BYOD: Bring Your Own Device

Rafael Ballagas[1], Michael Rohs[2], Jennifer G. Sheridan[3], and Jan Borchers[1]

[1] Media Computing Group, RWTH Aachen University
52072 Aachen, Germany
{ballagas, borchers}@cs.rwth-aachen.de
[2] Institute for Pervasive Computing, ETH Zürich
Zürich, Switzerland
rohs@inf.ethz.ch
[3] Ubiquitous Computing Group, Lancaster University
Lancaster, England
sheridaj@comp.lancs.ac.uk

Abstract. In this work, we examine the different types of user interactions and deployment issues surrounding large public displays. We then select and develop a usage paradigm in which people use personal devices to interact with large public displays. Lastly, we present a physical interface for interacting with large public displays based on camera-enabled mobile phones. Our interface uses both Visual Codes and optical-flow processing to support serendipitous interaction with large public displays.

1 Introduction

Interaction with large displays can be largely broken down into three application domains: personal, semi-public, and public. Personal large displays allow a single user to visualize and process large amounts of information at once [1]. Semi-public large displays are situated in a controlled access environment such as an office building or a meeting room where a limited amount of people regularly interact, usually collaboratively in single display groupware [2] applications. Large public displays are characteristically placed in locations that are open, usually in environments with high pedestrian traffic and extended wait times such as train stations, airports or theme parks.

Fig. 1. (Left) A group of semi-public displays in the iRoom, an interactive conference room at Stanford University. (Right) A large public display in a subway stop in Vienna, Austria.

In this paper we focus on enabling interactions with large public displays. We have identified a few design considerations that are specific to interaction techniques for large public displays including:

– **Serendipity** corresponds to users' ability to spontaneously interact with a large display. High serendipity means a very low threshold of use with any arbitrary display.

- **Portability** refers to users' ability to transport the tools necessary with which to interact with the display. Interaction mechanisms that only require what the users have with them on their body (including the body itself) have high portability, and systems that require users to carry awkward or inconvenient equipment around with them have low portability.
- **Sanitation** corresponds to the cleanliness and health considerations associated with an interaction technique. The physical condition of the display will affect interaction. In some cases, well-used displays will have a positive effect on usage, such as with arcade games. Conversely, a dirty public kiosk will have a negative effect. Ubiquity plays an important role here; even though public bank machines are used several times a day, most people do not seem to have any sanitation concerns.
- **Dexterity** refers to how many hands are required for operation. This is an important consideration for public environments where the user might need their hands to carry bags or other personal belongings.
- **Multi-User** corresponds to the ability of an interaction technique to support multiple users simultaneously and to arbitrate access to the display. This is important because large public displays are typically surrounded by many people at any given time.
- **Physical Security** Large public display systems must be protected from vandalism and theft. For example, in the subway station in Figure 1, the projector is protected in steel casing, and the display area is located across the rail tracks, making it inaccessible to vandals.
- **Information Security and Privacy** refers to the degree of security and privacy that a user can expect when interacting with a large display in public. The interaction technique has to ensure, e.g., that sensitive data, like name or phone number, are not accidentally shown on the large public display.
- **Social Acceptability** refers to the acceptability of an interaction technique in the presence of others, who passively observe the interaction. The interaction technique might be disturbing to observers, embarrassing to the user, or even raise the social status of a person.
- **Interruptability** refers to the fact that most interactions with large public displays tend to be short-term, and are often interrupted by external events. Candidate interaction techniques have to cope with these issues.
- **Intentional vs. Unintentional Interaction** denotes two modes of interaction initiation. The first case, which we focus on in this paper, requires the user to actively initiate the interaction. In the second case, the large display might initiate interaction upon sensing the presence of a user.
- **Maintenance** refers to the amount of regular service the system needs to stay operational and maintain an appearance that attracts user interaction.

2 Interaction Techniques

Today, the most common large display interaction paradigm involves users interacting directly with a large display surface. This can be accomplished with capacitive surfaces or DViT (Digital Vision Touch) technology like the SMART Board [3], or it can be done using active ultrasonic pens such as Mimio Virtual Ink [4]. In terms of usability, this paradigm has many positive aspects in that it has a direct one-to-one mapping, and a clear affordance. However, one problem with the direct mapping is scalability. Emerging display technologies such as OLED [5] promise displays that can be hung like wallpaper and large public displays such as these will inevitably extend beyond human reach, making the direct mapping less realistic and desirable. Direct interaction also puts users at an awkward viewing perspective: with very large displays, users might struggle between getting close enough to the display to interact with it, and getting away from the display far enough to put everything into their field of vision. The direct mapping paradigm also suggests that the interactive system is directly accessible to the user, reducing physical security. Touch-based input offers high serendipity, but the sanitation is low. (Whilst users might conceivably carry a stylus with them to improve on cleanliness, this lowers the system portability.) Lastly, the existing technologies in this paradigm do not handle multiple-user simultaneity very well, although there are a few emerging technologies[6] that address this problem.

The main issue with other implement-free mechanisms, such as gesture and speech, is that public areas are generally both audibly and visually noisy, making it very difficult for programs to distinguish between intentional interaction and random noise. While attempts have been made to reduce noise, such as with the Barehands [7] project which reduced noise by limiting gestures to those made directly on the display surface, having direct physical contact with the display leads to the problems described above. Speech utilizes sound and therefore inherently limits user privacy and has the potential to disturb others. Speech is a well studied input modality and there are many known concerns about switching between data input and control input [8]. Gesture input can be well suited for limited control command sets, however, complex gesture sets are difficult to learn. While combined speech and gesture modalities [9] hold some promise for single-user large display interaction in low-noise environments like offices, noise will continue to be an issue in public arenas such as train stations.

Device-based interaction is the other clear interaction paradigm. However, providing publicly available devices raises concerns about physical security, sanitation, maintenance, and multi-user simultaneity. Allowing people to interact with displays using personal devices that they already have on them resolves these issues. Since devices with communication capabilities like mobile phones, PDAs, and key-fobs are becoming ubiquitous, we can harness their capabilities for large public display interaction. For example, in Scott Minneman's project "Sunset: 200Mhz in a 35mph Zone" [10] a soap opera unfolded over several days on a pair of Sony JumboTron video billboards which were designed to be seen by the thousands of viewers who drive (or rather, sit in traffic) along the Sunset Strip in Los Angeles daily. Drivers could influence the outcome of the unfolding drama by clicking their garage-door openers or alarm system key-fobs. Several projects have used handheld PDAs to interact with large displays [11–14]. These projects benefited from users' familiarity with their own device. In terms of technology, supporting multiple users under this paradigm becomes more straightforward with connectivity standards such as GPRS, WiFi, and Bluetooth.

3 Device Details

In this paper, we elected to use a Nokia 6600 camera-enabled mobile phone as our platform in which to study interactions with large public displays. A "naive" look at this particular mobile phone as an input device would classify it as a 20 button keyboard combined with a joystick. Figure 2 shows its properties in a Card [15] style design space. The joystick has the same properties as the other keys, in that a single push in a certain direction results in one discrete key press event. However, the low-resolution integrated camera, found in today's mobile phones, can function as a special optical sensor and therefore provide an additional input channel. Previous work by one of the authors [16] shows how phone cameras can be used to detect 2-D bar codes (see Figure 4). Phones can detect codes from any arbitrary orientation, tilt, or rotation and the codes themselves can store up to 83 bits of information. This information can serve as a digital handle for a physical object, like a URL, or various other ID codes.

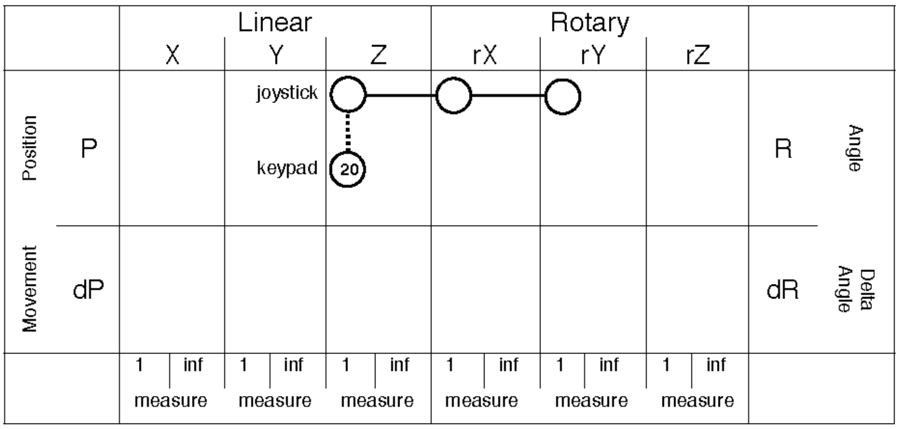

Fig. 2. Naive classification of mobile phone as an input device following the taxonomy presented in [15].

For our interactions, we use our phone cameras to enable both optical sensing and optical flow processing. Optical flow processing involves rapidly sampling successive images from a camera phone and sequentially comparing them to determine relative motion in the (x, y, θ) dimensions. This enables the camera to be used as a three degree of freedom (DOF) input device. As shown in the classification in Figure 3, the interaction properties of the device become richer by adding the camera as a relative movement sensor. Given a camera resolution that is fine enough, in principle arbitrarily small relative motion updates can be sensed, instead of the discrete values provided by the joystick. The three circles correspond to the 3 DOF and map to the (x, y, θ) dimensions. We did not draw circles for relative Z movement and relative X and Y rotation, in order not to clutter up the diagram and focus on the most

important properties. In our implementation, relative rotation around the X axis (dR:rX) is equivalent to linear Y motion and relative rotation around the Y axis (dR:rY) is equivalent to linear X motion. In addition, relative Z movement (dP:Z) could be mapped to a further input dimension.
To invoke this function, users press and hold a button, which acts as a clutch, to indicate to the system that they are actively controlling the cursor, and then they wave the phone in the air to control the (x, y, θ) input. Users can release the clutch button to reposition their arm, which is similar to the way a mouse can be lifted to be repositioned on a desktop surface. This means that the camera need not be pointed directly at the display but can be pointed at the floor to allow users a more comfortable arm posture.

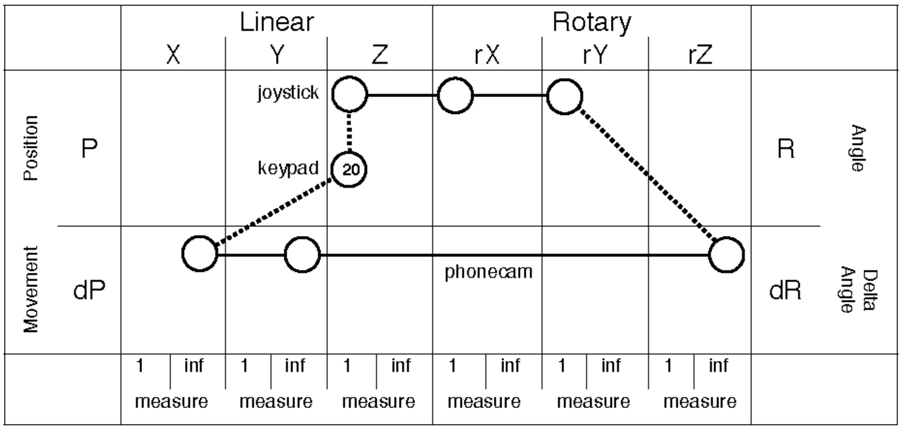

Fig. 3. Design space classification of our mobile phone prototype.

In terms of portability and dexterity, mobile phones are superior to PDAs in that they are smaller and they only require one-handed operation for both direct manipulation and text entry.
An advantage of invoking optical-flow processing on the phone rather than on the computer driving the display is user scalability; the interaction technique easily scales to a high number of users. A disadvantage however, is the high latency (about 200ms) that occurs when calculating the (x, y, θ) changes from successive images. Studies have shown that system lag has a multiplicative effect on task completion time [17]. Yet mobile computing trends indicate that in the not too distant future mobile phones will have the processing power necessary to create a fluid interaction experience.
In our implementation, we use Visual Codes (see Figure 4) to store the public display's Bluetooth address information thus enabling rapid association with the mobile phone. Users only need to take a picture of a Visual Code associated with the display and the phone will automatically connect to send (x, y, θ, text) information via Bluetooth. The same connection can be used to authenticate the user, to transfer information shown on the public display to the phone, and to store the current state of interaction. This creates a very low threshold of use and allows for highly serendipitous interactions.
In order to do Visual Code recognition and optical flow processing, our proposed device interactions require that users install special software on their mobile phone. The computer driving the public display requires a standard Bluetooth configuration to accept the mobile phone input. However, this software could potentially be installed during manufacturing, via the mobile phone network using over-the-air provisioning, or users could use Bluetooth to retrieve it directly from the computer driving the display. Fortunately, this software only needs to be installed once and therefore only slightly increases threshold of use for first time users.

4 Application Areas and Scenario

Using personal devices for large public displays can enable a rich set of applications, including games, interactive art, digital bulletin boards and advertising. In the public domain, games can provide entertainment to both active participants and to passive bystanders. For example, the PhotoPhone Entertainment Project [18] explored using mobile camera-enabled phones to play games in public places, like train stations and public squares, and to use available large public displays as output. However, our optical

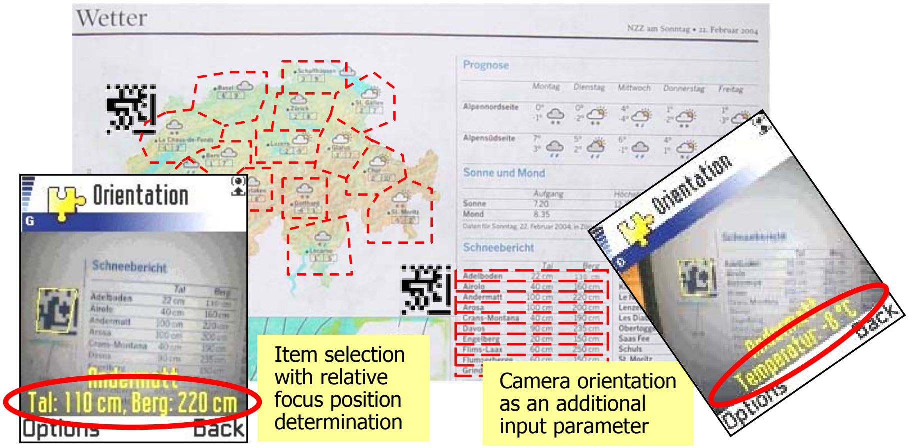

Fig. 4. Visual Codes can be used to augment physical media with camera accessible digital handles.

flow approach enables a more interactive gaming experience between users and the public display. The mobile phone's unique ID can enable a subscription-based gaming model, thus limiting access to paying customers only. Also, the phone can store game state in the event that gameplay is interrupted. Other application areas include interactive art installations [19] that often enrich our daily routines and encourage interaction with our familiar strangers [20] or public digital bulletin boards like those envisioned in the Plasma Poster Network [13] which provide an outlet for communities to share and disseminate news, announcements and ideas. While the most widely used application of large public displays to date is to display static information and advertisements, we typically have no method for automatically storing information from these displays. Instead we are forced to write the information down, or more commonly, just simply remember it. We can use mobile phones as a portable storage device to pull interesting information from the display and take it with us. However, in designing such applications, we should take heed when applying this personal device paradigm to essential application domains. For example, the concepts in this paper should not be applied to emergency services applications since we cannot guarantee that every person will have a phone camera.

To fully illustrate our interactive vision, here we present an interaction persona, scenario, and storyboard (shown in Figure 5).

Persona: Tom is an accountant in his late-twenties. He lives in New York and has a 30 minute commute on the subway to get from his apartment in Brooklyn to his office in Manhattan. Although he hates his job, he earns good money and it allows him to afford the lifestyle he enjoys. Tom is a young bachelor that is single with no pets and intends to stay that way at least until age 35. He's outgoing and likes to meet new people.

Scenario: Tom walks down the steps to his subway stop when he notices a familiar stranger that he refers to as ACE because he doesn't know his exact name. ACE has approximately the same commute schedule as Tom so they run into each other regularly. Tom doesn't particularly like ACE because he's really good at the multiplayer version of PacMan that is available on the large public display at the subway stop and he makes a lot of obnoxious fist movements whenever he beats somebody else. In fact, the reason Tom calls him ACE is because that's how he signs all of his high scores. ACE has just started a new game of PacMan and Tom decides that he's going to teach this guy a lesson. Tom pulls his phone from out of his pocket and joins the game by taking a picture of the Visual Code mounted next to the display. This code contains the Bluetooth information of the display and his phone automatically connects to the game. After 5 minutes of exciting maneuvers, Tom wins and the crowd of four people behind him erupts into cheers as ACE was finally beaten.

5 Additional Research Questions

There are a number of additional research questions that we are currently investigating.

– Is relative positioning using optical flow the best use of camera input? Another potential solution is using absolute position where we use a cursor on the phone's display to mark a selection. Then when a picture is taken the photo can be analyzed for the selection's on screen location.

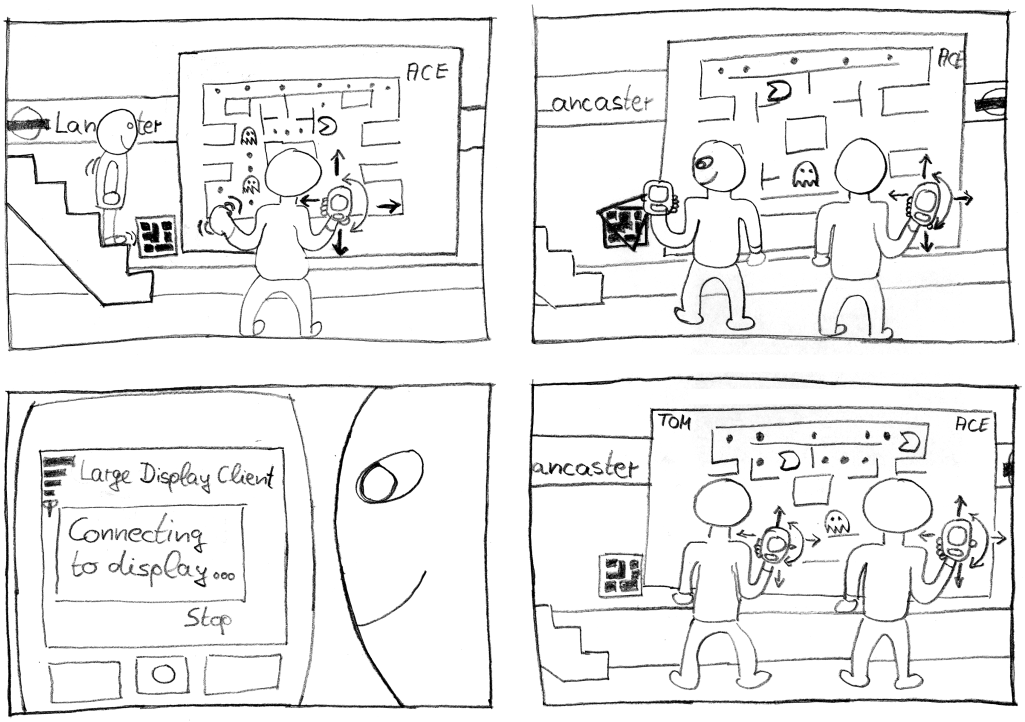

Fig. 5. A storyboard illustrating envisioned interactions between mobile phones and large public displays.

- How do we convey the affordance of an interactive public displays to potential users? Can the Visual Codes provide this affordance/information?
- Are there any interaction metaphors, other than cursor-based pointer input, that can mitigate the system lag we are experiencing in our current prototype?
- What are the consequences of choosing push versus pull data transfer techniques in large public display interactions?
- How do the aesthetics of the mobile phone itself including physical form, weight, size, material, and status symbol affect interactions with large public displays?
- What type of user is drawn to this kind of interaction? How do we analyse, evaluate and describe their interaction?
- How does this kind of interaction alter public space? What are the aesthetic, political, social and cultural implications [21]?
- How does distraction/interruption affect usage?

6 Related Work

The Remote Commander enables individuals to use a PDA to control the mouse and keyboard on a remote display using a touch sensitive display for mouse input and graffiti for text entry [11]. Similarly, PebblesDraw is a single display groupware system that captures input from PDAs and allows multiple users to draw simultaneously [11]. Both systems utilize touch sensitive displays on a PalmPilot and they require two hands for operation. Since the project focuses on semi-public display environments, it does not provide any mechanisms for serendipitous interactions.

The SharedNotes system [12] takes a different approach. Users employ a PDA to create and manipulate personal and public notes on a semi-public display. However, this solution uses replication of files across devices and synchronization of the modifications, and therefore limits the solution to a specific set of application domains.

Web Wall [14] and Digital Graffiti [13] allow users to post comments and to annotate a site, and therefore support more serendipitous interactions between a PDA and large public displays. User comments are constructed “off-line” and then applied to the system through a web-based interface. Users must use this web intermediary to issue commands instead of manipulating the GUI directly.

The VisionWand [22] system uses passive colored wands and stereo vision to provide multiple degrees of freedom. However, because the system uses camera tracking it is vulnerable to obstruction. Additionally, the system targets semi-public interactions and does not support multiple users.

Acknowledgements

The presented research is supported in part by the UK Engineering and Physical Science Research Council (EPSRC) under grant GR/S08848/01 ("Multi-Sensor Perceptive Interfaces in Wearable and Ubiquitous Computing") and as part of the Equator IRC (GR/N15986/01 – "Technological Innovation in Physical and Digital Life").

References

1. Patrick Baudisch, Nathaniel Good, and Paul Stewart. Focus plus context screens: combining display technology with visualization techniques. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 31–40. ACM Press, 2001.
2. E. A. Bier and S. Freeman. MMM: A user interface architecture for shared editors on a single screen. In *Proc. of the 4th Annual Symposium on User Interface Software and Technology (UIST'91)*, pages 79–86. ACM, 1991.
3. http://www.smarttech.com.
4. http://www.mimio.com.
5. J. Shaw and P. Seidler. Organic electronics: Introduction. *IBM Journal of Research and Development*, 45(1):3–10, 2001.
6. Paul Dietz and Darren Leigh. DiamondTouch: a multi-user touch technology. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 219–226. ACM Press, 2001.
7. Meredith Ringel, Henry Berg, Yuhui Jin, and Terry Winograd. Barehands: implement-free interaction with a wall-mounted display. In *CHI '01 extended abstracts on Human factors in computing systems*, pages 367–368. ACM Press, 2001.
8. G. Hardock. Design Issues for Line-Driven Text Editing/ Annotation Systems. In *Proc. of Graphics Interface '91*, pages 77–84. Morgan Kaufmann, 1991.
9. Francis Quek, David McNeill, Robert Bryll, Susan Duncan, Xin-Feng Ma, Cemil Kirbas, Karl E. McCullough, and Rashid Ansari. Multimodal human discourse: gesture and speech. *ACM Trans. Comput.-Hum. Interact.*, 9(3):171–193, 2002.
10. Scott Minneman, S. Joy Mountford, Natalie Jeremijenko, Krzysztof Wodiczko, Anthony Turner, and Mike Davis. Public information: documents, spectacles and the politics of public participation. In *CHI 98 conference summary on Human factors in computing systems*, pages 78–79. ACM Press, 1998.
11. Brad A. Myers, Herb Stiel, and Robert Gargiulo. Collaboration using multiple PDAs connected to a PC. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pages 285–294. ACM Press, 1998.
12. S. Greenberg and M. Boyle. Moving Between Personal Devices and Public Displays. *Workshop on Handheld CSCW, held at ACM Conference on Computer Supported Cooperative Work*, November 1998.
13. Scott Carter, Elizabeth Churchill, Laurent Denoue, Jonathan Helfman, and Les Nelson. Digital graffiti: public annotation of multimedia content. In *Extended abstracts of the 2004 conference on Human factors and computing systems*, pages 1207–1210. ACM Press, 2004.
14. Alois Ferscha and Simon Vogl. Pervasive Web Access via Public Communication Walls. In *Proceedings of the First International Conference on Pervasive Computing*, volume 2414 of *Lecture Notes in Computer Science*, pages 84–97, Zurich, August 2002. Springer-Verlag.
15. Stuart K. Card, Jock D. Mackinlay, and George G. Robertson. A morphological analysis of the design space of input devices. *ACM Trans. Inf. Syst.*, 9(2):99–122, 1991.
16. Michael Rohs and Beat Gfeller. Using camera-equipped mobile phones for interacting with real-world objects. In Alois Ferscha, Horst Hoertner, and Gabriele Kotsis, editors, *Advances in Pervasive Computing*, pages 265–271, Vienna, Austria, April 2004. Austrian Computer Society (OCG).
17. I. Scott MacKenzie and Colin Ware. Lag as a determinant of human performance in interactive systems. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 488–493. ACM Press, 1993.
18. Johan Thoresson. PhotoPhone entertainment. In *CHI '03 extended abstracts on Human factors in computing systems*, pages 896–897. ACM Press, 2003.

19. Ernest Edmonds, Greg Turner, and Linda Candy. Approaches to interactive art systems. In *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Austalasia and Southe East Asia*, pages 113–117. ACM Press, 2004.
20. Eric Paulos and Elizabeth Goodman. The familiar stranger: anxiety, comfort, and play in public places. In *Proceedings of the 2004 conference on Human factors in computing systems*, pages 223–230. ACM Press, 2004.
21. M. Jacobs. Preface to symposium proceedings. In *Outside In: Symposium on Emerging Expressions, Interventions and Participation in Public Space*, 2004.
22. Xiang Cao and Ravin Balakrishnan. VisionWand: interaction techniques for large displays using a passive wand tracked in 3D. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, pages 173–182. ACM Press, 2003.