To Dwell or Not to Dwell: An Evaluation of Mid-Air Gestures for Large Information Displays

Soojeong Yoo[1], Callum Parker[2], Judy Kay[1], Martin Tomitsch[2]
1. Faculty of Engineering and IT; 2. Faculty of Architecture, Design and Planning
The University of Sydney, NSW, 2006, Australia
{syoo6624, cpar9842, judy.kay, martin.tomitsch}@sydney.edu.au

ABSTRACT
This paper investigates user preferences for mid-air gestures to interact with large public information displays. We designed and implemented a public display application that allows people to navigate between Twitter feeds and to find details about particular tweets. The application supports selection and navigation through (1) point-and-dwell and (2) push and grab-and-pull. A within-subject evaluation with 10 participants found that although point-and-dwell was perceived to be more accurate, push was preferred for selecting items and grab-and-pull was preferred for navigation. Based on our findings we derive recommendations for designing gesture-based information displays.

Author Keywords
Public displays, mid-air gestures, push, point-and-dwell, grab-and-pull, natural user interfaces.

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation: User Interfaces.

INTRODUCTION
Digital displays are increasingly being deployed in public spaces, and a large number of HCI studies have focused on aspects of both non-interactive (e.g. Hwang et al., 2008) and interactive public displays, with interaction by touch (e.g. Peltonen et al., 2008), mid-air gestures (e.g. Müller et al., 2012), or both (e.g. Jurmu et al., 2013). While touch has been found to be an intuitive interaction mechanism (Ingram et al., 2012), it does not translate well to large displays, which require users to stand at some distance to take in the entire content. For such situations, the use of gestures has been proposed as a mechanism that can provide intuitive interaction over a distance (Tomitsch et al., 2014) while enabling greater visibility of the display for both actors and bystanders.

The research presented in this paper is motivated by the following observations: (1) While many public displays are used to broadcast informational content, much of the research on gesture-based public displays has focused on the evaluation of gestures in a playful context, such as popping bubbles (Alt et al., 2013) or hitting balls (Müller et al., 2013). (2) Point-and-dwell appears to be a commonly used technique for selection and navigation (Hespanhol et al., 2012; Walter et al., 2014; Hincapié-Ramos et al., 2014). This may seem to be an intuitive approach, as graphical user interfaces, which employ a cursor controlled through a pointing device for the manipulation of objects, are widely familiar to users. However, the cursor-based manipulation of objects does not translate well to large screen displays (Aigner et al., 2012; Vanoni 2014). (3) While using distinct mid-air gestures to trigger actions may reduce the chance of the system detecting them incorrectly (Hespanhol et al., 2012), this requires simplifying the set of user interface commands or people to be trained with a detailed gesture set to support complex actions (Baudel et al., 1993).

These observations led to the study described in this paper, which investigates different approaches for implementing mid-air interaction techniques at large information displays. More specifically, our aim was to shed light on user preferences regarding commonly used mid-air gestures, namely point-and-dwell versus their semaphoric or hybrid counterparts.

BACKGROUND
A number of classifications for mid-air gestures have been proposed in the literature (e.g. Quek et al., 2002; Nancel et al., 2011; Aigner et al., 2012). In this section we describe point-and-dwell, manipulative, semaphoric, and hybrid gestures, which informed the design of the application used in our study. We further describe previous studies comparing different types of gestures.

Point-and-Dwell Gestures
Point-and-dwell gestures work by controlling the on-screen cursor with the user's hand position, enabling the cursor to follow their hand ('point'). Hovering the cursor over a selectable on-screen element for a predefined amount of time ('dwell') triggers the action associated with that element. Much of the recent research on point-and-dwell interaction was enabled through the introduction of the Microsoft Kinect, which was released together with a redesigned Xbox 360 in 2010, and the subsequent release of the Kinect SDK in 2011. Games developed for the Xbox 360 with Kinect typically use point-and-dwell for navigation and selection (Figure 1). In 2014 Microsoft released the Xbox One along with a new Kinect sensor. The Xbox One support site lists 6

OzCHI '15 , December 07 - 10 2015, Melbourne, VIC, Australia

ACM 978-1-4503-3673-4/15/12... $15.00
http://dx.doi.org/10.1145/2838739.2838819

common mid-air gestures[1]. In particular, it supports a push gesture for selection and a grab-and-pull gesture for navigation. The SDK's support of these new gestures enabled the research described in this paper, as this made it possible to compare mid-air gestures against their semaphoric or hybrid counterparts using standard consumer hardware.

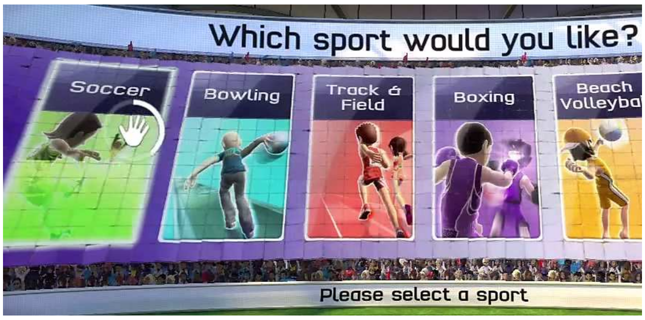

Figure 1. Selecting soccer using point-and-dwell in the sport selection menu of Kinect Sports.

Manipulative, Semaphoric and Hybrid Gestures

Quek et al. (2002) categorised gestures as either manipulative or semaphoric. They describe manipulative gestures as actual hand or arm movements with a direct link to the object being manipulated. For example, the pinch gesture on a multitouch device represents a manipulative gesture as the size of the object being manipulated is directly linked with the movement of the user's fingers. Semaphoric gestures carry meaning, which is often learnt through previous experience or using the system. For example, raising one's hand to vote on an item (Ackad et al., 2013) is a semaphoric gesture as it is based on the metaphor of raising one's arm e.g. in a classroom setting.

A hybrid approach incorporating manipulative and semaphoric gestures can allow for simplified interaction while still retaining the cursor to manipulate entities in the users direct field of view. In mid-air gesture controlled displays, hybrid gestures are often implemented using point gestures and a secondary gesture to denote an action (Mäkelä et al., 2014).

Studies Comparing Different Types of Gestures

Hespanhol et al. (2012) studied intuitiveness and learnability of four pre-designed mid-air gestures for selection and rearrangement in a cursor driven display. Intuitiveness of a gesture was determined by the amount of time taken to learn a gesture without user prompts. The study found that dwelling on an item was most intuitive for selection, followed by grabbing. However, visual feedback was provided for the dwelling gesture and not for grabbing, which may have hastened the speed of learning. The study further found that some participants preferred grabbing gestures for item rearrangement and dwelling gestures for dropping the item.

Ackad et al. (2013) reported an in-the-wild deployment of a large wall display, which offered four pre-defined mid-air gestures for interaction in a public space. The gestures consisted of two fluid manipulative gestures "swipe left" and "swipe right", and two semaphoric gestures "more" and "back", which were static poses. They reported that their manipulative gestures were consistently used by a range of 63% to 74% of the people who faced the display, while the semaphoric gestures were used by a range of 9% to 25% people. Their findings suggest that semaphoric gestures (poses in this case) may require additional time to learn the gesture and its meaning, although there was also an order effect and problems in recognition of the latter.

Summary

The merits of point-and-dwell as an interaction method are that it is intuitive and draws on the familiar mouse paradigm (Walter et al., 2014). Similarly, push also draws on from the mouse paradigm as it uses a familiar down and up motion, like a mouse click (Camp et al., 2013). What sets these gestures apart is that push involves more movement of the arm, whereas point-and-dwell requires a static pose, which can be tiring (Pyryeskin et al., 2014; Camp et al., 2013).

PROTOTYPE DESIGN AND IMPLEMENTATION

To develop a better understanding of user preferences regarding mid-air gestures identified in our literature review, we devised a gesture-based public display application allowing people to navigate Twitter feeds and to find out more about a specific "tweet" by selecting the corresponding item. To ensure authenticity of our findings, we designed the application to look and feel as real as possible, while balancing this with controlling the conditions in the experiment. For example, we used actual content from Twitter streams that were relevant to the participants and the location of the display (installed in a university building), but the content was kept exactly the same across all sessions. The front-end interface was developed as a Google Chrome web application to enable full screen mode across two projectors. A Microsoft Kinect version 2 (for Microsoft Windows) was used to detect users in front of the display and for gesture recognition. The data from the Kinect was sent to the front-end webserver via a local websockets server coded in C# and using the Kinect version 2 API.

The Tweet Wall User Interface

The application (Figure 2) displays nine Twitter feeds from the University of Sydney's faculties, with each feed being represented in one page that fills the entire screen. Arrow buttons located on both the left and right side of the screen allow users to traverse sequentially through the available feeds. The selected feed is displayed in the navigation bar, highlighted in yellow. Tweets appear in the form of tiles, for visual appeal and to provide an interface element large enough to be selected with a hand cursor. The number of tweets from each feed is limited to 15 of the latest, to fit the available space in the content area. The application includes three possible interactions: First, users can view more information about a particular tweet by selecting its tile to open a pop-up window, providing a larger image, description, date posted, and

[1] Xbox Support, Common Kinect navigation gestures on the Xbox One. http://support.xbox.com/en-AU/xbox-one/kinect/common-gestures

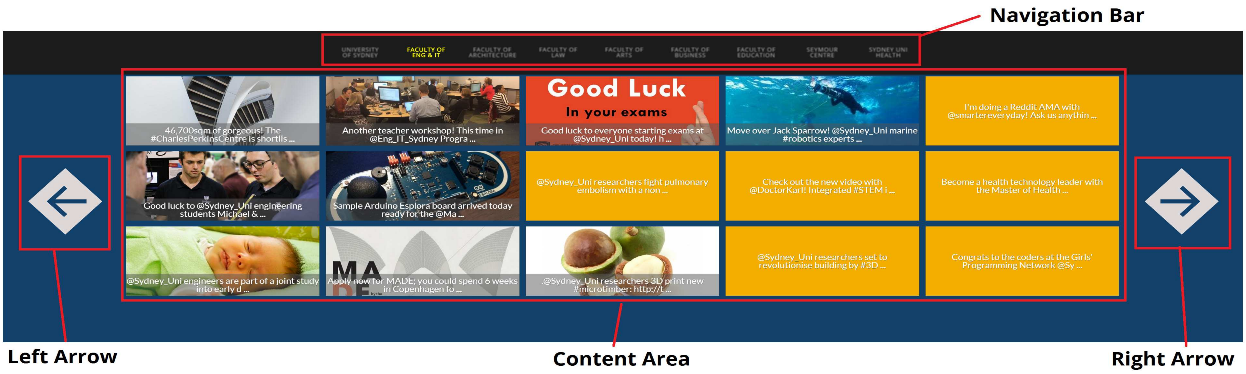

Figure 2. Tweet Wall user interface diagram.

name of the poster. Second, users can close the pop-up to return to the overview. Third, users can browse the available feeds in sequential order.

Gesture Design

We implemented the following mid-air gestures, which we identified in our literature review: (1) point-and-dwell, (2) push, and (3) grab-and-pull. In order to compare the different types of gestures, we designed the user interface so that each possible user action could be either performed using point-and-dwell or a semaphoric/hybrid gesture. The design of the gestures and the matching user interface elements were refined in an iterative process, which involved evaluating early versions of the prototypes with 10 participants. In particular, feedback from these tests determined the timing of the loading circle used for point-and-dwell, the feedback provided for the push gesture, and generally ensuring the interface always responded.

Selecting Items through Point-and-Dwell or Push

Selection of a tile can be performed using either push or point-and-dwell. Push here is a semaphoric gesture as it carries the learnt meaning of pushing a physical button. Using the push gesture, tiles can be selected by moving the cursor over the corresponding tile and subsequently performing a fluid pushing gesture motion (Figure 3). The popup window can be closed by repeating the push gesture anywhere on the screen.

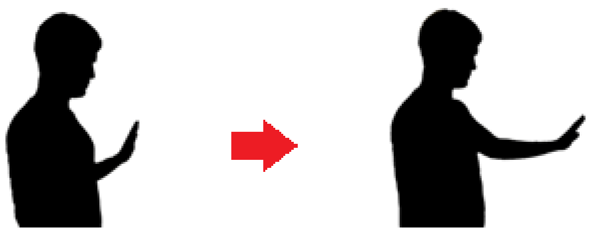

Figure 3. The stages of the push gesture, which uses a similar motion to that of tapping a screen.

Using point-and-dwell, tiles can be selected by hovering the cursor (represented in the form of a hand icon) over a particular tile. The tile behind the cursor is displayed with a white border to communicate that this is an interactive element. If the cursor is dwelled on top of a tile for half a second a loading circle appears (Figure 4). At this stage the user can decide to cancel the operation by moving the cursor away from the tile before the loading circle is completed. Upon the loading circle completing (2 seconds; the same timing as the Xbox Kinect version 2 point-and-dwell), the associated tweet is selected and displayed in a pop-up window. The pop-up window can be closed by hovering the cursor over the red "X" button located on both the top right and left corners of the screen (Figure 5).

Figure 4. Hand icon representing the cursor, which is mapped to the user's hand movement, and the loading circle providing visual feedback when hovering selectable items.

Figure 5. Pop-up window displaying a selected tweet. The popup window can be closed through dwelling on the red "X" button or by repeating the push gesture.

A loading circle indicates that the activation of the button is in progress and will be performed after 2 seconds, in a similar fashion to the activation of tiles. The replication of the button in the left and right-hand corners enabled users to activate the element with either hand.

Navigating with Point-and-Dwell or Grab-and-Pull

Navigation in the application is supported through dwelling on the navigational arrow buttons or using a grab-and-pull gesture. We used a hybrid gesture, consisting of a semaphoric grab gesture and a manipulative pull (or swipe) gesture to prevent accidently triggering a swipe (Figure 6).

After grabbing, the user can move their hand in a horizontal direction and the screen will follow. If the user has not pulled the screen far enough and lets go, the screen will ease back into the previous position;

otherwise it will ease into the next or previous category, matching the pull direction.

Figure 6. The process to trigger grab-and-pull.

STUDY SETUP

The application was evaluated at a large indoor projection wall, consisting of two projectors and spanning an area of 5 by 4 meters, with the Kinect sensor placed in the centre of the wall, underneath the projection. We evaluated the application with 10 participants (3 female, aged 19 to 46). Seven of the participants had used a Kinect with the Xbox previously. They were positioned 1.7 meters from the Kinect, to ensure reliable body tracking (Figure 7).

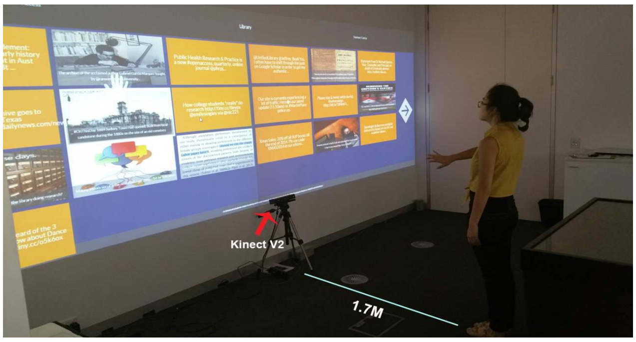

Figure 7. The Kinect version 2 tracking a user in the optimal standing range.

The two conditions were: point-and-dwell versus semaphoric/hybrid (push and grab-and-pull) interaction. We used a within-subject design and alternated the order of the two conditions for each participant. At the beginning of each condition, one of the researchers demonstrated the interface for 30 seconds. This tutorial was provided, as it was not the aim of this study to evaluate the learnability of the interface and the gestures. After the tutorial, participants were asked to explore the interface and content using the interactions from the particular condition, with each participant spending on average 5.06 minutes. Upon completion of both conditions, participants were invited to freely explore the interface using any of the available gestures. After this, participants completed a short interview inviting them to comment on what they liked about the application, the different types of gestures, and any frustrations.

We also recorded the number of problems that each participant encountered, categorised into type of gesture and type of user interaction (selection versus navigation). However, an ANOVA test showed no significant difference ($A = 0.05$, $F(3,9) = 1.4$, $p = .28$). We therefore focus on the qualitative findings in the following section.

RESULTS AND DISCUSSION

Overall, participants preferred grab-and-pull for navigating between feeds (10/10 participants) and push for viewing and closing more details contained within a popup window (8/10 participants). The literature commonly refers to point-and-dwell as the most intuitive gesture for selection (Hespanhol et al., 2012; Walter et al., 2013). However, we found that our participants preferred push over point-and-dwell due to its intuitiveness and perceived speed. This finding is surprising considering that point-and-dwell was more accurate than push (8 false triggers compared to 14), although these figures were not statistically significant. Indeed, the two participants who preferred point-and-dwell, stated that they felt it was more accurate.

The overall preference for push can likely be explained by the perceived delay occurring when using point-and-dwell, as it required participants to hold their arm still for 2 seconds. As pointed out in previous research this can also be tiring (Pyryeskin et al., 2014; Camp et al., 2013). At the same time, four participants commented that the point-and-dwell trigger was too fast, potentially leading to accidental activation of items. Another participant commented on point-and-dwell being fast at first but as they became more proficient in using the system it felt slow. We therefore suggest considering an adaptive approach, where the dwell time decreases as users become more proficient with using the system. However, supporting semaphoric or hybrid gestures as an alternative can offer faster interaction for proficient users.

All participants found grab-and-pull to be faster, more “fun”, and “intuitive” compared to point-and-dwell. Participants also liked the gesture because it felt familiar to swiping on multitouch devices, as one participant stated: “Grab-and-pull felt very natural and familiar to the iPad”. Interestingly, previous use of the Kinect did not seem to affect the number of problems experienced.

Additionally, every participant commented on grab-and-pull being less physically demanding than using point-and-dwell for navigation. It was deemed faster to perform than point-and-dwell, which requires users to stretch their arm to reach an arrow button on either side of the screen, while grab-and-pull can be performed anywhere on screen. Greater cursor acceleration could potentially reduce stretching.

CONCLUSION

Through the evaluation of our public information display application, we found that participants considered point-and-dwell to be more accurate but slower compared to push and grab-and-pull. All gestures draw on familiar interaction paradigms, ensuring they are perceived to be intuitive. However, push and grab-and-pull were described as more fun, suggesting that the cursor metaphor does not translate well to gesture-based large display applications. Proposed strategies for improving user acceptance of point-and-dwell interfaces include adaptive dwell times, greater cursor acceleration, supporting non-sequential navigation, and keeping interactive elements close to the center of the screen within reach of the user.

ACKNOWLEDGEMENTS

This research was supported by funding from the Faculty of Engineering & Information Technologies, The University of Sydney, Faculty Research Cluster Program.

REFERENCES

Ackad, C., Wasinger, R., Gluga, R., Kay, J., and Tomitsch, M. Measuring Interactivity at an Interactive Public Information Display. In Proc. OzCHI 2013, ACM Press (2013), 329-332.

Aigner, R., Wigdor, D., Benko, H., Haller, M., Lindbauer, D., Ion, A., and Koh, J. T. K. V. Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for hci. Microsoft Research TechReport MSR-TR-2012-111, (2012).

Alt, F., Schneegass, S., Girgis, M., and Schmidt, A. Cognitive effects of interactive public display applications. In Proc. PerDis 2013, ACM (2013), 13-18.

Baudel, T. and Beaudouin-Lafon, M. Charade: remote control of objects using free-hand gestures. Communications of the ACM 36,7 (1993), 28-35.

Camp, F., Schick, A. and Stiefelhagen, R. How to click in mid-air. In Distributed, Ambient, and Pervasive Interactions, volume 8028 of LCNS (2013), 78–86.

Hespanhol, L., Tomitsch, M., Grace, K., Collins, A. and Kay, J. Investigating intuitiveness and effectiveness of gestures for free spatial interaction with large displays. In Proc. PerDis 2012, ACM Press (2012),1-6.

Ingram, A., Wang, X. and Ribarsky, W. Towards the establishment of a framework for intuitive multi-touch interaction design. In Proc. AVI 2012, ACM Press (2012), 66-73.

Hincapié-Ramos, J. D., Guo, X., Moghadasian, M. and Irani, P. Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In Proc. CHI 2014, ACM (2014), 1063-1072.

Jurmu, M., Ogawa, M., Boring, S., Riekki, J., &Tokuda, H. Waving to a touch interface: Descriptive field study of a multipurpose multimodal public display. In Proc. PerDis 2013, ACM Press (2013), 7-12.

Mäkelä, V., Heimonen, T., Luhtala, M., and Turunen, M. Information wall: evaluation of a gesture-controlled public display. In Proc. MUM 2014, ACM Press (2014), 228-231.

Müller, J., Walter, R., Bailly, G., Nischt, M., & Alt, F. Looking glass: a field study on noticing interactivity of a shop window. In Proc. CHI 2012, ACM Press (2012), 297-306.

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.

Peltonen, P., Kurvinen, E., Salovaara, A., Jacucci, G., Ilmonen, T., Evans, J. and Saarikko, P. It's Mine, Don't Touch! interactions at a large multi-touch display in a city centre. In Proc. CHI 2008, ACM Press (2008), 1285-1294.

Pyryeskin, D., Hancock, M. and Hoey, J. Comparing elicited gestures to designer-created gestures for selection above a multitouch surface. In Proc. ITS 2012, ACM (2012), 1-10.

Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X.-F., Kirbas, C. and Ansari, R. Multimodal human discourse: gesture and speech. TOCHI 9. 3 (2002), 171–193.

Tomitsch, M., Ackad, C., Dawson, O., Hespanhol, L. and Kay, J. Who cares about the Content? An Analysis of Playful Behaviour at a Public Display. In Proc. PerDis 2014, ACM Press (2014), 160.

Vanoni, D. Human-Centered Interfaces for Large, High-Resolution Visualization Systems .In Research Exam. UC San Diego Computer Science and Engineering Department, retrieved, (2014).

Walter, R., Bailly, G. and Müller, J. Strikeapose: revealing mid-air gestures on public displays. In Proc. CHI 2013, ACM Press (2013), 841-850.

Walter, R., Bailly, G., Valkanova, N. and Müller, J. Cuenesics: using mid-air gestures to select items on interactive public displays. In Proc. MobileHCI 2014, ACM Press (2014), 299-308.