Defining Gestural Interactions for Large Vertical Touch Displays

Robin Andersson[1]([⊠]), Jonas Berglund[1], Aykut Coşkun[2], Morten Fjeld[1],
and Mohammad Obaid[3]

[1] Computer Science and Engineering, Chalmers University of Technology,
Gothenburg, Sweden
k.robin.andersson@gmail.com

[2] KUAR, Media and Visual Arts Department, Koç University, Istanbul, Turkey

[3] Department of Information Technology, Uppsala University, Uppsala, Sweden

Abstract. As new technologies emerge, so do new ways of interacting with the digital domain. In this paper, the touch interaction paradigm is challenged for use on large touch displays of 65 in. in size. We present a gesture elicitation study with 26 participants carried out on twelve actions commonly used on touch displays. The results and analysis of 312 touch gestures revealed agreement rates for each action. We report several findings including the results of a set of ten unique (and a few secondary) gestures, a taxonomy classifying the defined gestures, a pilot study on the defined gestures, and explicit design implications. We discuss the results and include several important factors for future considerations. We aim at helping future designers and engineers to design interactions for large touch displays.

Keywords: Large touch display · User-defined · Gestural interaction

1 Introduction

Gestural interactions have attracted researchers to enable the next generations of intuitive input. In the context of large touch display technologies, most research looks at platform specific gestural input, for example a study on multi-touch walls [3], and studies focusing on how to improve drag-and-drop operations on large displays [4,5,11]. However, little focuses on using a user-centred approach to investigate and define the gestural interactions for common actions such as moving an item, copying or accessing a menu item as in Fig. 1.

Saffer [27] writes that the following years will likely see designers and engineers define the next generation of inputs to be used for decades. He states that designing gestural interfaces is no different to designing any interface, where the needs of the user and preferences are to be defined first. Wobbrock et al. [32] have followed a user-centered approach when defining gestural inputs for surface top interactions, inspiring other researchers in other domains to define gestures
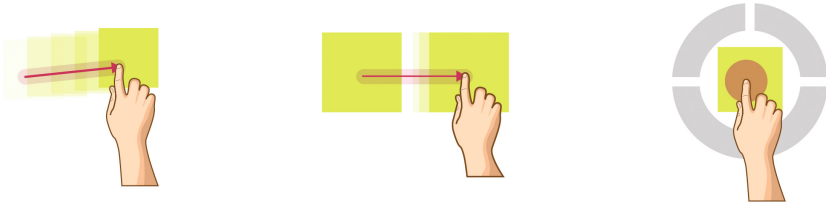
Fig. 1. Examples drawn from the user-defined gesture set: the three gestures *Move*, *Copy*, and *Menu Access* received high, intermediate, and low agreement scores.

in a similar process [13, 19, 23, 25, 32]. Related studies looking into improving the methodology [29], cultural differences [15], and legacy bias [12, 20], also exist.

With the emergence of large vertical touch displays, defining user preferences for gestural interactions has yet to be conducted. This paper attempts to remedy this gap by investigating user-defined gestures for large touch display. The word gesture is used as short for multi touch gesture in this paper if not stated otherwise. We contribute the following to the community: (1) a set of user-defined touch gestures for common actions on large touch displays, (2) a classification of the touch gestures through an existing taxonomy, (3) user agreement rates for each action, along with (4) design implications for future designers and engineers to use when designing for the platform. We confirm the contributions via a pilot evaluation with a high-fidelity prototype.

2   Background

Designing gestural interaction depends largely on the user and the technology at hand. Borrowing gestures from one technology to another (e.g. tabletops to mobiles devices) might not be appropriate [21]. Thus, in this section we present the bases of the related literature to our research which includes work on human gestures, user-defined gestures, and studies on large display interactions.

2.1   Human Gestures

There has been much research aimed at explaining and comprehending human gestures, and attempts at describing and categorizing them. Two commonly used taxonomies are employed in the examples below. Efron [6] groups gestures into five categories, namely *physiographics*, *kinetographics*, *ideographics*, *deictics*, and *batons*. Four different categories were later defined by Ekman and Friesen [7], namely *emblems*, *illustrators*, *regulators* and *adaptors*.

McNeill [16] groups gestures into five categories: *cohesive*, *beat*, *deictic*, *iconic*, and *metaphoric*. In later work, McNeill [17] defines four phases that gestures consist of; namely *preparation*, *stroke*, *hold*, and *retraction*. These phases describe the steps from first positioning the body so that it is ready to carry out the gesture (preparation phase), to performing the gesture (stroke phase), and lastly

returning to the original position (hold phase). In our work, we adopt the definition by McNeill due to its popularity currency in elicitation studies [23,24].

2.2   Large Displays and User-Defined Gestures

There has been much research done on large displays in general. For example, Chen et al. [3] introduced an approach to design interactions for Multi-Touch Interactive Wall displays, and Mateescu et al. [14] introduced *aWall*, which is an agile team collaboration tool for large multi-touch wallsystems. Nolte et al. [22], on the other hand, present a system that aims to improve collaboration when using a large interactive wall-mounted touch display. These examples present approaches and applications; however for interactions with large vertical touch displays there is a need to establish and understand users' preferences for gestural touch interactions with large displays.

Elicitation studies to define interaction is an approach that puts the user at the center of gestural design. Wobbrock et al. [31] introduced a guessability methodology to allow users, through elicitation, to suggest gestures for an interface that for them feels natural. The work was followed by another elicitation study by Wobbrock et al. [32] which lets users come up with appropriate gestures to execute a set of tasks on a tabletop platform. Many other studies have adopted the same approach of user-defined gestures into different areas of interest, such as single-handed microgestures [2], non-touch display gestures for smartwatches [1], free-hand TV control [28], navigating a humanoid robot [23], navigating a drone [24], and mid-air gestures on large wall-mounted displays [30]. Kou et al. [13] studied freehand gestures for vertical displays.

Of interest to this paper is the study by Mauney et al. [15] which investigates the effect culture has on gestures by using the same gesture-elicitation methodology. In their study, participants from China, India, the US, the UK, Germany, Finland, France, and Spain were examined to determine cultural differences in user-defined gestures on small hand-held touch displays. In our work, we also consider the cultural impact on the user-defined gestures.

In considering the literature to date, there seems to be no specific research into or design of user-defined gestural interaction for large vertical touch displays, opening a promising gap in the current state of research for large touch displays. Thus, we adopt a similar approach to [32] who studied interactions with a horizontal display (tabletops). In this context, the prospective usage of the interfaces (horizontal versus a vertical touch display) and possible applications inherently differ in several ways; for instance, while users are often seated at horizontal screens, users of vertical screens typically stand [26]. Users seated at horizontal displays often experience undesired effects of both parallaxis [18] and leaning [9]. Even though [32] did not study those effects, their existence serves as a "capstone argument" that gestures which have been found to work for horizontal displays cannot directly be transferred to vertical displays, motivating us to revisit the actions and gestures studied in [32]. Our contributions have the possibility of providing guidelines for interacting with the intended platform as well as a valuable foundation for future research to expand or investigate further.

3   User Study

While the overarching methodology is specified in the study by Wobbrock et al. [32], the details still need to be adapted to the context of a large touch display. The study included 12 actions. These were in part based on the previous work in [32] describing common actions performed on a touch screen and in part based on a brief observation session. The session consisted of video-recording user interactions with a large screen planning application during four stand-up meetings with approximately 10 participants each, followed by an analysis examining the most frequently used actions. From the observation session, four actions were found to be commonly used; specifically *Mark as complete*, *Enter Edit mode*, *Change Color* and *Create an Object*. In addition, we included eight actions featured in [32], namely *Delete*, *Long Copy*, *Short Copy*, *Fire up a context menu*, *Expand*, *Move*, *Undo single delete*, and *Undo multiple delete*. To enable the actions for the study we developed a tool that has a set of predetermined animations for the 12 actions. Figure 2 illustrates the initial state of the Move action that was shown on the display using the tool.

Two instructors followed a script to run the study; one guided the user through the session, and the other acted as a Wizard of Oz who remotely used the tool to activate the animated actions.
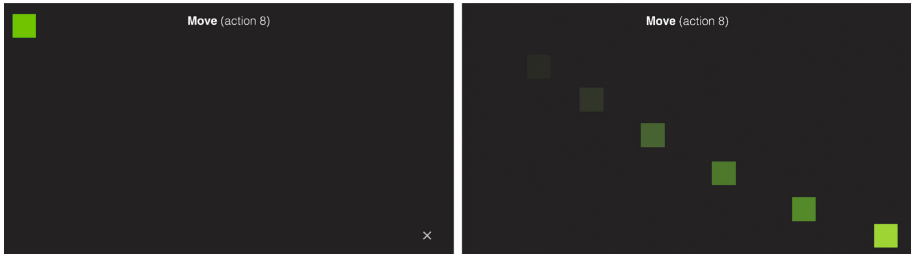

Fig. 2. Screenshot of how the move action was shown to participants (left) and illustration of its animation (right). The green object was animated in a continuous movement from the top left corner to the lower right corner. (Color figure online)

Each action contained a simple green square object displayed on a dark background, leaving out any unnecessary distraction during the pre-defined animation. The animations were designed to provide the user with the lead of what is required in the requested action, providing the user with before and after states for the green square. For the actions *Create* and *Delete an object* the object appeared and disappeared. The Long and Short *Copy* actions, and the *Move* action had a continuous animation showing the object moving along the display for a set duration to its position, see Fig. 2. For the action *Fire up a context menu*, a radial menu (also known as a pie menu) appeared around the object. The *Change Color* simply switched between green and red color abruptly.

On *Expand*, the green object contained an example text that was partially hidden until the object was expanded to twice its size. *Mark as complete* presented a check mark on the object. The *Enter Edit mode* action showed an already expanded object with an example text, and switched from a flat green background into the same background but also with darker green diagonal stripes, while showing a cursor-like line blinking for a few seconds. Both *Undo* actions (i.e. undo single delete and undo multiple delete) started by showing one or six objects respectively disappearing from the display (simulating the objects being deleted) and later re-appearing at their previous locations.

3.1   Participants

In total, 26 participants volunteered in the study (14 female). The average age was 24.3 years ($SD = 4.0$) and the average body height was 169.2 cm ($SD = 7.9$). Three participants were left handed. Most were students within the fields of Engineering, Design, Business or Psychology. Eighteen were of Turkish nationality and the remainder were comprised of one each from Singapore, England, Lebanon, Mexico, USA, and Sweden, and two from Germany. Twenty three of the participants self rated themselves as having either high or very high experience using touch displays (of any kind). It is important to note is that our participants were not chosen with their touch screen experience in mind; rather, the norm in our society is to own and fluently use smartphones with touch capabilities. Finally, participants were each offered a coffee voucher for their participation.

3.2   Apparatus

The study was set up in a studio room measuring 9×5 m, as can be seen in Fig. 3. The introduction area in the figure was used for introducing users to the study and having them read and sign the consent form before moving over to the main area. A 65 in. (165 cm) TV display was used to display the actions to the users - the exact size of the TV was selected for its suitability for use in an open office space. Two cameras - one Canon 700D DSLR camera and one Apple iPhone 6 camera - were set up on tripods at the sides of the display to record the participants' interactions throughout the study process.

3.3   Study Procedure

Each participant was first given a description of the study and asked to fill in the questionnaire. The participants were then asked to stand at a marker on the floor approximately three meters in front of the display. They watched simple animations, one at a time, of the 12 actions executed on the display, and after each one they were asked to perform a touch gesture on the display that they believed would repeat the demonstrated action. At the same time as the participant performed the gesture, the animation was triggered on the display.
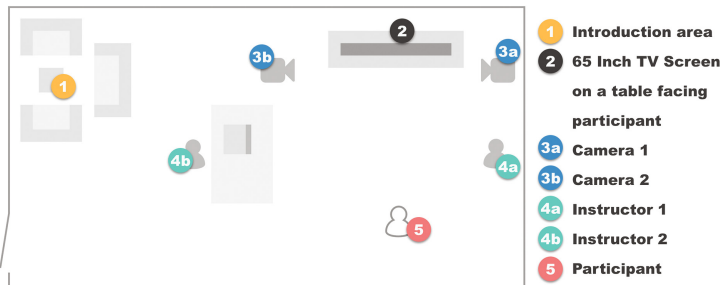
Fig. 3. Sketch of the user study setup.

After completing a gesture, the participants were asked to rate how easy it was to think of a gesture intended for the action executed on a 5-point Likert scale. The participants were also given the opportunity to add any extra comments about the completed gesture. This procedure was repeated for all 12 actions, but the order of the actions was randomized using an online random list generator[1] for each participant. This was done to avoid any order effect. One gesture was given for each action and with all 26 participants in the study, a total of $26 \times 12 = 312$ gestures were recorded for analysis. Each session took about 30–45 min to complete.

3.4   Measurements

Several aspects were measured during the study. An initial questionnaire asked for demographic data, such as nationality, age, gender, body height and dominant hand, along with technical background and profession or course of study. The video recordings were annotated using the ELAN Annotation Tool [8], marking each gesture with "start" and "end". Two researchers cross checked the annotated files. One researcher did $\approx 27\%$ of the data independent of the other researcher. The two researchers discussed and resolved any inconsistencies. In addition, all video files were independently examined by the authors multiple times and given descriptive classifications. Based on this data, the agreement rate, taxonomy and a gesture set were defined.

Agreement Rate: The agreement rate has been calculated based the formula presented in [29] and seen in Eq. 1:

$$AR(r) = \frac{|P|}{|P| - 1} \sum_{P_i \subseteq P} \left(\frac{|P_i|}{|P|} \right)^2 - \frac{|1|}{|P| - 1} \tag{1}$$

where P is a set of gestures for action r, $|P|$ is the size of the proposed set of gestures, and P_i represents the size of a subset of identical gestures within P.

---

[1] https://www.random.org/lists/.

The result is a value between 0 and 1, where 0 represents total disagreement and 1 absolute agreement. In the paper by Vatavu and Wobbrock [29] a possible interpretation of agreement rates is proposed, where 0.3–0.5 is considered a high agreement, and anything below 0.1 is considered a low agreement.

Taxonomy: All elicited gestures were defined in four taxonomy categories adopted from the paper by Wobbrock et al. [32], consisting of *Form*, *Nature*, *Binding* and *Flow* as seen in Table 1. In the *form* category, there are six different classifications; each can be applied to either hand irrespective of the other. Gestures defined as *static pose* imply that the hand is kept in the same spot and does not change its posture throughout the gesture. Gestures defined as *dynamic pose* have the used hand's pose change while it's still being held in the same position. The gestures classified as *static pose and path* gestures has the hand posture being the same through all of the gesture while the hand moves along some path. *Dynamic pose and path* gestures have the hand change its posture in the gesture while the hand moves. The *one-point touch* and *one-point path* gestures concern themselves only with a single touch point. These groups include gestures where the user might have touched the display with several fingers but with the intention of still being a single touch point. The *one-point touch* gestures contains *static pose* gestures but using only one finger. *One-point path* however is the same as the *static pose and path* form, but using only one finger.

Table 1. Taxonomy of touch display gestures adopted from the work by [32].

Taxonomy		
Form	Static pose	Hand pose is held in one location
	Dynamic pose	Hand pose changes in one location
	Static pose and path	Hand pose is held as hand moves
	Dynamic pose and path	Hand pose changes as hand moves
	One-point touch	Static pose with one finger
	One-point path	Static pose & path with one finger
Nature	symbolic	Gesture visually depicts a symbol
	Physical	Gesture acts physically on objects
	Metaphorical	Gesture indicates a metaphor
	Abstract	Gesture-referent mapping is arbitrary
Binding	object-centric	Location defined w.r.t. object features
	World-dependent	Location defined w.r.t. world features
	World-independent	Location can ignore world features
	Mixed dependencies	World-independent plus another
Flow	Discrete	Gesture occurs once
	Continuous	Gesture takes place over a period of time

Gesture Set: In a gesture set, a gesture can at most be nominated to one action; however one action can have more than one gesture assigned to it. This approach is based on previous research by Wobbrock et al. [31].

To finalize a gesture set based on the users' behaviour, the actions with higher agreement rates were prioritized to get their most frequently occurring gestures assigned to them. An action sharing the same top pick gesture with other actions - but where the action itself had a lower agreement rate - would be assigned the gesture with its second most (or third most, and so on) occurrence for that particular action.

4   Results

In this section we present the results on the agreement rates for all actions, a taxonomy for the collected data, and a user-defined set of gestures for the 12 actions included in this study. There were also other noticeable results, such as that the participants graded their previous experience with touch displays to be 4.38 ($SD = 0.68$) on a 5-point Likert scale, indicating that they were very well versed in the use of touch displays. Statements from the participants however show that gestures were inspired from both touch interfaces and the desktop (or laptop) paradigm. For instance, one participant said *"I was inspired from my MacBook. I use all the fingers and the menu appears" [par 24]*, while another said *"Easy since I have used it on my smartphone" [par 15]*.

4.1   Agreement Rates

The agreement rates for each action were calculated, following the procedure by Vatavu and Wobbrock [29]. The results show noticeable leaps in agreement rates which divide the actions into groups, such as the leaps between the actions *Move* and *Mark complete*, *Mark complete* and *Copy short*, and also between *Delete* and *Color change* (see Fig. 4).

4.2   Taxonomy

As previously explained, all elicited gestures were defined according to four taxonomy categories (see Table 1). The results of these definitions can be seen in Fig. 5. Out of all the 312 gestures provided during the study, 251 were performed using only the right hand, 20 with only the left hand, and 41 with both hands. Only two of the left handed gestures were performed by a right-handed user. In total; 23 right-handed participants provided two left-handed gestures, while the three left-handed participants provided seven right-handed gestures.

Looking at the collected data and the results represented in Figs. 4 and 5, it seems that actions within the Nature and Binding categories that are classified as physical and object-centric (also continuous from the Flow category) generally score higher. In fact, the four actions that were ranked the lowest in terms of agreement rates were also the only four that had predominantly gestures classified in the Abstract category.
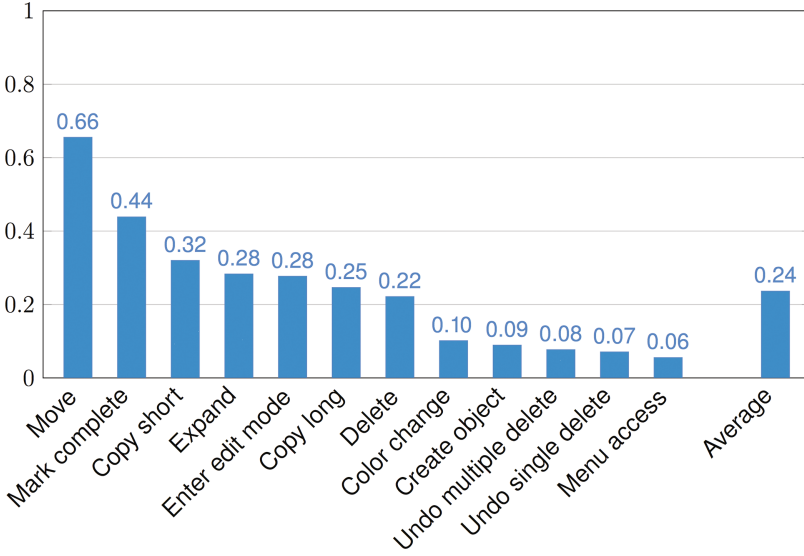
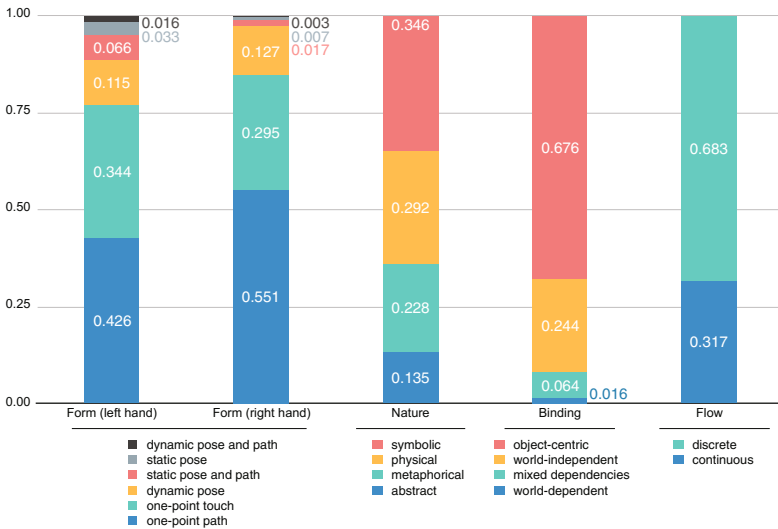**Fig. 4.** Agreement rates for the 12 actions in descending order.

Fig. 5. Taxonomy distribution, color coded by category: *Form, Nature, Binding, Flow.*

4.3   Gesture Set

A gesture set was produced as shown in Table 2 and illustrated in Fig. 6. Each single gesture corresponds to an action; a gesture is defined by its form, nature, binding, and flow (all defined in Table 2). Since the results did not show any major differences between the actions *Copy long* and *Copy short* they were

Table 2. The gesture set for the user study's actions with occurrence data. The gestures elicited for Copy long and Copy short barely differed and were thus joined into Copy. Undo single delete and Undo multiple delete were similarly joined into Undo. oc = object centric, md = mixed dependencies, wi = world independent. cont = continuous, disc = discrete. opp = one-point path, opt = one-point touch, dp = dynamic pose

Action	Gesture	Form	Nature	Binding	Flow	Occurrence
Move	Drag	opp	physical	oc	cont	81 %
Expand	Spread	dp	metaphorical	oc	cont	77 %
Mark complete	Draw Check	opp	symbolic	oc	disc	73 %
Copy	Pre-Action + Drag	opp	physical	oc	cont	62 %
	Tap S + Tap D	opt	abstract	md	disc	23 %
Delete	Flick	opp	physical	oc	disc	46 %
	Draw "X"	opp	symbolic	oc	disc	12 %
Edit mode	Double Tap	opt	metaphorical	oc	disc	42 %
Color change	Slide	opp	metaphorical	oc	disc	31 %
Create	Draw Square D	opp	symbolic	wi	disc	23 %
	Tap D	opt	abstract	wi	disc	12 %
Menu access	Tap and Hold	opt	abstract	oc	disc	19 %
Undo	Slide W	opp	abstract	wi	disc	21 %
	Rotate (CCW)	dp	metaphorical	wi	cont	10 %

combined into *Copy*. Similarly, *Undo single delete* and *Undo multiple delete* were treated the same way and were thus combined into *Undo*. This resulted in a set of 10 actions in total.

4.4   Cultural Differences

Breaking down the combined results as presented in Figs. 4 and 5 exposes the differences in agreement rate and taxonomy distribution on the cultural level. The international group (non-Turkish participants) scored overall higher agreement rates, measuring in at 0.38 compared to 0.22 for the Turkish group. Looking at individual actions, the agreement rates follow largely the same trends. There are however noticeable differences for both copy actions that show higher agreement rates for the international users (0.58 versus 0.17 for Copy Long, and 0.61 versus 0.24 for Copy Short). As for the taxonomy, both groups had almost identical numbers, with the exception of the international group not having performed any gestures defined in the forms *Dynamic pose and path*, *static pose & Static pose and path*, and instead having a larger share of one-point-path gestures.
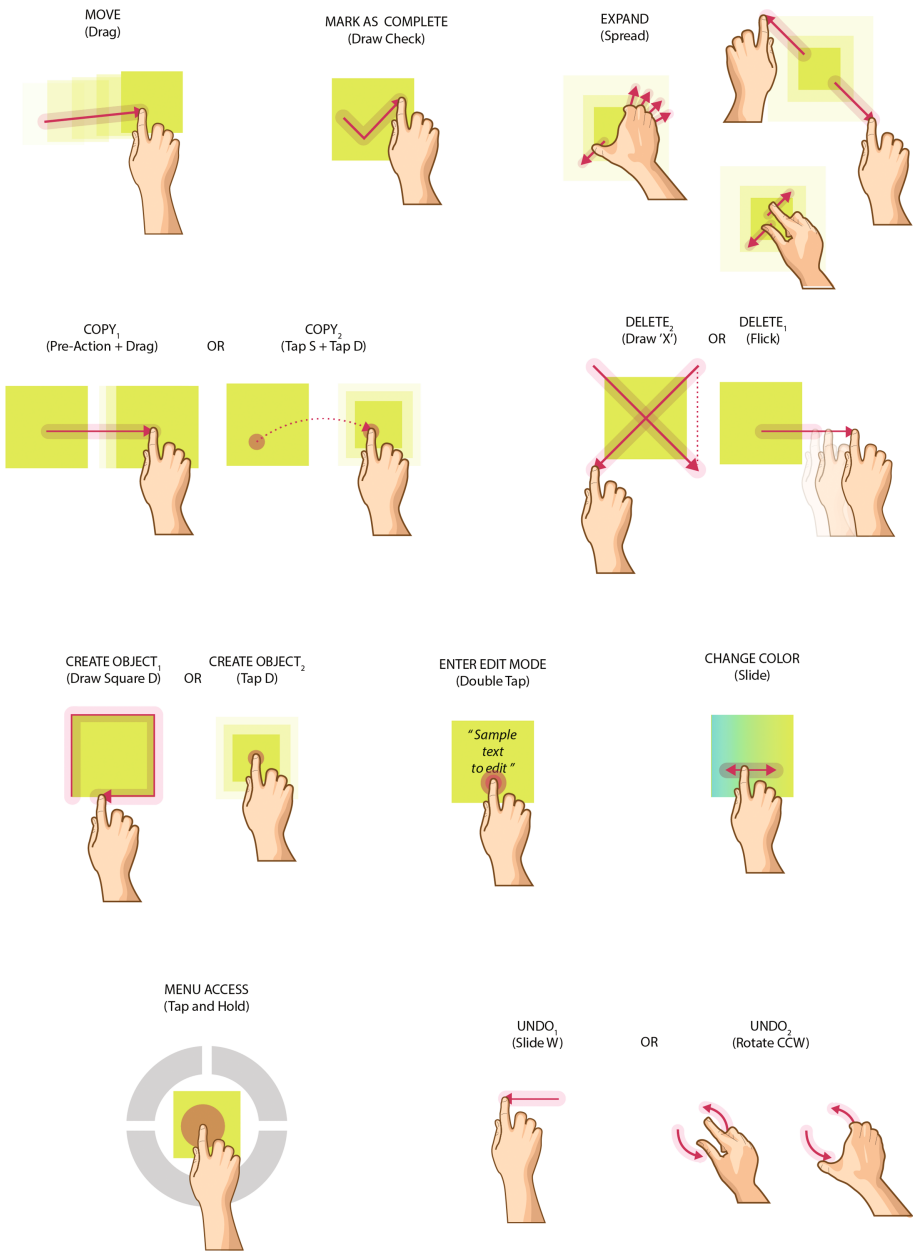
Fig. 6. Illustrations of the user-defined gesture set. Gestures illustrated with 1 finger can also be performed with 2 or 3 fingers. "S" is short for source and "D" for destination.

5   Pilot Study

The pilot study featured a fully functional system deployed on a 55 in. touch display placed in an open space at an office. The system consisted of pre-defined gestural actions deployed on a planning application (Visual Planning, by Yolean[2]) - a web based visual management tool used in organisations to visualise and organize employees' tasks on various platforms. We chose a 55 in. display as it allowed us to test the gestures on a large display of different size, suggesting how the results might be generalisable. This is considered to be a contribution to the base of several iterative cycles to validate the results, and further studies into generalisability are required and beyond the scope of the study presented in this paper.

To get an indication on whether the proposed gesture set had merit and would be well received, a prototype of four user-defined gestures was implemented. Four gestures considered to be of varying difficulty, from easy to hard, were selected based on their agreement scores (*Create an Object*, *Mark as complete*, *Enter Edit mode*, and *Delete*) that allowed for comparison with corresponding pre-defined gestures in the existing planning application. The pre-defined counterparts were all except one executed through a tap and hold to bring up a menu and then tapping the corresponding button. The one difference was *Create and Object* which were accomplished through dragging an object from the top right of the application.

Thus, the study had two conditions A and B, where condition A was the new version containing the implemented gestures (user-defined), and condition B had the pre-defined gestures from the planning application. We counter-balanced the conditions between each session to avoid any order-effect.

The hypothesis of the pilot study was that the majority of participants would favor their experience when interacting with the user-defined gestures (condition A) over the system's pre-defined gestures (B) both overall as well as for each individual metric. To test for our hypothesis, subjective ratings were collected of the users' performances via the commonly used NASA-TLX questionnaires [10] which evaluate the experience of the condition according to a set of workload performance qualities (mental demand, physical demand, temporal demand, performance, effort, and frustration). The procedure of the study had three steps:

In step one, the study instructors provided a demonstration and got the participant accustomed to the application and gestures used for the condition. Participants then got the chance to use the application, first by a guided scenario and then on their own for as long as they needed. The instructor also explained the details of the questionnaire forms that were to be completed after the session.

In step two, the participant carried out the main study task, which had the following set of instructions (kept consistent throughout the study conditions):

– "Create FOUR notes on week 43 for John Doe. Name them Task a, b, c, and d."

---

2 Yolean is a company located in Gothenburg (Sweden) that develop web applications to support lean methodologies.

- "Mark Tasks a, b and c, d, as well as TASKS 1, 2, 3 and 4 as complete."
- "Remove Tasks 1, 2 and 3, and a, b and c."

The third step consisted of collecting subjective evaluations using the NASA-TLX questionnaire. In addition, participants were asked which of the two conditions they preferred in general when interacting with the application.

5.1   Pilot Evaluation

In total 20 participants took part in the pilot study, of which 17 were male, two were female and one identified as *other*. The average age was 25.9 ($SD = 2.7$), and 19 were right-handed. Out of all participants, one was Chinese and the rest were Swedish. The participants all self-rated themselves as of high technical experience. Only two participants had any experience of the planning software.

The overall results from the collected NASA-TLX subjective data can be seen in Fig. 7. A Paired t-test was conducted to compare the NASA-TLX ratings for Condition A and Condition B. The results revealed an overall significant difference between Condition A ($M = 22.61$, $SD = 13.02$) and Condition B ($M = 28.9, SD = 15.76$); $t(19) = 2.91$, $p = 0.004$. This suggests that Condition A had lower overall NASA-TLX scores. Looking at the individual metric elements, the paired t-test shows a significant difference for the elements Physical Demand (A ($M = 26, SD = 20.49$) and B ($M = 33.25, SD = 22.08$); $t(19) = 1.91$, $p = 0.036$) and Frustration (A ($M = 21, SD = 16.19$) and B ($M = 33, SD = 20.42$); $t(19) = 3.29$, $p = 0.002$), while the other elements were not significant. When participants were asked about their preferences between condition A and B, the majority chose condition A, as shown in Table 3.

Table 3. The numbers of participants that preferred one condition over the other. Note that even though there were 20 participants some columns do not add up to 20; the reason being that some participants did not pick a preferred condition.

	Create	Edit	Complete	Delete	Overall
A	19	17	14	13	18
B	0	2	3	5	2

These results give us positive indications on the usability of the user-defined gestures, and allow us to further explore how they can be tested. In the following section we discuss the overall results and future directions for our investigation.

6   Discussion

Agreement Rates: The similarities between higher agreement rates and the Nature and Binding taxonomy categories seems to indicate that users more easily
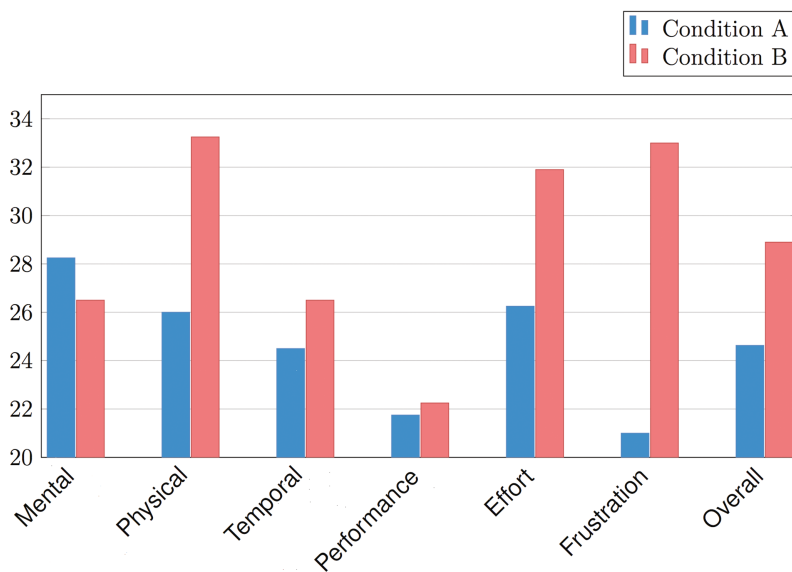
Fig. 7. The computed results from the NASA-TLX form used in the pilot study. Values range from 5 to 100. Lower is better in all metrics, and Condition A (blue) corresponds to the new gestures while Condition B (red) corresponds to the old gestures. (Color figure online)

think of gestures that relate to something they know, such as moving an object by dragging, rather than pairing their gesture with something abstract such as a tap and hold for opening up a menu. Looking at the group with lower rates (scoring <0.1), from the authors' own recollection these actions all seem to lack unique gestures on touch display devices. The actions are believed to generally be initiated through other means, such as through a dedicated button or accessed through a menu, this could be investigated in future work. If this is indeed true then the study's participants would not have been exposed to such gestures, thus resulting in the diverse set of elicited gestures as seen in the results.

Actions with low agreement rates for some actions seem to coincide to the seemingly low commonality of them having dedicated gestures. It also seems that actions with high agreement rates coincide with high commonality of dedicated gestures. For instance, *Move*, *Expand*, and *Delete* are all perceived by the authors to commonly have specific gestures devoted to them - as with the move action on iPhone which uses drag for positioning objects.

However, *Mark Complete*, *Copy Short*, *Copy Long* are not thought to commonly have dedicated gestures assigned to them. Furthermore, *Delete* is also recognized to often be situated in a menu or executed through a button. In these cases it is thought to be more fine-grained reasons for the agreement rates turning out the way they did. In the case of *Mark Complete*, the presentation of the action during the user study is perceived to be quite symbolic with the

showing of an explicit green check mark covering the whole object on display. This is thought to have been guiding the participants' gestures, hence more similar gestures and the high agreement rate. When it comes to the *Copy Short* and *Copy Long* gestures, the presentation in the user study is also believed to have played a higher part in their turnout. The presentation of the copy actions show many similarities to the *Move* action in which the new object is moved while the "old" object is left at the original position. This is thought to be the main reason why many used a pre-action (interpreted as initiating the copy) before conducting the drag gesture (interpreted as moving the copy to its intended location). As a side note, having a relatively high agreement rate for the copy actions is in line with Wobbrock et al. [32] which gives further validity to the result. Furthermore, the reason for *Delete* being situated at the bottom of its group is thought to be caused by the duality discussed previously, i.e. being present both with and without dedicated gestures in mobile touch devices.

Vertical Vs Horizontal Displays: As expected, the results confirm many of the findings from Wobbrock et al.'s work [32], where for example the patterns shown in the taxonomy categories only show a few noteworthy differences. Furthermore, agreement rates for similar actions between the studies are also quite similar. The actions might however be misleading to compare directly since explanations of them are often missing from [32], for example the animation and presentation of *Menu access* in both studies risk being too different even though the intention is most likely the same. Furthermore, the *Nature* category in this study showed a greater share of symbolic gestures.

There were some differences in the *Binding* category, which showed a very low value for *world-dependent*. The reason for these differences is believed to be related to differing legacy bias between the two studies; specifically from usage of mobile phones and touch interfaces, which in general have become more common in recent years [15, 27]. A few comments from the participants confirm this, for example *"I had to remember how to do it on my phone; had to relate to something I know" [par 11]*. It makes sense that the users relied less on gestures that made use of *world-dependent* widgets (e.g. buttons and toolbars) or other paradigms closely tied to the desktop design space. Looking at the given gestures in general, few of them are more complex or require more advanced body gestures compared to that of mobile touch displays. It seems as if the participants did not fully utilize the possibilities offered by the increased display size.

Something that is not present in the results compared to the study by Wobbrock et al. [32] is that there were no gestures elicited that moved objects to or from beyond the edges of the display. This most probably has to do with the differences in display formats between the two studies. Where the display in [32] had a wide flat edge outside of the active display suitable for such gestures, this study's display had a protruding bezel a couple of centimetres wide.

Interfering Gestures: One aspect in the proposed gesture set considered how some gestures might interfere with others so that one gesture would be identified

instead of another during an interaction. For instance, the *Mark as complete* action has the user draw a check symbol, which is essentially the same as tracing a check mark with the drag gesture. As such, the *Mark as complete* action interferes with the *Move* action. This problem can however be designed and developed around, which is why seemingly interfering gestures are still present in the gesture set (see Sect. 6.1). Looking at the study by Wobbrock et al. [32], there is no emphasis on interfering gestures, possibly because their gestures do not naturally overlap each other.

One-Handed Versus Two-Handed: Finally, the study allows users to choose if they want to perform the gesture with one or two hands. This was decided since whatever felt natural to the participants was the desired outcome. The fact that most people did provide one-handed gestures should then be seen as an indication for the preferred way to interact on large touch displays.

Successful Pilot Study Results: The pilot study results indicate that participants generally preferred the user-defined gestures over the pre-defined ones.

The pilot study was run with Swedish participants, which is a different culture from the majority of the data that constructed the user-defined set. Nevertheless, participants still leaned towards the user-defined gesture set over the pre-defined one; furthermore, as the display used in the pilot study was 55 in., which is smaller than the one used in the elicitation study, this shows promise for more general applicability. In addition, we suspect that using a planning application on a vertical display is not a common thing, and this might have affected some results for both A and B in regard to the NASA-TLX scores. However, the results indicate that user-defined gestures are less physically demanding and less frustrating, with an overall acceptance of the gestures. These are indications only, and further studies need to be conducted to validate the user-defined set on different large touch displays and the cultural impact on the user-defined set.

Cultural Impact: The higher agreement rates for the international users can in some regard be attributed to the lower participant count as stated by Vatavu and Wobbrock [29], i.e. high agreement rates correlate with low participant counts. The difference in user-defined gestures between cultures should be quite low, as stated by Mauney et al. [15]. Low differences between cultures could also be a direct result of culturally similar legacy bias. However, the nationalities present in both studies covered only four participants (two Germans, one British and one American), and as such these ratiocinations need to be investigated further.

6.1   Design Implications

Looking at the agreement rate it seems like the higher agreement rates coincide with the Nature and Binding categories. The higher agreement rates for *Physical* and *Object-centric* gestures gives a good indication that such actions are easier to think of for users, moreso compared to gestures classified towards the

Abstract spectrum. As such, designers should design interactions geared towards something more relatable; perhaps even similar to how real-world objects would behave. As there were low agreement rates for world-independent actions (and nonexistent gestures in the world-dependent *Binding*), designers should also preferably design gestures with respect to objects' positions on the display.

Participants often reached for previous knowledge when coming up with gestures and their feedback indicated this (especially referencing mobile devices), e.g. *"Seemed pretty easy, like on an iPhone" [par 12]*, *"[...] like on Android" [par 14]*, *"In every software (when editing text) you click on it" [par 16]*, *"Like Photoshop, crop and drag" [par 08]*, *"[...] like in Word or Powerpoint" [par 05]*. Designing gestures similar to those of applications in the same domain, or those on more common touch devices, should make the gestures easier to remember.

When designing interactions for even larger touch displays, tests and considerations have to be made regarding the move and copy gesture. Most user study participants did not think they would have chosen other gestures if the display was bigger, for example one participant indicated *"No I don't think so, even if it were super big" [par 21]*. A few however mentioned that drag gestures would not be suitable. For instance, one participant indicated that *"The only one that would be different is the one where I dragged. The rest would be the same." [par 12]*, another answered *"As long as the arms are able to cover the display. Might be changed otherwise. Might be abstract gestures if not reachable" [par 07]*.

Important to consider is the possibility of interfering gestures. While it is possible to develop software around interfering gestures, it should be done with care and with an understanding of its difficulties. A possible way for some of the user-defined actions could be to recognize several gestures at the same time, and then execute or discontinue one gesture on its completion or failure.

To compensate for situations similar to the pilot study's result showing higher mental demand for the suggested gesture set, both a unique gesture and a common menu-like alternative could be assigned. This could serve as a fallback if the user forgot the dedicated gesture.

6.2   Limitations

It was inevitable for the design of the gesture's animation presentation to have affected the results in some way, as showing some animation for an action could be expected to guide the user into performing a certain type of gesture. This aspect could have been improved by allowing participants to give not only one elicited gesture but multiple gestures.

For some gestures, the part of the gesture that involved "selecting" an object was removed from the interpretation of that gesture during the result computation. This was most commonly shown in the *copy*, *undo*, and *move* actions. The gesture set given by Wobbrock et al. [32] shows selection as a unique action that was incorporated into their research; thus the inclusion of selection in our study should have been considered as an action.

Our study had mostly participants from Turkish background (approx 70%), which makes our data and results dominated by the gestures performed in the Turkish culture. However, approx (30%) of our participants came from several

other countries, which to some extent had an influence on the gestural set. We elaborate on the cultural impact in our paper; however, to fully understand how to generalise the results in a more diverse cultural setup, it requires further investigations and study designs.

Furthermore, our study did not include any measures for reducing legacy bias; many gestures in the proposed set contains interactions that relate to those elicited on other platforms. While not reducing legacy bias can be advantageous as to not reinvent what does not need to be reinvented [12], there are techniques that can be used for doing so e.g. *priming* (preparing the participants for the new platform), *production* (letting participants propose multiple interactions) *partners* (letting participants participate in groups), which can yield more novel interaction techniques specific for the platform [20].

7   Conclusion and Future Work

This paper has presented a user study investigating user-defined touch gestures for large displays. In total, we analysed 312 elicited gestures from 26 participants for 12 actions commonly used on touch displays. The results and analysis reported agreement rate of user-gestures, a taxonomy of user-gestures, and a user-defined gesture set. We gave results of a pilot study evaluation of a gesture set used on a high fidelity prototype, which showed great promise in the user-defined gesture set. In addition, this paper discusses the results in detail and gives design implications to aid designers and engineers when developing interfaces for large touch displays.

This paper provides a basis for several future directions. The study by Kou et al. [13] shows that there are interesting findings to be made by allowing participants to provide multiple gestures for each action, and one can consider investigating the presented study by asking participants to provide multiple gestures for each action. Another angle for future work is considering several touch display sizes, which might reveal interesting gestural interactions depending on how large the touch display is. In addition, future studies could consider legacy bias in investigating differences in outcome by using techniques as presented by Morris et al. [20] or simply by actively selecting participants with low experience with touch screens. This could provide valuable insights into how legacy bias affects user-defined gestures on a touch display and perhaps provide more novel interactions. Finally, cultural impact on the user-defined gestures was only briefly touched upon in this paper, and future research could consider carrying out a full investigation on this important aspect that impacts upon how we interact with the interface.

References

1. Arefin Shimon, S.S., Lutton, C., Xu, Z., Morrison-Smith, S., Boucher, C., Ruiz, J.: Exploring non-touchscreen gestures for smartwatches. In: the 2016 Conference on Human Factors in Computing Systems, pp. 3822–3833. ACM (2016)
2. Chan, E., Seyed, T., Stuerzlinger, W., Yang, X.D., Maurer, F.: User elicitation on single-hand microgestures. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 3403–3414. ACM (2016)
3. Chen, W., Lao, S.Y., Lee, H., Smeaton, A.F.: Interaction design for multi-touch interactive walls. In: 2012 Second International Conference on Intelligent System Design and Engineering Application, pp. 310–313, January 2012
4. Collomb, M., Hascoët, M., Baudisch, P., Lee, B.: Improving drag-and-drop on wall-size displays. In: Proceedings of Graphics Interface 2005, GI 2005, pp. 25–32. Canadian Human-Computer Communications Society, School of Computer Science, University of Waterloo, Waterloo (2005)
5. Doeweling, S., Glaubitt, U.: Drop-and-drag: Easier drag & drop on large touch-screen displays. In: Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, NordiCHI 2010, pp. 158–167. ACM, New York (2010). http://doi.acm.org/10.1145/1868914.1868936
6. Efron, D.: Gesture and Environment. King's Crown Press, New York (1941)
7. Ekman, P., Friesen, W.V.: The repertoire of nonverbal behavior: Categories, origins, usage, and coding. Semiotica **1**(1), 49–98 (1969)
8. ELAN the language archive. https://tla.mpi.nl/tools/tla-tools/elan/. Accessed 21 Jan 2017
9. Gerken, J., Jetter, H.C., Schmidt, T., Reiterer, H.: Can "touch" get annoying?. In: Proceeding of the ACM International Conference on Interactive Tabletops and Surfaces, ITS 2010, pp. 257–258. ACM, New York (2010)
10. Hart, S.G.: NASA-Task Load Index (NASA-TLX); 20 years later. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 50, pp. 904–908. Sage Publications (2006)
11. Kobayashi, M., Igarashi, T.: Boomerang: Suspendable drag-and-drop interactions based on a throw-and-catch metaphor. In: Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology, UIST 2007, pp. 187–190. ACM, New York (2007)
12. Köpsel, A., Bubalo, N.: Benefiting from legacy bias. Interactions **22**(5), 44–47 (2015)
13. Kou, Y., Kow, Y.M., Cheng, K.: Developing intuitive gestures for spatial interaction with large public displays. In: Streitz, N., Markopoulos, P. (eds.) DAPI 2015. LNCS, vol. 9189, pp. 174–181. Springer, Cham (2015). doi:10.1007/978-3-319-20804-6_16
14. Mateescu, M., Kropp, M., Burkhard, R., Zahn, C., Vischi, D.: aWall: A socio-cognitive tool for agile team collaboration using large multi-touch wall systems. In: Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces, ITS 2015, pp. 405–408. ACM, New York (2015)
15. Mauney, D., Howarth, J., Wirtanen, A., Capra, M.: Cultural similarities and differences in user-defined gestures for touchscreen user interfaces. In: CHI 2010 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2010, pp. 4015–4020. ACM, New York (2010)
16. McNeill, D.: So you think gestures are nonverbal? Psychol. Rev. **92**(3), 350 (1985)

17. McNeill, D.: Hand and Mind: What Gestures Reveal about Thought. University of Chicago press, Chicago (1992)

18. Migge, B., Kunz, A.: Interaction error based viewpoint estimation for continuous parallax error correction on interactive screens. In: Proceedings of the IADIS International Conference: ICT, Society and Human Beings (2011)

19. Morris, M.R.: Web on the wall: Insights from a multimodal interaction elicitation study. In: Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces, ITS 2012, pp. 95–104. ACM, New York (2012)

20. Morris, M.R., Danielescu, A., Drucker, S., Fisher, D., Lee, B., Schraefel, M.C., Wobbrock, J.O.: Reducing legacy bias in gesture elicitation studies. Interactions **21**(3), 40–45 (2014)

21. Nielsen, M., Störring, M., Moeslund, T.B., Granum, E.: A procedure for developing intuitive and ergonomic gesture interfaces for HCI. In: Camurri, A., Volpe, G. (eds.) GW 2003. LNCS, vol. 2915, pp. 409–420. Springer, Heidelberg (2004). doi:10.1007/978-3-540-24598-8_38

22. Nolte, A., Brown, R., Poppe, E., Anslow, C.: Towards collaborative modelling of business processes on large interactive touch display walls. In: Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces, ITS 2015, pp. 379–384. ACM, New York (2015)

23. Obaid, M., Kistler, F., Häring, M., Bühling, R., André, E.: A framework for user-defined body gestures to control a humanoid robot. Int. J. Soc. Robot. **6**(3), 383–396 (2014)

24. Obaid, M., Kistler, F., Kasparavičiūtė, G., Yantaç, A.E., Fjeld, M.: How would you gesture navigate a drone?: a user-centered approach to control a drone. In: Proceedings of the 20th International Academic Mindtrek Conference, pp. 113–121. ACM (2016)

25. Ruiz, J., Li, Y., Lank, E.: User-defined motion gestures for mobile interaction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2011. ACM, New York (2011)

26. Ryall, K., Forlines, C., Shen, C., Morris, M.R., Everitt, K.: Experiences with and observations of direct-touch tabletops. In: Proceedings of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems, TABLETOP 2006, pp. 89–96. IEEE Computer Society, Washington, DC (2006)

27. Saffer, D.: Designing Gestural Interfaces: Touchscreens and Interactive Devices. O'Reilly Media, Inc., Sebastopol (2008)

28. Vatavu, R.D.: User-defined gestures for free-hand TV control. In: Proceedings of the 10th European Conference on Interactive TV and Video, EuroITV 2012, pp. 45–48. ACM, New York (2012)

29. Vatavu, R.D., Wobbrock, J.O.: Formalizing agreement analysis for elicitation studies: New measures, significance test, and toolkit. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI 2015, pp. 1325–1334. ACM, New York (2015)

30. Wittorf, M.L., Jakobsen, M.R.: Eliciting mid-air gestures for wall-display interaction. In: Proceedings of the 9th Nordic Conference on Human-Computer Interaction, NordiCHI 2016, pp. 3:1–3:4. ACM, New York (2016)

31. Wobbrock, J.O., Aung, H.H., Rothrock, B., Myers, B.A.: Maximizing the guessability of symbolic input. In: CHI 2005 Extended Abstracts on Human Factors in Computing Systems, pp. 1869–1872. ACM, New York (2005)

32. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2009, pp. 1083–1092. ACM, New York (2009)