Waving to a Touch Interface: Descriptive Field Study of a Multipurpose Multimodal Public Display

Marko Jurmu[1], Masaki Ogawa[1], Sebastian Boring[2], Jukka Riekki[3], Hideyuki Tokuda[1]

[1]Graduate School of Media and Governance
Keio University, Japan
marko.jurmu@ee.oulu.fi
{richie, hxt}@ht.sfc.keio.ac.jp

[2]HCC Group
Dept. of Computer Science
Univ. of Copenhagen, Denmark
sebastian.boring@diku.dk

[3]Intelligent Systems Group
Dept. of Computer Science and Engineering (CSE)
Univ. of Oulu, Finland
jukka.riekki@ee.oulu.fi

ABSTRACT

Multipurpose public displays are a promising platform, but more understanding is required in how users perceive and engage them. In this paper, we present and discuss results and findings from a two-day descriptive field trial with a multipurpose public display prototype called FluiD. Our main objective was to uncover emerging issues of interaction to inform future evaluations. During the field trial within a public research exhibition, people were able to freely interact with the prototype. Twenty-six persons filled out short questionnaires and gave free-form feedback. In addition, researchers in the vicinity of the display gathered observation data. Our main findings include the difficulties encountered with mid-air gesture commands, the lack of agency in case of larger interaction area, and the possibility for stepping out from the implicit-explicit continuum in the face of potential social conflicts.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems – *human factors.* H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *interaction styles, prototyping.*

General Terms

Design, Experimentation, Human Factors.

Keywords

Multipurpose public displays, multimodal interaction, proxemics, urban informatics, HCI, ubiquitous computing.

1. INTRODUCTION

Due to decreasing display hardware prices and positive market prospects of digital signage, interactive public displays are moving beyond the lab environments and are being deployed in real-world urban environments. Beyond digital signage, interactivity allows a shift from viewing experience to a user experience and enables public displays to be re-framed as an urban informatics tool, supporting citizens in their everyday activities through interactive services. While this development carries vast potential, it also raises tensions towards stakeholders, especially in the co-existence of digital signage and value-adding services on a single display platform. Within the research community, this new concept is called a *multipurpose public display* [22].

Due to the novelty of multipurpose public displays, their design, implementation and evaluation are challenging [2]. This culminates to following points: Lack of theoretical and practical *models*, variety of *objectives* for the services, and lack of *metrics*. Due to these reasons, the design space associated with multipurpose public displays is currently under debate and construction. It is also highly interdisciplinary, encompassing at least architecture, sociology and cognitive psychology besides urban informatics, HCI and ubiquitous computing. On the economic side, the recent discussion promotes infusion of informative content directly to advertisements [1] or addition of personalized services and application purchases [22]. Within the field of *pervasive advertising* (use of pervasive computing technologies for advertising purposes), Müller et al. present a comprehensive design space for public displays [19] spanning interaction phases [28], attention and motivation, mental models, as well as interaction modalities.

In this paper, we are presenting and discussing findings from a two-day field trial of a multipurpose, multimodal public display called FluiD. This display prototype incorporates principles from proxemics [12] to realize the implicit-explicit interaction continuum identified for interactive public displays [19], and accommodates both multipurpose and multimodal functionality. During the trial, people were freely allowed to interact with the display, get help from researchers on-site if necessary, and finally answer a short questionnaire with Likert-type questions as well as free-form feedback. Our main findings are summarized as follows:

- Communicating mid-air gestural commands accurately is challenging.
- Larger sensing area in front of the display leads people to doubt their agency over the public display.
- Without extrinsic motivation, people tend not to switch interaction modalities during interaction.
- Sensing area needs to accommodate a possibility for users to ‘opt-out’ from the interaction.

2. BACKGROUND

Contemporary experiments with interactive public displays have reported models and metrics adopted from other fields closely related to large digital displays, especially CSCW and more generally HCI. While these provide a solid overview of the display's effectiveness within the adopted model, researchers have recently called out for an emerging evaluation paradigm called *balanced triangulation* [2,11] where several methods are combined to investigate a research problem, and where the problem drives the selection of evaluation methods. This is seen as a necessary progress in the case of interactive public displays, since the interaction setting needs to be captured as a whole. Model-driven evaluation with high internal validity produces accurate results, providing a contribution that can be framed against the model itself. In case of public displays, we are now seeking for relevant questions and theoretical basis through holistic observations and so-called 'quasi-experiments' [23,26]. As a conclusion, deployments with high ecological validity drive the formulation of emerging questions and theories, with the aim of reaching multidisciplinary conceptual models and eventually model-based empiricism.

Our focus is a multipurpose, multimodal public display deployed in a public setting. Through this framing, our research background consists of the discussion regarding multipurpose functionality of public displays, the different interaction modalities and techniques incorporated within these displays, as well as the effects of space and place where the displays are deployed. For the sake of brevity, we shall consider three interaction modalities within this work: *Mobile*, *bodily* and *direct*. All of these modalities are supported by our prototype platform.

A significant body of HCI- and CSCW-oriented work has been conducted for evaluating public display interaction modalities and techniques. In the mobile case, several comparative studies of different techniques within this modality have been investigated. Ballagas reports a survey of different mobile input techniques [3]. Rukzio et al. compare three techniques for physical mobile interaction, albeit with a more generic focus of a smart object [27]. Boring reports a comparative study of cursor controls and a method based on direct video [5,6]. These techniques represent exemplary solutions for different phases of interaction with multipurpose public displays, for example pointing for selecting a target display, and live video for in-application manipulation.

In case of bodily interaction, the focus has been mostly in gaming and performative interactions, evidenced by contemporary game controllers such as MS Kinect or Nintendo Wii. The qualitative characteristics of this modality, namely performative nature and physical effort, strongly constrain the possible use cases. Some investigations have looked at the appropriateness of gestural interaction [10,25]. For public displays, mid-air gestures represent a potential interaction modality when their appropriateness and precision are accounted for. Design support can be found for example from the Kinect Human Interface Guidelines [16].

In the case of direct manipulation, a major focus has been in investigating the quantitative performance aspects of different (multi-) touch techniques. Through the current display-driven interaction paradigm of mobile devices, people are becoming increasingly accustomed to multi-touch interactions and gestures. From public display point of view, interesting non-functional aspects raised by direct manipulations are *territoriality* of the screen real estate and the resulting social conventions for turn taking [24].

Kurdyukova et al. [18] compared all of the abovementioned modalities in interactions with a personalized public display, and reported both quantitative and qualitative findings. Mobile modality is preferred for discrete and private operations such as login and logout, while browsing tasks on the public display controlled by a mobile interface are problematic due to constant macro attention shifts. Direct manipulation is feasible when the interaction area of the public display comfortably fits the person's viewing area. Bodily interaction is seen as supportive for use cases where fun, physical exercise and the act of performing are important. These findings are in line with our analysis above.

In a purely quantitative vein, Holleis et al. [13] provide extensions to the Keystroke-Level-Model (KLM) based on distributed interactions encompassing mobile devices and smart objects such as active posters. While KLM extensions provide a good basis for designing for the effectiveness of the user interface, it poorly accommodates other, non-functional issues such as social structure of the environment. For this reason, certain extension variables are purposefully categorized as 'additional noise', to be quantified based on each individual context.

The space and place dichotomization established by Dourish [9] gives a good starting point for framing the deployment locations of public displays. Aspects of *space* span issues related to space syntax, namely physical structuring of the space and the resulting movements of people that the structuring affords. Aspects of *place* encompass the social structure pertaining to that location, relying on the shared understanding of appropriateness within actions taken by people. Within this framing, aspects of space give information on what kind of physical structuring best affords people to stop for interactions with a public display, while aspects of place set a web of appropriateness constraints within which the interactions should fit.

3. FLUID PROTOTYPE

FluiD is a prototype platform of a multipurpose and multimodal public display. The hardware setup of the prototype consists of a 42" full HD display panel coupled with an IR frame capable of supporting up to six simultaneous touch points, and an MS Kinect depth camera positioned right above the display and observing the area in front of the display. The system is powered with a contemporary Mac-Mini computer running Windows 7 operating system.

The basis of the FluiD middleware is to utilize a depth camera in combination with face detection software to capture *operationalized proxemics* [12] in front of the display in real-time, and then feed this data to pluggable sliding window algorithms that produce higher level abstractions of human presence in front of the display. Both display behavior as well as interaction can be built on top of the depth sensing middleware.

For display behavior control, the system currently uses temporary IDs assigned for each tracked person, orientation of their head (towards / not towards display), real-time distance from the display as well as dwell time of each tracking. For interaction purposes, the sensing middleware implements mid-air sweeping gestures. This is realized by tracking the relative motion trajectories of elbow and wrist skeletal points of each tracked hand. In addition, single touch events and two-finger pinching gestures are supported through the touch panel.

4. FIELD TRIAL

The first application design carried out on top of the FluiD platform was for an annual Open Research Forum organized by Keio SFC in order to showcase the campus' research activities (see Figure 1). For this purpose, we designed a concept that would give visitors an overview of the platform's functionality, while at the same time providing enough interactive facilities to double as an observational field trial. The field trial was *descriptive* in nature, i.e. not designed with any specific hypotheses in mind, and focused in observing emergent behavior.

Figure 2 illustrates the interactive behavior designed on the prototype platform. First, the digital signage service is emulated through a carousel of digital images being scheduled in a static manner. As the depth camera detects a person's skeleton, the digital signage content is overlaid with two *service previews,* which appear from the sides of the display. Next, if the person is in front of the display *and* his/her face is oriented towards the display, i.e., a person is looking at the display in front, we display *service activation cues* within the previews. These visual cues serve to explain to the person how the service previews can be activated to make them full-screen and interactive. After activating either of the previews, the corresponding service slides into full view (in this case filling the 42" screen), and remains open for further interactions. In any of these phases, the user can quit the interaction by walking away or orienting him/herself away from the display. The previews are nevertheless shown as long as there are persons in front of the display. Finally, when a person leaves the display during an activated service, a short timer starts counting down and if there are no persons in front when the timer fires, the display goes back to digital signage mode.

The services implemented through the previews were as follows: The first one was a slot machine application, which upon winning provided a discount coupon to an imagined nearby store, claimable via a QR code. The slot machine was used since its animated content was assumed to catch the attention of the passers-by. The second application was a map-based view of recent tweets around the display. We assumed the map would be used by touch gestures, by relying on a priori knowledge that people have from operating maps in touch-enabled personal mobile devices.

It should be noted that this interaction design was implemented mainly for demonstration purposes as well as for observational data gathering, and should *not* be seen as a final design. There are several issues related to the interaction that are still left open in the design, including how to activate a new service on top of another one, how to browse services that can be activated etc. The current interaction design nevertheless provided an interesting observational basis for the research.

The campus event ran for two consecutive days, during which several hundred people visited the exhibition and examined the research booths. During the two days, we collected both Likert type data as well as free-form textual feedback from twenty-six users (seven female, one person preferred not to disclose gender information), ranging in age from 18 to 59 years (six persons preferred not to disclose age information). In addition, we conducted general diary-based observations, making notes of interaction behaviors exhibited by the persons. Due to the crowded nature of the event, organizing video recordings was challenging and was left out from the observation methods. We consider video-based analysis as an important part of future work with this prototype.

4.1 Likert-Type Questions

After interaction with the prototype, we asked persons to fill out a short questionnaire. The Likert-type questions asked in the questionnaire were as follows (range for all questions is 1=Strongly Disagree and 5=Strongly Agree).

Figure 1. Field trial setting. Main flow of people was along the arrows augmented to the figure.

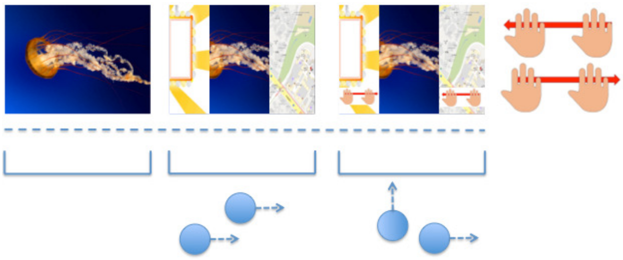

Figure 2. Phases of interaction. Continuous presence brings out the service previews, which are augmented with activation cues when correct orientation is detected. Right: Detail of the activation cues.

Question #1: *I understood that the display reacted to my presence.*
Question #2: *I understood that I can control the display.*
Question #3: *The visual cues for interaction (service activation cues) were easy to understand.*
Question #4: *Invoking the applications through gestures was easy.*
Question #5: *Despite the public setting, I was in control of the interaction.*
Question #6: *During the interaction, control of the display was shifting between me and others in front.*
Question #7: *Public displays are a potential platform for provisioning context-aware applications.*

4.2 Results

4.2.1 Likert-Type Data

For the Likert-type data, we report the mode of each question [15] along with notes regarding the deviation and tendency of answers.

For questions #1, #2 and #7, the mode was five, equaling 'strongly agree', with low deviation towards other scores. Results of question #7 thus indicate that after experimenting with the prototype, people were in general positive about the provisioning of interactive services through multipurpose public displays. Modes to questions #1 and #2 on the other hand indicate that people easily understood that the display's behavior was attributed to their presence in front. In this context it needs to be noted that the demonstration setup also included a secondary display where we demonstrated the view that the depth camera sees. When viewing this display, it was easy for people to associate one of the tracked skeletons to their own body [20].

As for the other questions, the deviation between scores was more pronounced, making the mode less representative of the overall result. For question #3, the mode was divided between 4 and 5 and some deviation for scores 3 and 2. This indicates that people understood the service activation cues quite well, i.e., that to activate the services, a gestural command needs to be performed. Certain bias is introduced here as well, as researchers were standing by to help users in performing the correct gesture when clear difficulties were observed.

Question #4 achieved the lowest mode, 3, along with the highest deviation. This is a clear indication of people having difficulties in translating the activation cue into a corresponding gestural command (see later sections for more discussion). The activation cues were still images describing an image of a hand, joined with an arrow to another hand. This was a static representation for a mid-air sweeping gesture used for activating the service previews (see Figure 2). Furthermore, as the gesture recognition was implemented based on the skeletal points of the depth camera view, it was sensitive regarding the angle of the elbow, requiring the wrist to be higher than the elbow while taking the wrist from one side of the elbow to the other.

Question #5 had a mode of 5 with an almost equal deviation between scores 2 to 5. This indicates that although the person occupying the immediate front of the display understood being in control, the other people constantly passing through the detection area of the depth camera caused many users to feel a loss of control. As the Kinect camera can quickly pick up a skeleton even from walking speed, there were rarely points in the interaction when the person in front was the only skeleton tracked by the depth camera.

Question #6 had a mode of 4 with high deviation among all scores between 1 and 4. This means that almost every participant felt that the control was shifting away from him or her during the interaction, mostly during the service activation phase. Besides from seeing other tracked skeletons in the secondary display, we suspect that a major contributor for this result was the difficulty in correctly translating the gestural command from the cue, resulting in a feeling that the system is trying to track someone else's gestures instead.

4.2.2 Free-Form Feedback

The free-form feedback provided by the questionnaire participants indicated a variety of issues, and the most interesting comments are summarized below.

"I'd like to get a notification of some kind when the operational authority moves to me. That would make controlling the display less stressful." (N/A, N/A) This indication is challenging to realize in practice, except in cases where the representation of the user is continuously shown, as in Looking Glass [20]. As the skeleton tracking effectively democratizes the gesture-based control, it's an interesting future work to consider how to enable turn-taking and negotiation of control in this setting.

"I couldn't understand that I can control the map application with touch." (Male, 18) This comment can have two reasons: Either the user was expecting a visual cue for touch-based operation, or the user could not leverage knowledge from map operations on mobile devices. We also suspect that in their interactions with the display, people were relying on the strategy of minimal cognitive effort, i.e., relying on the need to learn one of the interaction modalities, namely the one they started the interaction with.

"I felt bad about the reaction of the display. It is hard to understand, how the display can be returned back to the digital signage mode." (Female, 19) This user was surprised by non-warranted implicit interactivity, and instead of seeing it as a possibility for further interactions, wished she could revert the display back to its 'normal' operation mode. Allowing persons to 'opt-out' from the sensing is an interesting point of future research.

"I want to operate all the display contents without touch, if I started operating without touch." (Male, 21) This user was clearly unwilling to switch from one modality to another, and was displeased about the map not responding to gestural commands. This raises a host of interesting questions, including *should there be modality changes during interactions with a public display*, and if yes, *how should they be motivated and communicated to users?* As this functionality would introduce additional cognitive load, the switches need to be well motivated. Kurdyukova et al. report [18] that for example login and logout operations are preferred on a mobile device for privacy reasons, suggesting that another modality could dominate the interactions in between. Contemporary desktop systems are designed multimodal for the purpose that every user can find the preferred modalities and stick to them, so perhaps public displays should also aim for overlapping modalities, even when running a risk of lessened feel of control. Personalization in this sense would also mean discovering interaction modalities preferred by each person.

"When I want to obtain personalized information from a public display in a situation where the public display is big and I can't get closer to the display, then I think it's useful." (Female, 23) We assume that the personalization refers to the fact of having additional services on public displays. Gestures are here correctly seen as one possible solution for situations where the display's touch distance cannot be reached due to some physical obstacle such as protective glass.

4.2.3 Observations

As noted before, the difficulty of setting up video-based capturing meant that researchers needed to rely on simple diary-based methods to record emerging behavior in front of the display. In this section, the most pronounced behaviors are listed and shortly discussed, with a follow-up discussion in section five.

The first observed behavior was that through continuous sensing, users had difficulties in comprehending where the precise point of 'in-front-of' is in relation to the display. We attribute this behavior to a model where users assumed a relatively small spatial 'sweet spot' in front of the display, from where the mid-air gestures can be exclusively given. This assumption also quickly leads to lessened overall agency: As no confirmation for the sweet spot exists and skeletons of other people within the sensing area are tracked as well, people start to doubt the control they have over the display. This feeling is further amplified with incorrect executions of the activation gestures.

Second observation was the willingness of people to stick with a single interaction modality after the explicit interaction had commenced. In the case of the store coupon, people understood the mobile interaction by seeing the QR code on the display. In the case of the map, no explicit cues of modality change were given. Due to this, we observed a variety of strategies from people in trying to control the map through mid-air gestures. Zooming was attempted with pushing gestures and panning through sweeps. It is important to notice here that all attempts were effectively constrained within the interaction language learned by users during the service activation, namely the mid-air sweeps. This is an interesting observation that warrants further research in how interaction modalities are adopted unsupervised.

5. DISCUSSION

As the descriptive field study was not directly model-driven, we discuss findings related to several different issues. These findings are representative of an observation study, and serve to stimulate further discussions regarding interaction design for multipurpose, multimodal public displays.

Depth cameras such as Kinect allow continuous person tracking in front of the display. When coupled with face detection software, this combination allows presence and orientation to be tracked separately for each person. Several studies of public displays have shown how displays are enacted in groups of 2-5 persons, and that face-to-face communication takes place in front of the display. Instead of having fragmented sensory view of this behavior through face detections with web cameras and AVA-(anonymous video analytics) methods [14], depth cameras allow the display to accommodate this natural behavior of people to its own behavior, thus appearing more coherent towards users.

The display enactment process seems to follow the path of least cognitive effort. This would explain why the modality switch from gestural to touch-based interaction was not attempted by anyone: People did not even realize that after one modality has been adopted, another should be adopted as well. This resulted in people first activating the map view through gestures, waving for the map view in hope of a gesture control, getting desperate and increasing the speed and magnitude of the gestures, and finally giving up. This clearly indicates that *communicating the possible modalities as well as their correct use* is an interesting future work challenge for multipurpose and multimodal public displays. In addition, the willingness to utilize different modalities can be linked to the expected and perceived utility of the interactions, thus the finding bears similarities to that of Rukzio et al. [27].

An interesting finding was that due to the continuous sensing of the area around the front of the display, people did not have a clear model of *what does it mean to be in front of a display*. People made the assumption that commands to the display can be given only when the exact 'in front' location has been found, but since other people frequently occupied the sensing area as well, a clear model was inhibited. This can lead to lessened sense of agency, and needs to be tackled in future designs.

As an identified research challenge in [7], people will appropriate public displays in different ways than designers originally intended, and this should be acknowledged when designing the display's behavior. In our demonstration, one clear implication is that service activation should have also been implemented through touch. Investigating the question of designed interaction vs. enacted interaction within the context of multipurpose public displays remains future work. Content-wise, Kukka et al. report findings of discrepancies between services desired by users prior to usage versus the actual services used [17].

When transitioning from implicit to explicit interaction, some people expressed a wish to default the display back to its digital signage mode, i.e., to resign from the interaction. In some cases such as the Looking Glass [20], interaction is accommodated directly by having the signage mode show an empty template, but in this case people sensed that interaction by their presence was causing disturbance to the viewing experience of other people. This behavior has several implications. First is another reverberation of the finding that during continuous presence sensing with a depth camera, people don't have a clear understanding of 'in front of the display'. Nevertheless, people are willing to find this spot, as indicated in [4].

Second, the concept of *non-interaction* refers to situations where people discover interactive behavior in their immediate environment without any willingness to actually interact [21]. Since the interaction was built to start with an implicit, presence-based interaction through the depth camera, some people expressed a concern about transferring the display back to the digital signage state. This indicates that even though displays might provide advanced interactions for the users, sometimes people just want to passively observe the displays. Finally, the conflicts caused by implicit interaction to the display's behavior can be linked to conflicts of *content* and *pace*, identified by Dix et al. [8]. For other persons just willing to observe the digital signage, a person causing the previews to appear can be a nuisance, thus presenting a possibility for a social conflict between people in the vicinity of the display. In this case, the display can also be seen counterproductive from the viewpoint of shared understanding of appropriateness, i.e., the concept of place.

People tried very discrete swiping gestures in order to maintain with the appropriateness of the place and not to disturb people around them. Immediately after starting to give the gesture controls, people assumed a performative role, accepting that the rest of the people around were observing both the display and the person interacting with it as a combined performance. This should be taken into account in design regardless of the chosen modality, and it is especially pronounced in gestural interactions. Finally, with only a crude visual sign, people were having difficulties gauging the magnitude of the gestures to be given. In nearly all the cases, the researchers needed to intervene to guide people in giving the correct gesture. This clearly implies the difficulties in accurately communicating gesture commands, and that to give precise information people should for example see a video of an example person performing the desired gesture. We assume that this rationale is also behind the short videos shown in the OS X settings section related to the track pad controls: As the user sees the exact execution of the gesture by a real human hand from the video, it is much easier to mimic in actual use.

Depth cameras are well suitable for realizing presence-based functionality in public displays, but their suitability for outdoor conditions remains unknown. If we establish that hardware for outdoor deployments should be *affordable*, *weatherproof* and *foolproof*, we can analyze that for example MS Kinect is currently affordable and can be also made foolproof with appropriate casing, but its performance especially in varying lighting conditions and direct sunlight remains a challenge. While the FluiD prototype was not directly tested in outdoor conditions, prior research has demonstrated difficulties in optical sensing in the face of varying sunlight levels.

6. SUMMARY

In this paper, we contribute to the emerging field of multipurpose public displays with a descriptive field study, which prioritized external and ecological validity and featured a fully implemented prototype. Through this setting, issues related to user experience and display performance that are challenging to discover in lab studies were identified. These findings can be seen as detailing the concept of *interaction blindness* established in [22]. We acknowledge that the findings uncovered are to a certain degree mentioned also in other works related to interactive public displays. However, we want to emphasize the need to discuss these issues also in the context of multipurpose public displays, and see this work as a necessary opening for investigating emerging interaction models such as proxemics in this setting.

7. ACKNOWLEDGMENTS

First author would like to thank the Academy of Finland for financial support during his research visit to Keio University. This research was also partly supported by Japan National Institute of Information and Communications Technology (NICT).

8. REFERENCES

[1] Alt F, Schmidt A & Müller J. 2012. Advertising on public display networks. *IEEE Computer 45(5): 50-56.*

[2] Alt F, Schneegaß S, Schmidt A, Müller J & Memarovic N. 2012. How to evaluate public displays. *Proc. PerDis'12*.

[3] Ballagas R, Borchers J, Rohs M & Sheridan JG. 2006. The smart phone: A ubiquitous input device. *IEEE Pervasive Computing 5(1): 70-77.*

[4] Beyer G, Alt F, Müller J, Schmidt A, Isakovic K, Klose S, Schiewe M, Haulsen I. 2011. Audience behavior around large interactive cylindrical screens. *Proc. CHI 2011,* 1021-1030.

[5] Boring S, Baur D, Butz A, Gustafson S & Baudisch P. 2010. Touch projector: Mobile interaction through video. *Proc. CHI 2010*, 2287-2296.

[6] Boring S, Jurmu M & Butz A. 2009. Scroll, tilt or move it: Using mobile phones to continuously control pointers on large public displays. *Proc. OzCHI 2009*, 161-168.

[7] Dalsgaard P & Halskov K. 2010. Designing urban media façades: Cases and challenges. *Proc. CHI 2010,* 2277-2286.

[8] Dix A & Sas C. 2008. Public displays and private devices: A design space analysis. *Proc. CHI 2008 Workshops.*

[9] Dourish P. 2006. Re-Space-ing Place: "Place" and "space" ten years on. *Proc. CSCW 2006*, 299-308.

[10] Grandhi SA, Joue G & Mittelberg I. 2012. To move or to remove? A human-centric approach to understanding gesture interpretation. *Proc. DIS 2012, 582-591.*

[11] Greenberg S & Buxton B. 2008. Usability evaluation considered harmful (some of the time). *Proc. CHI 2008*, 111-120.

[12] Greenberg S, Marquardt N, Ballendat T, Diaz-Marino R & Wang M. 2011. Proxemic interactions: The new ubicomp? *ACM Interactions Magazine, Jan/Feb 2011*, 42-50.

[13] Holleis P, Otto F, Hußmann H & Schmidt A. 2007. Keystroke-level model for advanced mobile phone interaction. *Proc. CHI 2007,* 1505-1514.

[14] Intel AVA solutions. URL: http://tinyurl.com/cb9aape. Accessed: 2013-04-15.

[15] Kaptein M & Robertson J. 2012. Rethinking statistical analysis methods for CHI. *Proc. CHI 2012*, 1105-1113.

[16] Kinect for Windows Human Interface Guidelines v1.7.0, URL: http://msdn.microsoft.com/en-us/library/jj663791.aspx. Accessed: 2013-04-12.

[17] Kukka H, Kostakos V, Ojala T, Ylipulli J, Suopajärvi T, Jurmu M & Hosio S. 2013. This is not classified: Everyday information seeking and encountering in smart urban spaces. *Springer PUC Journal 17(1): 15-27.*

[18] Kurdyukova E, Obaid M & André E. 2012. Direct, bodily or mobile interaction? Comparing interaction techniques for personalized public displays. *Proc. MUM 2012,* #44.

[19] Müller J, Alt F, Michelis D & Schmidt A. 2010. Requirements and design space for interactive public displays. *Proc. Multimedia 2010 (MM'10)*, 1285-1294.

[20] Müller J, Walter R, Bailly G, Nischt M & Alt F. 2012. Looking glass: A field study on noticing interactivity of a shop window. *Proc. CHI 2012,* 297-306.

[21] O'Hara K. 2010. Interactivity and non-interactivity on tabletops. *Proc. CHI 2010,* 2611-2614.

[22] Ojala T, Kostakos V, Kukka H, Heikkinen T, Lindén T, Jurmu M, Hosio S, Kruger F & Zanni D. 2012. Multipurpose interactive public displays in the wild: Three years later. *IEEE Computer 45(5): 42-49.*

[23] Oulasvirta A. 2012. Rethinking experimental designs for field evaluations. *IEEE Pervasive Computing 11(4): 60-67.*

[24] Peltonen P, Kurvinen E, Salovaara A, Jacucci G, Ilmonen T, Evans J, Oulasvirta A & Saarikko P. 2008. It's mine, don't touch! Interactions at a large multi-touch display in a city centre. *Proc. CHI 2008,* 1285-1294.

[25] Rico J & Brewster S. 2010. Usable gestures for mobile interfaces: Evaluating social acceptability. *Proc. CHI 2010, 887-896.*

[26] Rogers Y. 2011. Interaction design gone wild: Striving for wild theory. *ACM Interactions Magazine, Jul/Aug*, 58-62.

[27] Rukzio E, Leichtenstern K, Callaghan V, Holleis P, Schmidt A & Chin J. 2006. An experimental comparison of physical mobile interaction techniques: Touching, pointing and scanning. *Proc. UbiComp 2006, LNCS 4206,* 87-104.

[28] Vogel D & Balakhrishnan R. 2004. Interactive public ambient displays: Transitioning from implicit to explicit, public to personal, interaction with multiple users. *Proc. UIST 2004*, 137-146.