Eliciting Mid-Air Gestures for Wall-Display Interaction

Markus L. Wittorf
University of Copenhagen
Copenhagen, Denmark
markuswittorf@gmail.com

Mikkel R. Jakobsen
University of Copenhagen
Copenhagen, Denmark
mikkelrj@di.ku.dk

ABSTRACT
Freehand mid-air gestures are a promising input method for interacting with wall displays. However, work on mid-air gestures for wall-display interaction has mainly explored what is technically possible, which might not result in gestures that users would prefer. This paper presents a guessability study where 20 participants performed gestures for 25 actions on a three-meter wide display. Based on the resulting 1124 gestures, we describe user-defined mid-air gestures for wall-display interaction and characterize the types of gesture users prefer for this context. The resulting gestures were largely influenced by surface interaction; they tended to be larger and more physically-based than gestures elicited in previous studies using smaller displays.

Author Keywords
Mid-air gestures; Wall displays; Gesture elicitation.

ACM Classification Keywords
H.5.2. Information interfaces and presentation: User Interfaces

INTRODUCTION
For interacting with wall-sized displays at a distance, freehand mid-air gestures (e.g., [13]) is a promising alternative to using dedicated pointing devices such as mice or laser pointers. Compared to using a dedicated input device, freehand mid-air gestures allow several people to walk up and use a display simultaneously. Tethered input devices in contrast may constrain users from the benefits of moving in front of a wall-display (e.g., [1]).

The potential of freehand input for wall-display interaction requires mid-air gestures to be intuitive and easy to learn in this particular context. However, work on mid-air gestures for wall-display interaction has mainly explored what is technically possible, using gestures developed by system designers. Although such work often evaluates users' satisfaction with the gestures, the gestures suggested by the literature do not necessarily represent those a user would prefer. This may be problematic, as users may experience difficulties using the gestures or find it hard to learn gestures that do not match what they expect, and as a result may cause dissatisfaction.

NordiCHI '16, October 23-27, 2016, Gothenburg, Sweden
© 2016 ACM ISBN 978-1-4503-4763-1/16/10...$15.00
DOI: http://dx.doi.org/10.1145/2971485.2971503

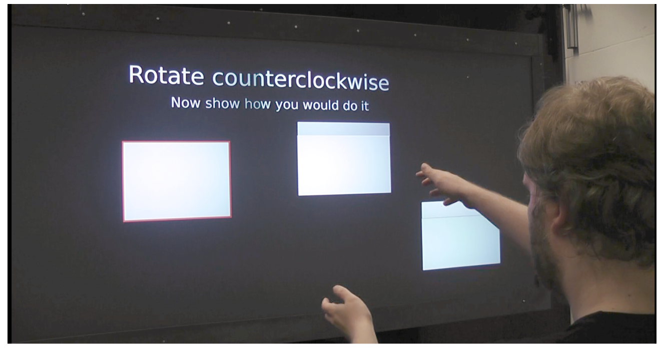

Figure 1. The wall-display used for the study showing the end state of the Rotating counterclockwise referent.

We aim to help the development of freehand mid-air gestures suitable for wall-display interaction by eliciting gestures from users. Elicitation has been used many times in the literature (e.g., [6, 10, 11, 15, 16]). However, it is unclear whether the knowledge gained from these studies can be applied to mid-air interaction with wall-sized displays. In particular, the question is how the size of the display might influence the type of gestures users would find intuitive.

This paper contributes a set of gestures elicited from 20 participants. The data may help support the design of suitable input methods for wall-sized displays, which is important as such displays become more widespread. We discuss how mid-air gestures that people find suitable for wall-display interaction may differ from mid-air gestures elicited with smaller displays (e.g., [6, 11]) or augmented reality [10], and other types of gesture-based input (e.g., [16]).

RELATED WORK

Mid-Air Wall-Display Interaction
We focus on freehand mid-air gestures, which have the benefits that they do not require a dedicated device and can be used straight away. The design of freehand gestures comes with several challenges including the lack of physical buttons to click or a surface to touch [13], and a trade-off between accuracy and speed exists in choosing between relative pointing and ray casting techniques [3, 13]. Also, a range of actions may need to be supported, from object selection [13] over pan-and-zoom [9] to text entry [5].

Prior studies often report participants' subjective satisfaction and preference with different techniques: Concerns such as ease of use and physical demand are weighted against performance. However, research rarely investigates how easy gestures are to learn for new users (although some research has studied how to "reveal" gestures on a large display [14]).

Gesture Elicitation Studies

Several studies have elicited gestures from users with the aim of maximizing the guessability of gestures while disregarding technical concerns (e.g., [15]). Typically, an effect (known as a referent) is shown and the user is asked to perform actions to cause the effect. A study comparing a user-created gesture set with one created by designers found that users preferred the user-created gesture set [8]. Several studies have elicited gestures for different interaction contexts, including surface computing [16], augmented reality [10], television [11], web interaction [6]. One rare study describes gestures by participants asked to imagine interacting with a wall-display [4]. Gestures from related work may or may not transfer to the different tasks that users may perform on wall-sized displays.

We thus elicit mid-air gestures from users focused on interacting with a wall-sized display. We compare the gestures to those in the above studies in order to understand differences and commonalities in gestures for wall-display and other applications; and also relate user-defined gestures to those suggested in large-display research.

STUDY

We elicited mid-air gestures using an adaptation of the guessability method from earlier studies [10, 16]. Informed by earlier work [7], we *primed* participants before eliciting gestures with the aim of promoting diversity in the gestures produced, and we asked participants to produce several gestures for each referent (i.e., *production*). We used *kinetic priming* where participants performed exercises to make them aware of their body potential and *visual priming* where participants were shown four video clips of different types of gestures [2] to inspire them to consider the different possibilities. Our use of production means that participants were prompted to create as many as possible before picking their favorite gesture.

Participants

Twenty volunteers (3 female), between 21 and 59 years old (M = 30.7), participated in the study; 17 were right-handed, one left-handed, and two ambidextrous. Participants had varying experience using computers and touch interfaces: 17 of them owned a smartphone. All but one participant said they rarely or never use midair technology (e.g., Kinect).

Apparatus

We used a 2.8×1.2m back-projected display with 12 HD projectors arranged in a 4×3 grid. A Java application showed the referents and kept a timestamped log of each referent shown. Each referent was shown as a sequence of images showing a beginning state, intermediate steps, and the end state (showing the effect); an on-screen message prompted participants to make gestures (as seen in Figure 1).

Sessions were recorded using three video cameras: two placed at different angles facing participants, and one placed at an angle behind them to record how they gestured in relation to the screen.

Referents

Participants were presented 25 referents (illustrations are available at http://mikkelrj.dk/research/mid-air):

Selection (6): Single select, Single deselect, Select multiple, Deselect multiple, Select all, Area select.

Element Manipulation (9): Short move, Long move, Move multiple, Rotate clockwise, Rotate counterclockwise, Shrink element, Enlarge element, Delete element, Duplicate element.

Environment Manipulation (4): Zoom in, Zoom out, Local zoom, Pan.

Video Navigation (4): Play, Stop, Rewind, Fast forward.

Collection Navigation (2): Next and Previous.

Procedure

Participants were first given an introduction to the study, including priming as previously explained. After the introduction, participants were presented with the referents in random order. Each referent was shown as a sequence of images, after which the beginning state was shown along with a text prompting the participant to suggest a gesture. Participants were then repeatedly prompted to come up with more gestures. When they could not come up with any more gestures, they were asked to show the gesture they preferred. Participants were instructed to think aloud and the experimenter asked participants for further explanations when necessary. After participants had produced gestures for all referents, they were interviewed. Here, we discussed participants' thoughts and inspirations for their gestures.

RESULTS

Participants made an average of 2.25 gestures per referent (SD = .98). Participants produced the most gestures for the Delete Element referent (M = 2.85) and the least for the Local Zoom referent (M = 1.60). Most (69%) of the gestures that participants preferred were the first gesture they produced.

Based on the recordings as well as the post-study interview we grouped gestures and derived a gesture set (following [15]); we classified the gestures based on an existing taxonomy [10]; and we characterized the gestures.

Gesture set

The study resulted in a gesture set with an agreement score of .26, similar to that of Wobbrock et al. [16]. The referents with the highest agreement in our study were Next (A = .61) and Previous (A = .51), similar to Piumsomboon et al. [10]. Select all received the lowest agreement score (A = .11). Illustrations and agreement scores for all the gestures are available at http://mikkelrj.dk/research/mid-air.

Classification of the Gestures

We classified the elicited gestures using the taxonomy of Piumsomboon et al. [10] (except *Locale* was left out as it does not fit with the parameters of this study). This taxonomy is appropriate since it was created for in-air gestures, although

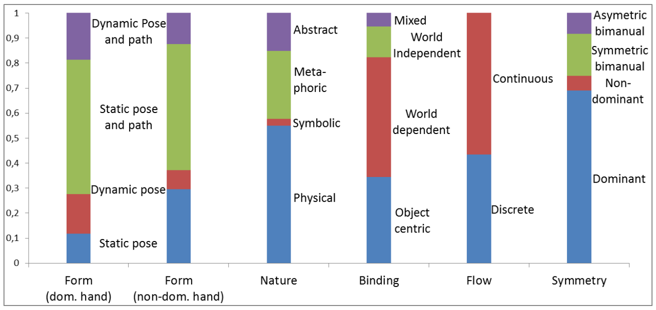

Figure 2. Share of gestures in each taxonomy category.

in an AR environment. Furthermore using a taxonomy similar to the one of prior studies allows us to compare our results to theirs. As can be seen in Figure 2, most of the gestures were performed with a static pose and path (53.8% and 50.3%), meaning that the hand was held in a specific pose while moving throughout the gesture. Also, most of the gestures were physical in nature (55.0%), meaning that participants gestured as if they were manipulating physical objects (e.g., by grabbing or pushing an object). Most gestures were either world dependent (47.8%) or object centric (34.5%), meaning that the position of the surroundings or of the object being manipulated in the gesture was important. There was a fairly even split between gestures requiring continuous feedback (56.5%) and gestures requiring discrete feedback (43.5%). It should be noted that for the majority of the referents, participants agreed on producing gestures that were either discrete or continuous: For 20 of the 25 referents, 90% or more of the gestures were discrete or were continuous. The majority of the proposed gestures were performed using only the participant's dominant hand (69.0%).

Observations

Larger variations of touch gestures. For the nine referents that resemble actions associated with common touch interactions, participants suggested gestures similar to the touch interactions. Most of these gestures typically were larger versions of common touch gestures. For the Previous referent, for example, 12 participants suggested swiping using their whole hand, while only two suggested a single-finger swipe similar to that typical for touch devices. Similarly, for Zoom in, 13 participants suggested two-handed spreads, while only three suggested finger spreads.

Pose is less important. Participants most often used unspecific or indistinct hand poses. For instance when making a swipe gesture the participants' fingers would be in a relaxed pose and simply follow the movement of the swipe rather than be in a specific pose. The only referent for which we noted a distinct pose was the Stop referent, where eight participants would hold up their hand in a stop gesture with the palm facing the screen.

Target Selection. A majority of participants (14) selected objects either by simply pointing at the target without making a specific movement to indicate the selection, or by making a small movement of the hand *towards* the target. Only one participant made a thumb click motion; and one made a circling motion to select the intended target.

Comments from Post-Study Interview

Inspiration from touch interaction. A number of participants made comments relating their gestures to touch gestures, such as saying how a gesture was "*...like on a smart phone*" (P4, P12 and P18) or "*...an iPad maneuver*" (P15).

Exact poses hold specific meaning only for some gestures. When asked about making varying indistinct poses, participants commented that the important part of their gesture was not the hand pose, but the overall direction or expression of the gesture. Participants expressed an expectation that systems allow them to vary the poses of their gestures as long as the overall impression remains the same; they only expected their hand pose to matter for gestures where the hand pose held a specific meaning. For example, participants expected to be able to vary gestures for pointing, grabbing and swiping, while they expected precision to matter for gestures such as holding up the fingers in a specific pose or drawing a symbol in the air. Some participants made comments to the contrary, however: P12 suggested having different meanings for clicks with different fingers and overall expected the pose to have meaning.

DISCUSSION

In this section we discuss how the results relate to prior work, how our work may be used to guide designers, and suggestions for future research.

Comparison to Prior Work

Preference for physical manipulation. We found that the majority of the proposed gestures (55%) were physical in nature. Piumsomboon et al. [10] in their work on augmented reality and Wobbrock et al. [16] in their work on surface technology both found high preference for physical gestures, although they were lower (39% and 44%, respectively) than the preference for physical gestures in the present study. In contrast our results show much fewer symbolic gestures (2.6%) than the other studies (10.6% and 10.0%).

Gesture size. Our findings that participants made gestures that resemble those found in other areas, but larger, is related to findings in prior studies: Piumsomboon et al. noted that the size of gestures seemed to depend on the size of the object being manipulated [10]; Vatavu et al. saw a connection between the size of the input area and the size of the gesture in their work on touch gestures [12]; and Knudsen et al. found very large gestures related to display space [4]. We extend this prior work with data suggesting that the size or extent of gestures is related to the size of the display.

Variation in gestures. Participants said they would prefer to be able to vary how they perform gestures, which has the implication that systems should allow for some variation. Prior work reports similar findings for augmented reality [10], where it was found that users would use variations of hand poses, and for surface interaction [16], where users generally did not care about the number of fingers used when gesturing.

Implications for Design

Prior work [10, 16] has shown that users often prefer using gestures resembling physical manipulation. Our study showed that this is even more pronounced for wall-sized displays. We therefore suggest that such gestures be used, whenever a natural real-world parallel exists, such as for rotating or moving objects.

Participants indicated that they only expected the exact hand pose to matter in situations where the exact pose has a specific meaning. This might suggest that cheaper, coarse-grained sensors may suffice for recognizing most gestures, whereas more fine-grained sensors may be needed to reliably recognize gestures where exact pose has such specific meaning.

Participants expressed a wish to be able to vary how they performed gestures. We thus suggest allowing for some variation in the gestures recognized by wall-display systems.

Last, our findings suggest a place for dwelling-based selection techniques: Only few of our participants used visible triggers for selections.

Implications for Research

Participants showed a stronger preference for physical manipulation (more than half the gestures), and they produced symbolic gestures less often than in prior work [10, 16]. This suggests that a larger display (and large actions as represented by the referents used) may lead to more physically-based gestures. This would be interesting to investigate in an experiment varying the size of the display and of the referents.

Also, it is clear that participants generally made large gestures, but the reason is not clear. Prior work [10, 12] gives two possible explanations. It could be due to the size of the screen or it could be due to the size of the objects shown. Whether it is one or the other, or perhaps a combination, should be investigated. If larger displays lead to larger gestures, how can we design gesture-based interactions that are intuitive without requiring excessive physical effort that cause fatigue?

CONCLUSION

We have presented a study of mid-air gestures for wall-displays with results based on 1124 gestures. The resulting gesture set and our characterization of the gestures can help understand how users prefer to use gesture-based input for wall-display interaction. A key take-away is that users make larger and more physically-based gestures than have been found in earlier work that have studied smaller displays. The gestures bear similarity to existing gestures known to users of touch technology, which might help ease the transition to gestures for mid-air input. We have also seen that participants care less about the specific hand pose or movement for some gestures and more about the general idea or expression of the gesture. We have discussed the implications this might have for gesture recognition and for future research.

REFERENCES

1. Ball, R., North, C., and Bowman, D. A. Move to improve: Promoting physical navigation to increase user performance with large displays. In *Proc. CHI 2007*, ACM (2007), 191–200.
2. Figueiredo, L. S., Gonçalves Maciel Pinheiro, M. G., Vilar Neto, E. X., and Teichrieb, V. An open catalog of hand gestures from sci-fi movies. In *Proc. CHI EA 2015*, ACM (2015), 1319–1324.
3. Jota, R., Nacenta, M. A., Jorge, J. A., Carpendale, S., and Greenberg, S. A comparison of ray pointing techniques for very large displays. In *Proc. GI 2010*, Canadian Information Processing Society (2010), 269–276.
4. Knudsen, S., Jakobsen, M. R., and Hornbæk, K. An exploratory study of how abundant display space may support data analysis. In *Proc. NordiCHI 2012*, ACM (2012), 558–567.
5. Markussen, A., Jakobsen, M. R., and Hornbæk, K. Vulture: A mid-air word-gesture keyboard. In *Proc. CHI 2014*, ACM (2014), 1073–1082.
6. Morris, M. R. Web on the wall: Insights from a multimodal interaction elicitation study. In *Proc. ITS 2012*, ACM (2012), 95–104.
7. Morris, M. R., Danielescu, A., Drucker, S., Fisher, D., Lee, B., schraefel, m. c., and Wobbrock, J. O. Reducing legacy bias in gesture elicitation studies. *interactions 21*, 3 (May 2014), 40–45.
8. Morris, M. R., Wobbrock, J. O., and Wilson, A. D. Understanding users' preferences for surface gestures. In *Proc. GI 2010*, Canadian Information Processing Society (2010), 261–268.
9. Nancel, M., Wagner, J., Pietriga, E., Chapuis, O., and Mackay, W. Mid-air pan-and-zoom on wall-sized displays. In *Proc. CHI 2011*, ACM (2011), 177–186.
10. Piumsomboon, T., Clark, A., Billinghurst, M., and Cockburn, A. User-defined gestures for augmented reality. In *INTERACT 2013*. Springer Berlin Heidelberg, 2013, 282–299.
11. Vatavu, R.-D. User-defined gestures for free-hand tv control. In *Proc. EuroiTV 2012*, ACM (2012), 45–48.
12. Vatavu, R.-D., Casiez, G., and Grisoni, L. Small, medium, or large?: Estimating the user-perceived scale of stroke gestures. In *Proc. CHI 2013*, ACM (2013), 277–280.
13. Vogel, D., and Balakrishnan, R. Distant freehand pointing and clicking on very large, high resolution displays. In *Proc. UIST 2005*, ACM (2005), 33–42.
14. Walter, R., Bailly, G., and Müller, J. Strikeapose: Revealing mid-air gestures on public displays. In *Proc. CHI 2013*, ACM (2013), 841–850.
15. Wobbrock, J. O., Aung, H. H., Rothrock, B., and Myers, B. A. Maximizing the guessability of symbolic input. In *Ext. Abstracts CHI 2005*, ACM (2005), 1869–1872.
16. Wobbrock, J. O., Morris, M. R., and Wilson, A. D. User-defined gestures for surface computing. In *Proc. CHI 2009*, CHI '09, ACM (2009), 1083–1092.