

Developing Intuitive Gestures for Spatial Interaction with Large Public Displays

Yubo Kou^{1(✉)}, Yong Ming Kow², and Kelvin Cheng³

¹ Department of Informatics, University of California, Irvine, CA, USA
yubok@uci.edu

² School of Creative Media, City University of Hong Kong, Hong Kong, China
yongmkow@cityu.edu.hk

³ Keio-NUS CUTE Center, National University of Singapore,
Singapore, Singapore
idmkksc@nus.edu.sg

Abstract. Freehand gestures used in gestural-based interactive systems are often designed around technical limitations of gesture capturing technologies, resulting in gestures that may not be intuitive to users. In this paper, we investigated freehand gestures that are intuitive to users with common technical knowledge. We conducted a gesture solicitation study with 30 participants, who were asked to complete 21 tasks on a large display using freehand gestures. All gestures in the study were video-recorded. We conducted in-depth interviews with each participant to ask about the gestures they had chosen and why they had chosen them. We found that a large proportion of intuitive freehand gestures had metaphoric origins from daily uses of two-dimensional surface displays, such as smart phones and tablets. However, participants may develop new gestures, particularly when objects they are manipulating deviated from those commonly seen in surface technologies. In this paper, we discuss when and why participants developed new gestures rather than reusing gestures of similar tasks on two-dimension surface displays. We suggest design implications for gestures for large public displays.

Keywords: Large displays · Gestures · Freehand · Spatial interaction

1 Introduction

In recent years, the popularization of more robust and affordable depth-sensing cameras, such as the Microsoft Kinect, has made it easier for large display developers to include freehand gestures as user-interaction options. While significant progress has been made in hardware technologies for tracking and capturing human and their gestures [1], few research studies had been done to identify interactions that would be intuitive to ordinary users [2–4]. In addition, the gestures chosen by system designers are often restricted by technical implementation limitations such as freehand recognition ability.

Intuitiveness of gestural-based and interactive systems is particularly important in the contexts of public displays. Here, system users are very likely to have chanced upon the system, are interacting with the system for the first time, and for a relative short

period of time. This is a challenging situation since these users would quickly lose interest if interaction with the system proves difficult. Adding to this is the fact that performing freehand gestures requires more effort when compared to computing devices like mouse and keyboard [5]. Thus, users would want to complete their tasks as efficiently as possible.

In this study, we elicited freehand gestures from 30 participants who were asked to complete 21 gestural tasks in front of a large display. All gestures were video-recorded and analyzed. In addition, we conducted in-depth interviews to understand how participants derived their gestures in relation to the given tasks.

2 Related Work

Freehand gestures studies to-date have tended to begin by designing a set of gestures, and then conduct experiments to elicit performance measures, such as the accuracy of those gestures [5–8]. However, in these studies, factors such as intuitiveness of the gestures to users themselves have been taken for granted.

In studies around gestures of two-dimensional surface devices, Wobbrock et al. [9] coined the term *guessability*, to refer to “quality of symbols which allows a user to access intended referents via those symbols despite a lack of knowledge of those symbols.” They argued that in many contexts, it is unrealistic to assume that users are willing to learn the system. And therefore, systems must be designed such that users will be encouraged by early “success despite the user’s lack of knowledge of the [design intents]” [4, 9].

In Wobbrock et al.’s [4] study, users were first presented with the intended effects of gestures. They were then asked to perform gestures that would lead to these effects. By repeating this among many users, an intuitive gesture set can be derived. Other studies had utilized this approach to design gesture sets for web browser commands on large televisions [10], multi-display environments involving large surfaces and tablet devices [11], as well as motion gestures with mobile phones [12].

Recently, researchers have begun to adopt the approach to examine natural and intuitive gestures in other contexts. For example, Grandhi et al. [3] asked users to manipulate physical objects without physically touching them (freehand), such as coins, papers, and cups, to identify how users may interact with embodied systems. While our research interest is also on freehand gestures, we are more interested in their uses with large public displays.

3 Method

We adopted the guessability methodology for our user study. In each study session, we asked each of our participants to conduct a randomized series of 21 tasks using freehand gestures (Fig. 1). To generate the study tasks on large display, we adopted simple tasks from Wobbrock et al. [4] that are commonly used on the large display, and included additional map-related operations. We then extended the task set after a pilot test of five colleagues in our lab.

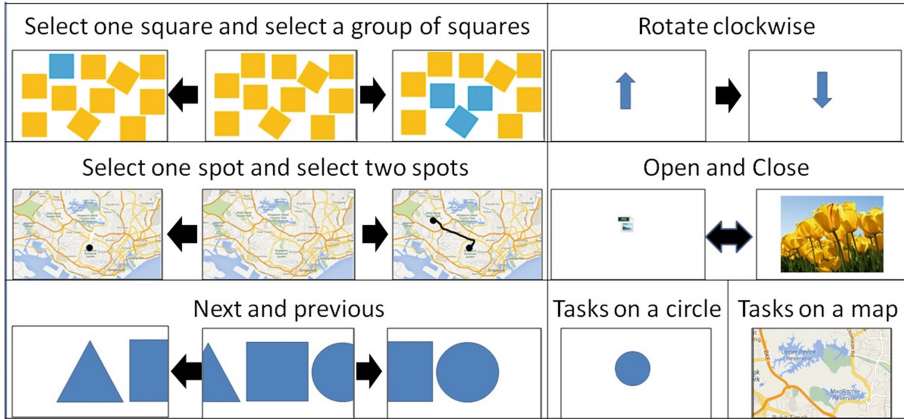


Fig. 1. 21 tasks (tasks on a circle include enlarge/shrink and move up/down/left/right and; tasks on a map include zoom in/out and pan up/down/left/right).

We recruited 30 university students to participate in the actual study. The participants aged between 18 and 31 years old, with an average age of 22.3. Among the participants, 19 were females and 11 were male. 28 participants were social science students, whereas 2 were from computer sciences. All participants were right handed. All participants, except one, owned a multi-touch smart-phone. Eleven participants also owned a tablet.

At the start of each session, we asked the participant to stand 3.5 m, or 11.5 ft, away from the projection screen. The projection screen was 2 m \times 1.5 m, or 6.6 \times 4.9 ft, in size with a resolution of 1024 \times 768. For each task, we presented the participant a Microsoft PowerPoint animation on the screen. We asked the participant to use their hands and arms in any way they like to make that animation happen again. This process was repeated until each participant had completed all 21 tasks. All of these gestures were video-recorded. At the end of each session, we conducted in-depth interviews to ask users how they came up with each gesture. If they mentioned gestures on surface technology, we further asked them what the differences or similarities were between their own gestures and those used on surface technology. All interviews were voice-recorded, transcribed, and analyzed to identify common themes.

The video-recordings include a total of 630 video-recorded gestural sequences. We analyzed all the gestures using the taxonomy of spatial gestures developed by Wobbrock et al. [4], see Table 1.

In this taxonomy, gestures may be classified based on three dimensions: hand pose, hand path, and hand number. Hand pose is defined by the physical appearance of each finger and the orientation of the hand. Static hand pose means that the hand pose does not change even when the hand moves. For example, when people use one index finger to tap, their hand pose does not change but their hand moves. Dynamic hand pose means hand pose changes while the user is performing a gesture, for example, index finger and thumb on one hand moves while closing the gap between the two fingers. Hand path refers to the trajectory and the direction of hand's movement. Hand number refers to whether participants used one or two hands to perform the tasks.

Table 1. Taxonomy of spatial gestures

Hand pose	Static hand pose	Hand pose does not change.
	Dynamic hand pose	Hand pose changes.
Hand path	Still hand	Hand stays still in the space, hand pose may change.
	Moving hand	Hand moves from one place to another in the space.
Hand Number	One hand	Using only one hand.
	Two Hands	Using two hands.

4 Findings

Our findings are presented in two sections. The first reports how these participants tried to build connections between large displays and surface technologies. The second describes the moments when our participants realized that they must create new gestures that are different from surface gestures.

4.1 Learning from Surface Technologies

In the interviews, the participants often refer to surface technologies in their daily life when they were asked to explain why they had used a gesture. For example, a 20-years-old female participant, who had only used one-index-finger in every task, told us: “*I think [the large display] is the same as my iPhone so one finger is enough.*” Another example was how some participants used one finger to do the task move the circle right because it was what they commonly do on their smart phones. We correlated this finding with our video-recorded data, and found that in our user-generated gestures, 400 or 63.5 % of these gestures were learnt from those used on multi-touch smart phones and tablets.

4.2 Diverging from Surface Technologies

We identified several factors that influence participants’ choices of gestures on large displays.

Size Matters. Our study included six tasks that involved manipulating a circle (*move the circle right/left/up/down and zoom in/out the circle*), and six tasks that involved manipulating a map. When participants were asked to manipulate maps, they used less finger-based gestures when compared to circles, from 67.8 % to 48.3 %, and more palm-based gestures, from 30 % to 51.1 %.

In our interview, a 25-years-old male participant explained that he needed two palms to *move the map* because, “*the map looks much heavier than the circle.*” Many participants who held similar opinion and explained that the circle was small, and thus one finger was enough to manipulate it. However, the map was so large that they believed they would need to use one palm to move it. Due to the perceived “size” and “weight” of the map, participants used palms, rather than fingers, as pointers to manipulate these objects. The larger the object was, the more often that hands appeared to be preferred, as a 22-years-old male participant said “*I will use two hands if the display is huge, like panorama.*”

Despite palms being seen as more appropriate for manipulating large objects, our participants expressed concerns about the perceived “accuracy.” For example, when a female participant was asked to *select one spot on a map*, she said, “*the spot is too small on the map, I’m afraid I will miss it if I don’t zoom in.*” Many participants expressed the desire to alternate between palm and finger (i.e., a finger to select spots on a map while using the palm to pan the map), e.g. “*one finger is more accurate than a whole hand!*” said a 20-year-old female participant.

With the use of palms on large objects like maps, we observed that participants interesting ways in which they would indicate direction to which they wanted to move them. When a palm was held up to the map, participants perceived that a friction existed between the palm and the map. This perceived friction guided their subsequent actions, thus deviating from their experience with surface technologies. For example, when participants were asked to move a map left, right, or down, seven (23.3 %) participants also turned their palm facing the same direction—as if this would help to slide the object in the right direction. However, this only happened twice (6.7 %) in the “pan the map up” task. One 30-year-old male participant explained, “*It is [an unnatural posture for] my palm [to] face up. I don’t usually do that [in daily life].*”

Mapping Tasks to Actions in Daily Life. Participants frequently mapped tasks to actions they often did in their daily life. For example, the task *rotate clockwise* asked participants to operate an arrow on the display, so that the arrow would rotate around its center by 180 degrees and stop when pointing towards the bottom. Ten participants (33.3 %) posed their hands in a way that looked like they were holding on to an object - they rotated their hands to perform the task. We asked them why they used their whole hand rather than merely using one finger. A 23-year-old male participant explained, “*It is similar to rotating a knob on a door. One finger is not strong enough to move the knob.*” Another participant said, “*I can’t remember I have ever done this [rotate] on my phone, so I pick [the gesture of] opening a door.*” When participants faced tasks which they did not find correspondences on surface technologies, they mapped them to what they did during their daily life.

Simplified and Complex Mental Models. Participants had varying understandings of the given tasks. Some participants made simplified assumptions. For example, they considered *close* the same as *zoom out*. This finding is consistent with Wobbrock et al.’s findings [4] where participants tended to use one simple gesture to accomplish two different tasks, such as *zoom in* and *enlarge*. However, we also observed complex mental model from other participants where they considered the tasks complex. They devised two-steps gestures for those tasks. For example, when they were asked to pick a spot on a map, they feared that they might pick a wrong spot if they were to simply point their fingers at it. Therefore, they would first zoom-out the map, and then pick the right spot. The explanation given by the participants suggested one important characteristic of large displays – *untouchability*. People cannot touch any physical objects when they are operating on public large displays. Because participants could not touch the objects on large displays as they would usually do on their touchscreen devices, they believed that tasks on large displays were more difficult and they must devise multiple steps to achieve certain goals.

5 Discussion

Our findings point to prior user experience with surface technologies as being the provider for the tacit skills transfer to the context of spatial interaction with large displays. While surface technologies were widely used, and provided convenient analogies to freehand gestures, users are able to innovate where objects to be manipulated appeared to be dissimilar to those found in previous technologies. In particular, when it comes to manipulating *small objects* on screen (e.g., a spot), designers can draw analogies from surface technology use. In fact, previous studies have also shown that this type of *one-index-finger gesture* was originally a mouse gesture that had transferred from desktops into surface technology use [4, 13, 14]. Fikkert et al. found that simple gestures based on the act of pressing buttons, a common everyday action, was the most intuitive [15]. This legacy interaction method had remained highly relevant even for spatial interactions.

For actions that were analogous to those found in surface technologies, most of our participants had opted to use one-index-finger gestures; but for large objects like maps, the same participants had opted for palm-based gestures, as these objects were perceived to be “heavier.” While one-index-finger gestures were borrowed from analogies that could be traced to the mouse pointer, and thus more *guessable*, palm-based manipulation may be a new interaction approach for computing systems utilizing spatial interactions. In this case, analogies for palm-based manipulation had not come from technology use, but instead from ordinary notions such as weight, size, and the ways such objects are used in daily life. Another example that derived from daily experience was when completing the task *rotate*, our participants mapped the arrow to a door knob and acted as if they were holding the knob. In both cases, they drew these gestures from their daily experience of interacting with objects they were familiar with. This suggests that designers should be aware of varying referents which users are likely to draw to give meanings to their freehand gestures.

When interacting with large displays at a distance, our participants were seeking a sense of physically touching the objects. They frequently derived gestures from their physical experience with surface technologies or their daily life, in which they could either touch the screen or manipulate the physical objects. Some participants realized that they could not actually touch the objects shown in large displays, as oppose to how they would interact with touchscreen devices and daily physical objects. This untouchability caused uncertainty to our participants, who were sometimes unsure whether they had successfully selected the operable objects. Therefore, they felt they must create ways to select the objects, which sometimes resulted in complicating some steps, especially in the task *next*. This suggests that designers should devise mechanisms that can provide users certainty and confidence in manipulating objects. For example, many users mentioned that they expected there would be a cursor on the large display, which followed their hands’ movement.

6 Limitations and Future Work

In this study, we have varied a few object variables including object number, size, and movement. However, other object variables may have influenced their choices, such as higher granularity of object's size, trajectory, and movement speed. In future work, we will consider more of these variables and see how they affect participants' choices.

Our current experiment setup asked participants to define one gesture for each task. Therefore, when some participants designed multiple gestures, we asked them to pick only one gesture for one task. In future study, we should consider multiple gestures and see how these can influence development of intuitive freehand gestures.

Acknowledgement. This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centre @ Singapore Funding Initiative and administered by the Interactive and Digital Media Programme Office.

References

1. Williamson, J., Murray-Smith, R.: Rewarding the original: explorations in joint user-sensor motion spaces. In: Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI 2012, pp. 1717–1726. ACM Press (2012)
2. Lee, J.C.: In search of a natural gesture XRDS Crossroads. *ACM Mag. Students* **16**, 9 (2010)
3. Grandhi, S.A., Joue, G., Mittelberg, I.: Understanding naturalness and intuitiveness in gesture production. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI 2011, pp. 821–824. ACM Press, New York (2011)
4. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI 2009, pp. 1083–1092. ACM Press, New York (2009)
5. Nancel, M., Wagner, J., Pietriga, E., Chapuis, O., Mackay, W.: Mid-air pan-and-zoom on wall-sized displays. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI 2011, pp. 177–186. ACM Press, New York (2011)
6. Malik, S., Ranjan, A., Balakrishnan, R.: Interacting with large displays from a distance with vision-tracked multi-finger gestural input. In: Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology - UIST 2005, pp. 43–52. ACM Press, New York (2005)
7. Vogel, D., Balakrishnan, R.: Distant freehand pointing and clicking on very large, high resolution displays. In: Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology - UIST 2005, pp. 33–42. ACM Press, New York (2005)
8. Fikkert, W., van der Vet, P., Nijholt, A.: Gestures in an intelligent user interface. In: Shao, L., Shan, C., Luo, J., Etoh, M. (eds.) *Multimedia Interaction and Intelligent User Interfaces*, pp. 215–242. Springer, London (2010)
9. 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 - CHI 2005, pp. 1869–1872. ACM Press, New York (2005)
10. 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 Press, New York (2012)

11. Kurdyukova, E., Redlin, M., André, E.: Studying user-defined iPad gestures for interaction in multi-display environment. In: Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces - IUI 2012, pp. 93–96. ACM Press, New York (2012)
12. Ruiz, J., Li, Y., Lank, E.: User-defined motion gestures for mobile interaction. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI 2011, pp. 197–206. ACM Press, New York (2011)
13. Fikkert, W., van der Vet, P., van der Veer, G., Nijholt, A.: Gestures for large display control. In: Kopp, S., Wachsmuth, I. (eds.) *Gesture in Embodied Communication and Human-Computer Interaction*, pp. 245–256. Springer, Heidelberg (2010)
14. Cheng, K., Pulo, K.: Direct interaction with large-scale display systems using infrared laser tracking devices. In: Proceedings of the Asia-Pacific Symposium on Information Visualization - APVis 2003, pp. 67–74. Australian Computer Society, Darlinghurst (2003)
15. Fikkert, W., van der Vet, P., Nijholt, A.: User-evaluated gestures for touchless interactions from a distance. In: 2010 IEEE International Symposium on Multimedia, pp. 153–160. IEEE (2010)