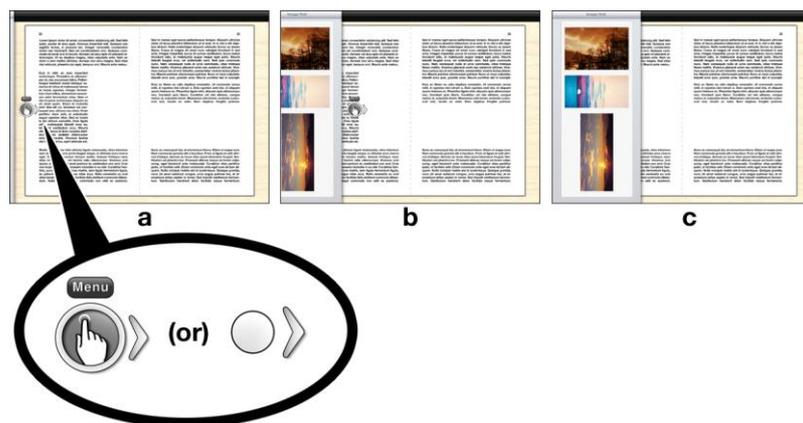
Figure 2: *Interaction 2 'Pull hidden menu'. The SPG appears on the left hand side of the screen (a). The animation demonstrates the menu moving sideways along with the gesture (b) and then returning to its place. The user is expected to swipe horizontally from the left bezel towards the centre of the screen to reveal the menu (c). Either Design 1 'Iconic' or Design '2' Symbolic are shown in a randomised sequence.*
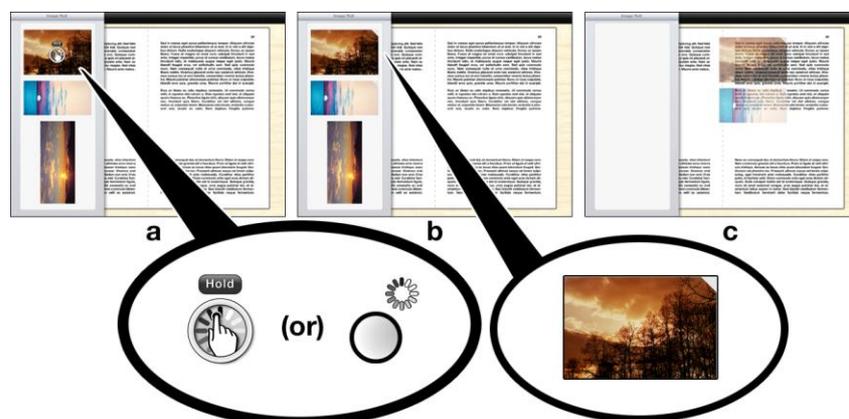
Figure 3: *Interaction 3 'Touch and hold'. The SPG appears over the top picture (a). In the animation, the top right corner demonstrates that the picture is selected (b) before fading out. The user can select as many pictures as desired and then drag and drop them over the booklet (c). Either Design 1 'Iconic' or Design '2' Symbolic are shown in a randomised sequence.*

horizontally from the left bezel towards the centre of the screen to reveal the menu (c).

3. Touch and hold a picture: Norman and Nielsen (2010), Vermeulen et al. (2013) and Norman (2014) note that this type of gesture is particularly difficult for inexperienced users of gestural interfaces. To avoid learning effects, we used the touch-and-hold gesture in a context that would be unfamiliar to most users. First, the gestural affordance appears over the top picture (shown in zoom in Figure 3-a).

In the animation, the top right corner demonstrates that the picture is selected (b) before fading out. The user can select as many pictures as desired and then drag and drop them over the booklet (c). This interaction requires the user to complete two steps. Therefore, it is a 'sequential'[2] affordance (Gaver, 1991, p. 82), as the picture 'affords' selection as a response to a touch-and-hold interaction. This affordance takes the shape of a small bent corner on the picture's top right (Figure 3-b), which remains active following a successful interaction to indicate a change of mode to 'selected' and available for dragging.

The next section describes the gesture-and-effect model used to assess the SPG.

4. THE GESTURE-AND-EFFECT (GEM) MODEL

The new forms of visual interventions that depict touch over a UI require a new framework of analysis to investigate the effectiveness of the SPG technique. To provide the theoretical foundations for such a framework, the works of Norman (1998), Wu et al. (2005) and Golod et al. (2013) were referenced to create a gesture-and-effect model (GEM) for evaluating touch interfaces. Figure 4 shows the GEM.

The initial approach when creating the model was to separate the user's planning and action into smaller steps. Norman's (1988, pp. 45–53) Theory of Action provides a generalized view of a person interacting with 'the world'. A user must form an intention and plan an action sequence in order to execute actions to fulfil his or her goal (the stage of execution). Following execution, a new phase of evaluation occurs, in which the user reassesses the environment to check whether his or her goal has been achieved. Thus, evaluation and execution form a cycle. Using a 'self-previewing' interface, which shows actions in relation to context(s), it is possible to reconsider Norman's (1988) theory beginning with the evaluation stage.

---

[2] According to Gaver (1991, p. 82), *sequential* affordances are 'sequential in time (i.e., acting on an affordance leads to information indicating new affordances).'

Wu et al. (2005) focus on the execution of gestures. The authors created the registration–continuation–termination (RCT) model to break down the execution of a gesture into 'micro' parts. In a similar fashion, Golod et al. (2013, p. 17), describe a 'gesture phrase' model, which segments the execution of a gesture into 'microinteractions'. Golod et al.'s (2013) 'gesture phrase' and Wu et al.'s (2005) RCT model were adapted into the GEM execution phase. The term 'micro-phase' was also adopted and used to describe the phases within the model.

4.1 Restore Status (Undo a gesture)

NUIs, which differ fundamentally from the predominant desktop metaphor paradigm, present a challenging context for undo actions (see Dix, 1996; Shneiderman, 2010), and mainstream touch devices (e.g. smartphones and tablets) rarely provide 'undo' buttons or options within their 'edit' menus.

However, Norman's (1988) model does not separately address the issue of an undo option. We therefore needed to add a separate 'Restoration' phase to the model, especially since the work of performing an undo is so complex in the domain of gestural interfaces. The micro-phase 'undo' or 'restore' is represented in the model with the arrowed pathways labelled 'C' and 'D'. This phase implies a new evaluation of the system status and the need for a new gesture (execution phase)—different from the initial gesture—to undo an action or restore the system to its previous state.

4.2 Tap to Preview

As can be seen in the model (bottom portion), another feature was included to represent an upcoming interaction technique in gestural interactions, which we call 'tap-to-preview'. This interaction is depicted in the GEM as an 'alternative execution' (arrowed pathways 'E' and 'F') that triggers a new evaluation of the available gesture. Wigdor and Wixton's (2011, p. 153) 'just-in-time chrome' technique and Hofmeester's (2012) 'teaching gesture' for Windows 8 touch were highly influential in creating this variant within the GEM.

The next section examines previous work using a rating system relevant to the study goals.

5. RATING SYSTEM

One measure of the difficulty users experience with an interface is the number of attempts required to successfully execute a gesture. Freeman et al. (2009) and Bragdon et al. (2009) defined *a priori* rating criteria to judge success or failure in participants' attempts to perform gestures.
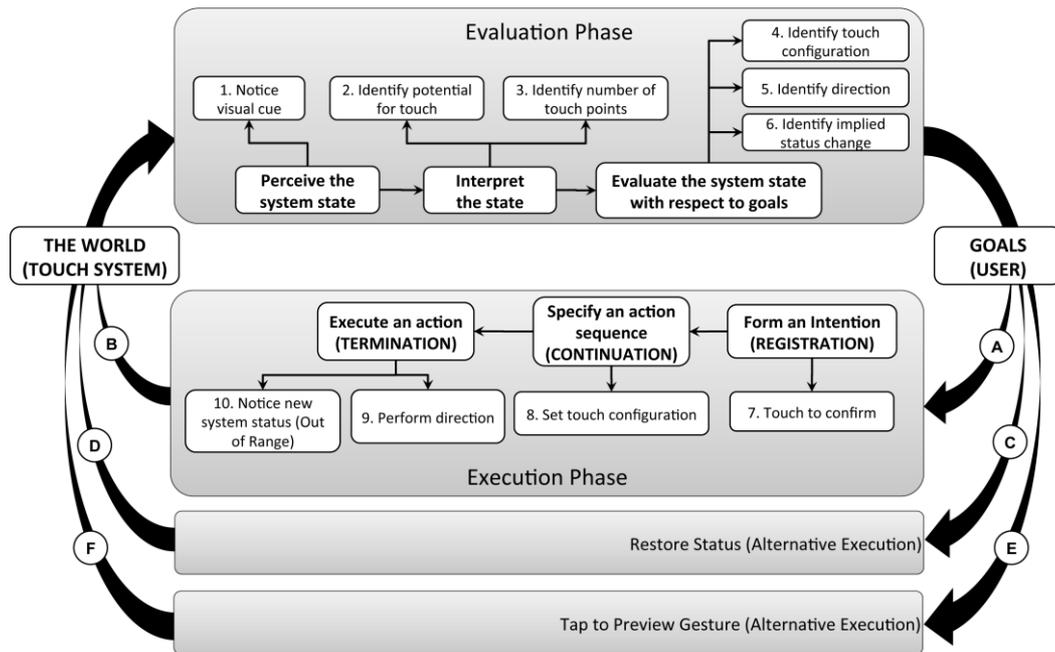
Figure 4: *The gesture-and-effect model (GEM) for touch interactions. The GEM is based on Norman's Theory of Action (1988) by considering the evaluation and execution stages, but adapting it to touch-based interactions.*

The study reported here sought to discriminate between participants' 'quick success' versus success obtained at great cost (e.g. after many attempts) in a prototype design with untaught gestures. Thus, Bragdon et al.'s (2009) criteria seemed directly relevant. Applying Bragdon et al.'s (2009) rating scheme at a more strategic level, this study counted the number of attempts to determine a nominal coding. For instance, up to six attempts could indicate success, but more than six could indicate significant effort to understand a visual prompt or execute an action.

A pilot study (n=10, 6F, 4M, age 22 to 54) was organised with acquaintances from other departments within City University (excluding colleagues from the Centre for HCI Design). It was used to validate the pre-set rating criteria. It did not intend to produce a formal coding or statistical analysis. A preliminary set of two designs was used to validate the number of attempts necessary for participants to succeed or fail in understanding and executing gestures.

No participants' verbalisations indicated a need for change the rating criteria. It was observed that most participants who fully understood the visual prompt produced an acceptable description and executed the gesture in the first three to six attempts. Participants who struggled to comprehend the prompt managed to describe and perform it in up to seven attempts, rarely more. However, some participants failed to describe or execute the gesture at all. The participants' struggles to execute the gestures showed that a 'partial' rating may provide

finer grained information. It is important to note that the partial rate of assessment, although it stemmed from a successful description, was defined in such a way as to differentiate it from full success.

Therefore, in the final rating scheme, a 'success' (1) was measured by the clarity of the user's description of the meaning of what he or she saw before a successful execution. Both the precision of the description and the number of attempts taken to arrive at the final assessment were considered. An accurate description was considered to include the number of touch points required, the motion to be performed and, at the most rudimentary level, the ability to perform a gesture within the first six attempts. A 'failure' (2) was considered to be a complete inability to describe the visual cue or execute the gesture. A 'partial success' (3) was considered to be a correct assessment from the seventh attempt onwards up to a successful execution.

The participants' evaluations and executions were assessed separately, meaning that a participant might succeed at one phase but fail at another. For instance, a user might correctly identify a gesture as requiring two touch points but fail to identify the movement he or she should follow. This would result in a 'correct' assessment for the number of fingers, but a failed physical execution of the gesture (e.g. due to swiping in the wrong direction).

6. STUDY HYPOTHESES

Hypothesis (a) stated: Depicting the touch points in the user interface will improve gestural learning and reduce user error in executing gestures. Following the approach of laboratory-based research *through* design (RTD) *for* design (Cross, 1999; Frayling, 1993; Koskinen et al., 2011; OECD Frascati Manual, 2015; Zimmerman et al., 2007), two visual designs described in section 3 were deployed.

Second, this study sough to support or reject the second study hypothesis (b), which stated: Displaying automatic visual cues before interaction will facilitate the discovery of gestures and reduce errors in execution. Its aim was to explore the SPG combinations that yield fewer error rates in participants' executions of gestures.

Third, hypothesis (c) stated: A rating system that segments users' gestural interactions into smaller phases will help to reveal issues with users' evaluations and executions of gestures. This study used the GEM rating system to evaluate prototype designs in a within-subjects experimental design and explore the specific moments within assessment and physical executions that posed the most difficulties for participants.

The next section describes the methodology used to undertake the empirical study, including the criteria for the selection and recruitment of participants, the materials utilised and the study design.

7. METHODOLOGY

Motivated by the review of previous research in designing gestural interactions, the first design consideration was how to trigger the appearance of the cue (the visual prompt) that informed the user of an available gesture. For this study, five versions of the UI were prepared and evaluated. The iconic and symbolic designs were used as visual prompts to indicate the touch gestures. The three interactions described in section '3' were used as tasks for participants to undertake. In short, these five versions included:

Version 1 (V1). 'Static gestures': The visual cue that depicts touch points on the screen only fades in and out for an appointed time. Thus, version 1 provided an industrial baseline, representing a common form of cue found in contemporary software on iOS, Android and other touch-screen operating systems. Mobile applications still rely on static (and inefficient) visual prompts that are either displayed only once upon the first run of an application or presented in step-by-step tutorials, which are often ignored by users.

Version 2 (V2). 'Animated gestures and effect': Version 2 was the SPG technique. This version showed the number of touch points and the effect of the gesture (e.g. a hidden menu) before the user touched the screen.

Version 3 (V3). 'Tap-to-preview animated gesture': Version 3 provided a research baseline using an existing and proven method for guiding users during the execution of a gesture. It displayed the registration pose (the touch points over the screen) after a user began touching the screen. This version drew from Freeman et al.'s (2009) ShadowGuides technique and the work of Hofmeester (2012). The tap-to-preview feature was included to prevent participants from missing the automatic presentation of visual cues.

Version 4 (V4). 'Tap to reveal animated UI response': Version 4 provides an additional research baseline. It draws from Bau and Mackay's (2008) Octopocus and Bennett et al.'s (2011) SimpleFlow, which lack the visual prompts necessary to start a gesture and show only the effect of the gesture in the form of a 'gesture-completion path'. The effect is shown only to users touching and holding target objects.

Version 5 (V5). 'Complete set': Version 5 was one of the representatives of SPG. It was the most complete set, combining animations that 'self-preview' visual cues for gestures and their effects. Different from V2, it also included the tap-to-preview feature, which was incorporated to determine whether this feature could further improve gesture discoverability by allowing participants to *play back* the gesture and effect in the event of missing the automatic cue.

7.1 Participants

The study recruited a total of 45 participants. Participants were recruited via leaflets placed across City University. The recruitment process sought participants from diverse backgrounds to ensure that the designs were assessed across a broad pool of users. All participants were coded according to their participant number, age, gender (e.g. P1, 37, M) to ensure anonymity.

Study participants were between 19 and 64 years old. This sample selection sought to avoid ethical issues, since participants under 18 or over 65 years old require special protection consent, as mandated by the university's ethical regulations. In more detail, 27 participants were between 19 and 33, 12 were between 34 and 48 and 6 were between 49 to 64. With respect to gender, 18 were male and 27 were female. Regarding desktop computer use, 27 participants used Windows, 14 used MacOS and 4 did not report an OS preference. In terms of smartphones, 18 owned iPhones, 17 owned Android devices, 2 owned Blackberry devices, 7 had regular cell phones and 1 did not specify a phone platform.

For tablet devices, 25 participants were iPad users, 4 used other brands and 16 did not possess tablets.

7.2 Study Design

The study took place at the Interaction Lab (Centre for HCI Design), City University of London. The test was set up with an iPad running iOS 7 attached to a metal stand for testing mobile devices. A Microsoft camera with a built-in microphone was positioned to record the screen and comments. Only the participant and the facilitator (the researcher) were present in the room during the test.

The prototype application was developed by a third party on Linux and OS/X using Xcode, and it ran on iOS 7 (which was, at the time, the current version of iOS). The application was implemented in JavaScript.

A within-subjects experimental design was used. The first set of independent variables (IVs) consisted of the five versions of the application. The second set of IVs comprised the 10 micro-phases in the GEM, including 'tap-to-preview'. Each micro-phase was rated accordingly to the *a priori* criteria (Incorrect, Correct and Partial ratings).

The first micro-phase within 'evaluation' (Notice visual cue) was removed from the analysis due to a low error rate and a limited error scale. Specifically, of the 45 participants, 42 (94%) participants detected the visual prompt when the study session started. Three did not see the cue and had to be prompted, but were still considered able to continue the study. The low error rate for this phase might have been due to a learning effect. Removing the data reduced the risk of producing spuriously significant results from small sample sizes (i.e. of errors).

7.3 Randomization set

The designs (2), interactions (3) and application versions (5) were randomised using a Latin square set, yielding a total of 30 combinations. To verify the methodology and to check if the designs or versions required any further improvement, a pilot study (similar to the set used to validate the rating system) was organised (n=4, 2F, 2M, age 26 to 43).

The participants produced no relevant comments suggesting the need for design alterations. Furthermore, most participants managed to produce adequate descriptions of the SPG within the first 10 presentations, indicating that showing all possible combinations throughout the test would be unnecessary. Therefore it was decided to display only 20 combinations per sequence and to expose each participant to each of the five versions four times in a balanced fashion.

Finally, to avoid biasing results by showing the same sequence to all participants, we organised three different randomised sequences. It was considered that 15 participants per set would provide a sufficient sample.

7.4 Elicitation Method

An 'oral structured interview' (Geiwitz et al., 1990; Hudlicka, 1997) was used to interviewing the study participants. The oral structured interview method combines situational and behavioural question types and, unlike 'contextual inquiry', can be used in a controlled laboratory environment. Furthermore, unlike long, conventional post-task interviews (e.g. those used in 'contextual inquiries'), this method uses short questions to elicit micro-responses, rather than including a time-consuming addendum at the end of the test.

Videos of the participant sessions were systematically reviewed in detail. Verbalisations were transcribed, and the performance of evaluations and executions for each micro-phase was rated and recorded in a spreadsheet for analysis. The application logs were consulted, to further assess participants' success at each step of evaluation and execution.

8. FINDINGS

The results of a generalized linear model (GLM) were assessed by employing different tests: a log-linear analysis, a chi-square test, a Mann-Whitney test and a Kruskal-Wallis test. Initially, a Shapiro-Wilk normal distribution test was conducted, and the results indicated that the H0 could be rejected ($p<0.05$), which demonstrated that the distribution of results was non-normal. The difference between the designs was then calculated using a Mann-Whitney test.

To determine any reliable significant differences and as a first step in the analysis, a global log-linear analysis was conducted for the evaluation phase (descriptive values in Table 6 and log results in Table 8 within the appendices). The analysis of the scores used all three dimensions: a) the three ratings of user performance, b) model micro-phases 2 to 6 and c) the five separate versions. The results of the global test across all factors were statistically significant: G^2=514.4, df=64, $p<0.001$.

Following the same procedure used to assess the evaluation phase, a global log-linear analysis was conducted to verify the significance between the independent and dependent variables for the execution phase (descriptive values in Table 7 and log results in Table 9 within the appendices). The analysis used three dimensions: a) the three ratings of user performance, b) model micro-phases 'T' to 10 and c) the five separate versions. The results of the global test across all factors were statistically significant: G^2=1364.26, df=64, $p<0.001$.

8.1 Addressing hypothesis 'a'

The first hypothesis states that **depicting the touch points in the user interface will improve gesture learning and reduce user error in executing gestures.** To either support or reject the null hypothesis, it was necessary to compare Version 4, which did not visually depict touch points, to all versions that did depict touch points (V1, V2, V3 and V5). The dependent variable (DV) was the number of errors (signalled by correct, partial and incorrect ratings) per version per 10 micro-phases. This section examines whether there were marked differences in micro-phase results between versions.

Version 4 is typical of established research in gestural interactions, such as gesture-completion paths, in which users have no direct support to identify available and hidden interactions. It does not depict touch points to initiate gestures, only the touch path or movement. As can be observed in Table 2 (appendices), Version 5 produced the most correct evaluations for micro-phase 4 (Identify touch configuration), with 88%, χ^2= 31.154, p<0.001 (8, N=900). The results also show that Version 5 produced the most correct executions for micro-phase 8 (Set touch configuration) (89.4%), and micro-phase 9 (Perform direction), with 98.3%. Versions 4 and 1 produced the least correct responses to micro-phases 2 (Identify potential touch) (6.7% and 6.1%, respectively), 3 (Identify touch points) (21% and 14%, respectively) and 4 (Identify touch configuration) (21% for both). Micro-phase 6 was different and did not show significant results (χ^2=3.969, df = 8, p= 0.860).

This section examines whether there are marked differences in the micro-phase results between versions. As can be observed in Table 4 (appendices), Version 4 produced the second most errors in execution for micro-phases 8 (Touch configuration) (18%) and 9 (Perform direction) (4%).

Table 1: *Expected and actual executions for versions (system status).*
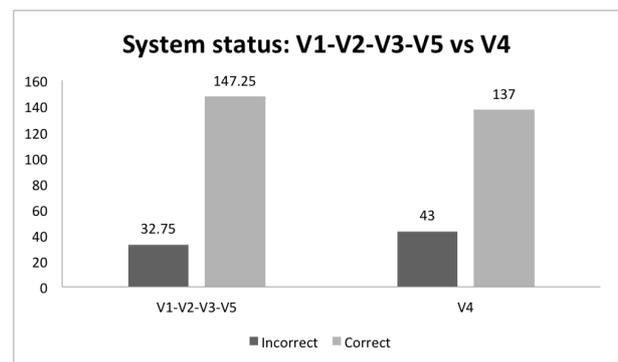
		System status	
		Incorrect	Correct
V1	**N**	43	137
	Expected	34.8	145.2
	%	23.9%	76.1%
V2	**N**	29	151
	Expected	34.8	145.2
	%	16.1%	83.9%
V3	**N**	35	145
	Expected	34.8	145.2
	%	19.4%	80.6%
V4	**N**	43	137
	Expected	34.8	145.2
	%	23.9%	76.1%
V5	**N**	24	156
	Expected	34.8	145.2
	%	13.3%	86.7%

χ^2=10.145, df = 4, p<0.05

Micro-phase 10. (System status), is treated separately, since the result was simply true or false, with no option for 'Partially correct'. This micro-phase is critical because it is the final determination of whether a participant succeeded or failed in executing a given gesture. A Kruskal-Wallis test was used to analyse versions across this micro-phase, showing a significant difference (H=2.551, df = 895, p<0.05). In order to assess the differences in system status between the versions, each version was tested for the likelihood of correct and incorrect executions. In this final analysis, the general performance for all versions across both evaluation and execution was 81%.

As can be seen in Table 1 there was a statistically significant difference in performance across all versions (χ^2=10.145, df = 4, p< 0.05), demonstrating that the results of all later analyses are statistically significant (p<0.05). Table 1 shows that Version 5 had the most correct executions across the board (87%), while Versions 1 and 4 had the least (76%).

Now that significance was verified across the various micro-phases and application versions, it is safe to focus on supporting or rejecting the hypothesis by comparing the versions that show visual depictions of touch points (V1, V2, V3 and V5) with those that do not (V4). Figure 5 shows the descriptive (N) and expected values for correct and incorrect responses regarding system status for the compared groups. Version 4 had more incorrect executions (24%) than the group of versions that show visual touch (V1, V2, V3, V5; 18%). As reported above, the difference is statistically significant (p<0.05).

Figure 5: *Comparison of Versions V1, V2, V3 and V5 versus V4 in terms of system status.*

In summary, the evidence supports this hypothesis by demonstrating significance across the model. The versions that visually depict the gesture (V5 and V2) produced the lowest error rates, outperforming the version that does not include a visual depiction (V4). The qualitative evidence further demonstrates the benefits of providing visual depictions to users.

8.1.1. Selected participants' comments

Clear evidence was observed that the visual cue supported the user's identification of an unfamiliar gesture. For instance, Interaction 1 required the participants to use an unfamiliar two-finger gesture to achieve what is otherwise a commonplace interaction (with one finger). Even faced with a challenge to both unlearn an existing association and learn a new association, the versions could help participants successfully make a difficult leap: "...but thanks to the interactive description otherwise I wouldn't really try to use 2 fingers" (P2, M, 30); "To activate it I have to do what the hand is doing" (P29, M, 48); and "It's actually easy looking the way you're doing it better than reading the instruction. If you want to show someone things it's better to show someone a picture or video…I know I have to put two fingers and move it down" (P32, F, 27).

By contrast, Version 4, which did not display gestural affordances, only the effect of the action, yielded high error rates for evaluation and execution. This was clearly observed in Interaction 1 and Interaction 2, in which six participants complained about the lack of visual cues for gesture or touch points. As an example: "Similar to the one before but with no dots…still unclear" (P40, M, 42); "I'd try to see if the sign comes back" (P46, M, 54); "Doesn't seem to have much point in that. Doesn't tell you anything" (P34, F, 35); and "This time I got the same symbol but without the fingers circles" (P46, M, 54).

8.2 Addressing hypothesis 'b'

This hypothesis stated that **displaying automatic visual cues before interaction will facilitate the discovery of gestures and reduce errors in execution.** In order to support or reject this hypothesis, it was necessary to compare the application versions (the IVs) that self-preview visual depictions of touch points (V1, V2 and V5) with the application version (V3) that uses 'tap-to-preview' (requiring user interaction to display the visual cues). Version 4 was removed from the analysis because it does not use visual depiction for touch. The DV is the number of errors (signalled by correct, partial and incorrect ratings) per version per 10 micro-phases.

A simple mean was drawn from the execution rates (see Table 4, appendices), and a larger success rate was found for Versions 1, 2 and 5 (82.23%) than for Version 3 (80.6%). Similar results were observed when analysing micro-phases within execution in detail. For instance, micro-phases 7 (Touch to confirm) (χ^2=30.289, df = 8, p<0.001), 8 (Touch configuration) (χ^2 = 24.940, df = 8, p<0.05) and 9 (Perform direction) (χ^2 = 49.924, df = 8, p<0.001) yielded similar results. In fact, all versions that displayed automatic SPG yielded lower error rates for these micro-phases. Noticeably, the versions that self-preview visual touch slightly outperform the versions that require touch interaction. The self-preview group (V1, V2 and V5) had fewer incorrect executions (18%) than Version 3, which uses tap-to-preview (19.4%).

This hypothesis aimed to provide a detailed examination of the micro-phases. The analysis demonstrated significant results for three of them; however, the final analysis of system status revealed no statistical significance between the grouped versions (V1, V2 and V5) and Version 3, so the null hypothesis could not be rejected. Given the small difference in measurements (Figure 6), it appears that there was no effect. This is seen in the similarity in error rates achieved at the end of the interaction, since, in both cases, more than 80% of participants completed the gesture successfully.
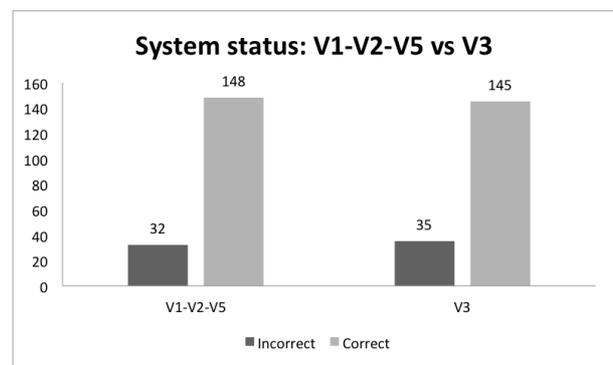
Figure 6: *Comparison of V1, V2 and V5 versus V3 in terms of system status.*

8.2.1. Selected participants' comments

A few comments show participants' reactions to automatic events, followed by correct descriptions of the implied actions of a given visual prompt. Comments include: "This came up before I touch it this time. The corner thing again. Maybe it says it is selected, no?" (P36, F, 24); and "Something showed and disappeared...two circles...maybe zoom perhaps? Ah, that was two dots, guess had to bring down" (P43, F, 35). However, five participants expressed surprise in the event of affordances being displayed automatically without any interaction from their side Comments include: "It did that because I tapped or would come anyway?" (P36, F, 24); "I didn't touch that" (P39, F, 56); "Will this always appear in the program?" (P43, F, 35); and "But I haven't done anything! Feels like it was doing something I didn't ask for" (P43, F, 35).

8.3 Addressing hypothesis 'c'

Hypothesis 'c' stated that **a rating system that segments users' gestural interactions into smaller phases will help reveal issues with users' evaluations and executions of gestures.** The null hypothesis states that a statistical analysis will show

no significant differences between phases or between the evaluation and execution of gestures.

Statistical significance has already been observed across micro-phases within evaluation and execution (see 8.1 Addressing hypothesis 'a' and 8.2 Addressing hypothesis 'b'). These results support the current hypothesis. However, this hypothesis also applies to each micro-phase individually, considering a total of 10 evaluation and execution phases. The null hypothesis cannot be rejected for micro-phases 6 (Effect on system-status; $\chi^2=3.969$, df = 8, 0.860) or T (Tap-to-preview; $\chi^2=5.663$, df = 2, p = 0.59), since these did not show significant differences.

Several additional findings that emerged from the analysis support the current hypothesis. The proportion of success and failure between the evaluation and execution phases had the added distinction of the 'partial success' rating (as explained in '5. Rating System'). A partial success outcome indicates that a user was eventually successful, but only after a number of errors in initially assessing the visual cues. Ideally, a good design will have not only a low error rate, but also a low rate of partial success.

To assess how the five different versions fared across the board in relation to partial success rates, 'observed' (obs.) and 'expected' (exp.) results were drawn for both evaluation and execution (Table 3 and Table 5 respectively, within the appendices). The 'observed' values are the real data and the 'expected' values are equally distributed across each version, thus reflecting the null hypothesis that differences are randomly discovered in different conditions. The bottom row shows the average results for all versions, demonstrating that the expected proportions are valid. Each micro-phase has different levels of expected partial success; however, we do not balance for each version, since the variation of each version from the average level is what we are testing. Figure 7 (within the appendices) shows the total 'Observed' values from each version compared across the evaluation and execution phases. In each case, the number of attempts to execute a gesture was fewer than the number of attempted evaluations.

For instance, in Version 1, a mean partial success rate was observed for 20 evaluations versus 7 executions (20/7). Thus, participants' attempts at execution were a third of their attempts to evaluate the visual prompts. The same rate was observed for Versions 2 and 5. Versions 3 and 4 showed a more marked difference, with the number of execution attempts representing a fifth of the number of evaluation attempts (19/4). In terms of the total number of attempts, Versions 2 and 5 required fewer evaluations than the other versions. Across all versions, the number of participant executions was a quarter of the number of evaluations.

9. DISCUSSION

Due to the lack of guidelines on effective design practices for communicating gestures to users, the visual solutions for the self-previewing gestures were each created with a certain degree of risk. However, the design choices were not random; they were informed by research for design. Furthermore, none of the versions used in the empirical study was intended to be ideal. Without empirical evidence, any preference would be mere speculation.

There were good reasons to suspect that some approaches would be less optimal if based on the experiences and insights gained from previous study. For example, Version 1 (static), while representing much of standard common practice, is potentially less effective in communicating direction of movement. The study also reveals that animated feedforward, in the form of self-previewing gestures, is superior to static affordance. Furthermore, tap-to-preview, on its own, proved to have limited effectiveness, although it is more effective than a basic static approach. Self-previewing affordances were found to be more effective, and it appears that the option of tapping to repeat a recent affordance may be beneficial. However, more evidence is required to determine the importance of this option to users.

10. CONCLUSIONS AND FUTURE WORK

Study hypothesis (a) was supported, demonstrating significance across the model. The versions showing a visual depiction for the gesture outperformed the version that did not. The qualitative evidence further demonstrates the benefits of providing visual depictions to users. Study hypothesis (c) was also supported, as the three-way rating criteria revealed statistically significant differences between micro-phases. The null hypothesis for hypothesis (b) could not be rejected; thus, this requires further research. The model-based rating system proved helpful in distinguishing different aspects of user performance during both the evaluation and execution stages of the interaction. The following discusses the limitations of each contribution and offers suggestions for future work:

1. GEM model: A deductive process defined the stages and micro-phases pertaining to the GEM. In order to refine the GEM, independent researchers should apply the method themselves. The data from these studies should then be compared to more closely examine the utility of the short questions corresponding to the model's micro-phases. This could identify any redundant micro-phases within the model or support the development of improved questions to elicit user responses.

2. Rating system: Ideally, any evaluation system should be validated and proved reliable by other independent researchers. Further data on the threefold rating criteria would allow the rating scheme to be tested for inter-rater reliability and, in cases of vague definitions or ill-defined or debatable heuristics, to be revised. Similarly, data collected from participants in other studies could be used to further refine the definition of partial success.

11. ACKNOWLEDGMENTS

12. REFERENCES

Bærentsen, K. B., & Trettvik, J. (2002). An activity theory approach to affordance. In *Proceedings of NordiCHI* (pp. 51–60). ACM New York, NY, USA.

Bau, O., & Mackay, W. (2008). OctoPocus: A dynamic guide for learning gesture-based command sets. In *Proceedings of UIST2008* (pp. 37–46). ACM New York, NY, USA.

Beaudouin-Lafon, M. (2000). Instrumental interaction: an interaction model for designing post-wimp user interfaces. In *Proceedings of SIGCHI conference on human factors in computing systems* (pp. 446–453). ACM New York, NY, USA.

Bennett, M., McCarthy, K., O'Modhrain, S., & Smyth, B. (2011). Simpleflow: Enhancing gestural interaction with gesture prediction, abbreviation and autocompletion. In *Proceedings of INTERACT 2011* (pp. 591–608). ACM New York, NY, USA.

Bragdon, A., Zeleznik, R., Williamson, B., Miller, T., & LaViola, Jr., J. J. (2009). Gesturebar: Improving the approachability of gesture-based interfaces. In *Proceedings of CHI '09* (pp. 2269–2278). ACM New York, NY, USA.

Chueke, J. (2016). *Perceptible affordances and feedforward for gestural interfaces: Assessing effectiveness of gesture acquisition with unfamiliar interactions* (Unpublished doctoral dissertation). City, University of London, London, UK. Retrieved from http://ethos.bl.uk/OrderDetails.do?did=3&uin=uk.bl.ethos.697556

Cross, N. (1999). Design research: A disciplined conversation. Design Issues, 15 (2), (pp. 5–10).

Dam, A. (1997). Post-WIMP user interfaces. *Communications of the ACM*, 40 (2), (pp. 63–67).

Dix, A. Mancini, R. & Levialdi, S. (1996). Alas I am undone—Reducing the risk of interaction? In *HCI'96 adjunct proceedings* (pp. 51–56). Imperial College, London. ACM New York, NY, USA.

Djajadiningrat, J. P., & Overbeeke, C. J. (2004). Interaction Frogger: A design framework to couple action and function through feedback and feedforward. In *Proceedings of DIS '04: the 5th conference on designing interactive systems: Processes, practices, methods, and techniques* (pp. 177–184). ACM New York, NY, USA.

Djajadiningrat, T. Overbeeke, K. & Wensveen, S. (2002). But how, Donald, tell us how? On the creation of meaning in interaction design through Feedforward and inherent Feedback. In *Proceedings of the 4th conference on designing interactive systems: Processes, practices, methods, and techniques* (pp. 285–291). ACM New York, NY, USA.

Frayling, C. (1993). Research in art and design. *Royal College of Art Research Papers*, 1 (1), 1–5.

Freeman, D. Benko, H. Ringel, M. & Wigdor, D. (2009). ShadowGuides: Visualizations for in-situ learning of multi-touch and whole-hand gestures. In *Proceedings of ACM Intl. Conf. on interactive tabletops and surfaces* (pp. 165–172).

Freitag, G., Tränkner, M. & Wacker, M. (2012). Enhanced feedforward for a user aware multi-touch device. In *Proceedings of NordiCHI '12: Making sense through design* (pp. 578–586). ACM New York, NY, USA.

Gaver, W. (1991). Technology affordances. In *Proceedings of CHI* (pp. 79–84). ACM New York, NY, US.

Geiwitz, J., Kornell, J. & McCloskey, B. (1990). *An expert system for the selection of knowledge acquisition techniques* (Technical Report 785-2). California: Anacapa Sciences.

Golod, I., Heidrich, F., Möllering, C., & Ziefle, M. (2013). Design principles of hand gesture interfaces for microinteractions. *Proceedings. of the 6th International Conference on Designing Pleasurable Products and Interfaces—DPPI '13*, 11, 11–20.

Gutwin, C., Cockburn, A., Scarr, J., Malacria, S., & Olson, S. C. (2014). Faster command selection on tablets with FastTap. In *Proceedings of CHI '14: 32nd annual ACM conference on human factors in computing systems* (pp. 2617–2626). ACM New York, NY, US.

Hartson, H. (2003). Cognitive, physical, sensory, and functional affordances in interaction design. *Behaviour & Information Technology*, 22 (5), 315–338.

Hofmeester, K., & Wolfe, J. (2012). Self-revealing gestures: Teaching new touch interactions in

Windows 8. In *Proceedings of CHI'12 extended abstracts on human factors in computing systems* (pp. 815–828). ACM New York, NY, USA.

Hudlicka, E. (1997). *Summary of knowledge elicitation techniques for requirements analysis* (Course material for human computer interaction). Worcester Polytechnic Institute.

Kaptelinin, V., & Nardi, B. (2012). Affordances in HCI: Toward a mediated action perspective. In *Proceedings of CHI '12* (pp. 967–976).

Koskinen I., Zimmerman, J., Binder, T., Redstrom, J., & Wensveen, S. (2011). Design research through practice: From the lab, field, and showroom. San Francisco, CA: 1st Morgan Kaufmann Publishers Inc.

McGrenere, J. & Ho, W. (2000). Affordances: Clarifying and evolving a concept. In *Proceedings of Graphic Interfaces 2000* (pp. 179–186). Toronto, Canada: Canadian Human-Computer Communications Society.

Norman, D. (1988). *The design of everyday things*. Basic Books. New York, NY, USA.

Norman, D. (1999). Affordance, conventions and design. *Interactions*, 6 (3), 38–43.

Norman, D. (2010). Natural user interfaces are not natural. In *Interactions, 17* (3), 6–10.

Norman, D. (March 2012). *Microsoft has a real opportunity with Windows 8*. Retrieved from http://tabtimes.com/feature/ittech-os-windows/2012/03/23/don-norman-microsoft-has-real-opportunity-windows-8

Norman, D. & Nielsen, J. (2010). Gestural interfaces: A step backward in usability. In *Interactions, 17* (5), 46–49.

OECD. (2015). *Frascati manual 2015: Guidelines for collecting and reporting data on research and experimental development*. Paris: OECD Publishing.

Shneiderman, B. & Plaisant, C. (2010). Designing the user interface: Strategies for effective human-computer interaction. Pearson Higher Education. ACM New York, NY, USA. Retrieved from https://www.cs.umd.edu/users/ben/goldenrules.html

Sorensen, M. (2010). Usability and affordances: Examinations of object-naming and object-task performance in haptic interfaces. *World Academy of Science, Engineering and Technology, 63*, 228–233.

Spindler, M., Stellmach, S., & Dachselt, R. (2009). PaperLens: Advanced magic lens interaction above the tabletop. In *Proceedings of the ACM international conference on interactive tabletops and surfaces (ITS'09)* (pp. 69–76). ACM New York, NY, USA.

St. Amant, R. (1999). User interface affordances in a planning representation. *Interact, 14* (3), 317–354.

Turner, P. (2005). Affordance as context. *Interacting with Computers, 17* (6), 787–800.

Vermeulen, J., Luyten, K. Van den Hoven, E., & Coninx, K. (2013). In *Proceedings of CHI '13 conference on human factors in computing systems* (pp. 1931–1940). ACM New York, NY, USA.

Wigdor, D. (2010). Architecting next-generation user interfaces. In *Proceedings of AVI '10: The international conference on advanced visual interfaces* (pp. 16–22). ACM New York, NY, USA.

Wigdor, D., & Wixon, D. (2011). *Brave NUI world: Designing natural user interfaces for touch and gesture.* Morgan Kauffman Publishers. Burlington, MA, USA.

Wigdor, D., Williams, S., Cronin, M., Levy, R., White, K., Mazeev, M., & Benko, H. (2009). Ripples: Utilizing per-contact visualizations to improve user interaction with touch displays. In *Proceedings of the 22nd Annual ACM symposium on user interface software and technology (UIST'09)* (pp. 3–12). ACM New York, NY, USA.

Wu, M., Shen, C., Ryall, K., Forlines, C., & Balakrishnan, R. (2005). Gesture Registration, Relaxation, and Reuse for Multi-Point Direct-Touch Surfaces. In *Proceedings of TABLETOP '06: The first IEEE international workshop on horizontal interactive human-computer systems* (pp. 185–192). ACM New York, NY, USA.

Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 493–502).

13. APPENDICES

Table 2: *Percentage scores and Chi-square for each micro-phase (columns) across versions (rows) of the evaluation phase.*

	2. Potential touch			3. No. of touch points			4. Touch config			5. Direction			6. Effect on system status		
	I	**P**	**C**	**I**	**P**	**C**	**I**	**P**	**C**	**I**	**P**	**C**	**I**	**P**	**C**
Vers	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%
V1	6.7	25.0	68.3	21	6	74	21	8	71	14	2	84	7	16	78
V2	1.7	16.7	81.7	12	3	85	10	4	86	3	1	96	4	14	82
V3	6.1	22.8	71.1	9	6	84	15	9	76	2	1	98	7	16	77
V4	6.1	25.0	68.9	14	4	82	21	4	74	2	1	97	6	17	77
V5	1.1	20.0	78.9	9	6	84	8	3	88	2	2	96	6	12	82
Tot	**4.3**	**21.9**	**73.8**	**13.1**	**5.0**	**81.9**	**15.1**	**5.8**	**79.1**	**4.6**	**1.3**	**94.1**	**6.0**	**14.8**	**79.2**

χ^2	df	p	χ^2	df	p	χ^2	df	p	χ^2	df	p	χ^2	df	p
19.9	8	.01	16.2	8	.04	31.1	8	.0	53.8	8	.0	3.96	8	.86

Table 3: *Expected and Observed results for the evaluation phase.*

	2. Potential touch		3. No. of touch points		4. Touch config		5. Direction		6. Effect		Total	Total
Vers.	**Obs.**	**Exp.**	**Obs.**	**Exp.**	**Obs.**	**Exp.**	**Obs.**	**Exp.**	**Obs.**	**Exp.**	**Obs.**	**Exp.**
V1	45	39.4	10	9	15	10.4	3	2.4	28	26.6	20.2	17.56
V2	30	39.4	5	9	7	10.4	2	2.4	25	26.6	13.8	17.56
V3	41	39.4	11	9	16	10.4	1	2.4	28	26.6	19.4	17.56
V4	45	39.4	8	9	8	10.4	2	2.4	30	26.6	18.6	17.56
V5	36	39.4	11	9	6	10.4	4	2.4	22	26.6	15.8	17.56
Aver.	**39.4**	**39.4**	**9**	**9**	**10.4**	**10.4**	**2.4**	**2.4**	**26.6**	**26.6**	**17.56**	**17.56**

Table 4: *Percentage scores and Chi-square for each micro-phase (columns) across versions (rows) of the execution phase.*

	T. Tap-to-preview			7. Touch to confirm			8. Set touch configuration			9. Perform direction			10. System status	
	I	**P**	**C**	**I**	**P**	**C**	**I**	**P**	**C**	**I**	**P**	**C**	**I**	**C**
Vers.	%	%	%	%	%	%	%	%	%	%	%	%	%	%
V1	NA	NA	NA	17.2	6.1	76.7	20.6	6.7	72.8	15.0	0.0	85.0	23.9	76.1
V2	NA	NA	NA	10.6	1.1	88.3	10.0	5.0	85.0	3.3	0.0	96.7	16.1	83.9
V3	51.7	1.7	46.7	6.7	0.0	93.3	14.4	7.2	78.3	2.2	0.6	97.2	19.4	80.6
V4	62.2	1.7	36.1	8.3	2.8	88.9	18.3	5.0	76.7	3.9	0.0	96.1	23.9	76.1
V5	62.8	0.6	36.7	9.4	5.0	85.6	5.6	5.0	89.4	1.1	0.6	98.3	13.3	86.7
Tot	**58.9**	**1.3**	**39.8**	**10.4**	**3.0**	**86.6**	**13.8**	**5.8**	**80.4**	**5.1**	**0.2**	**94.7**	**19.3**	**80.7**

χ^2	df	p	χ^2	df	p	χ^2	df	p	χ^2	df	p
5.66	2	.59	30.2	8	0.001	24.940	8	0.002	49.9	8	0.001

Table 5: *Expected and Observed results for the execution phase.*

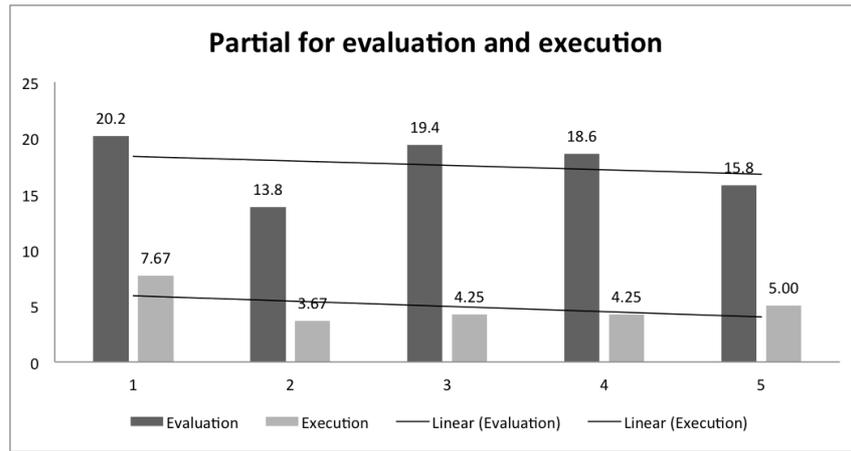|	T. Tap-to-preview		7. Touch to confirm		8. Set touch config		9. Perform direction		Total	Total
Vers.	**Obs.**	**Exp.**	**Obs.**	**Exp.**	**Obs.**	**Exp.**	**Obs.**	**Exp.**	**Obs.**	**Exp.**
V1	NA	NA	11	5.4	12	10.4	0	4	7.67	6.60
V2	NA	NA	2	5.4	9	10.4	0	4	3.67	6.60
V3	3	2.3	0	5.4	13	10.4	1	4	4.25	5.53
V4	3	2.3	5	5.4	9	10.4	0	4	4.25	5.53
V5	1	2.3	9	5.4	9	10.4	1	4	5.00	5.53
Aver.	**2.3**	**2.3**	**5.4**	**5.4**	**10.4**	**10.4**	**0.4**	**4**	**4.63**	**5.53**

Figure 7: *Comparison of Observed 'Partial' results for evaluation and execution phases.*

Table 6: *Descriptive table for the evaluation phase: scores for the success of participant identification for each micro-phase (columns) and version (rows).*

Ver	2. Potential touch			3. No. of touch points			4. Touch config			5. Direction			6. Effect on system status		
	I	P	C	I	P	C	I	P	C	I	P	C	I	P	C
V1	12	45	123	37	10	133	38	15	127	26	3	151	12	28	140
V2	3	30	147	22	5	153	18	7	155	6	2	172	7	25	148
V3	11	41	128	17	11	152	27	16	137	3	1	176	13	28	139
V4	11	45	124	25	8	147	38	8	134	3	2	175	11	30	139
V5	2	36	142	17	11	152	15	6	159	3	4	173	11	22	147
Tot	39	197	664	118	45	737	136	52	712	41	12	847	54	133	713

Table 7: *Descriptive table for the execution phase: scores per micro-phase (columns) and version (rows).*

Vers	T. Tap-to-preview			7. Touch to confirm			8. Perform swipe			9. Perform direction			10. System status	
	I	P	C	I	P	C	I	P	C	I	P	C	I	C
V1	NA	NA	NA	31	11	138	37	12	131	27	0	153	43	137
V2	NA	NA	NA	19	2	159	18	9	153	6	0	174	29	151
V3	93	3	84	12	0	168	26	13	141	4	1	175	35	145
V4	112	3	65	15	5	160	33	9	138	7	0	173	43	137
V5	113	1	66	17	9	154	10	9	161	2	1	177	24	156
Tot	318	7	215	94	27	779	124	52	724	46	2	852	174	726

Table 8: *Log-linear analysis for the evaluation phase.*

	G2	df	P
ABC	514.4	64	<0.0001
AB	397.6	8	<0.0001
AC	64.06	8	<0.0001
BC	0	16	1
AB [C]	450.34	40	<0.0001
AC [B]	116.8	40	<0.0001
BC [A]	52.74	48	0.2959

Table 9: *Log-linear analysis for the execution phase.*

	G2	df	P
ABC	1364.26	64	<0.0001
AB	724.88	8	<0.0001
AC	63.2	8	<0.0001
BC	503.94	16	<0.0001
AB [C]	797.12	40	<0.0001
AC [B]	135.44	40	<0.0001
BC [A]	576.18	48	<0.0001