How to evaluate Public Displays

Florian Alt[1], Stefan Schneegaß[1], Albrecht Schmidt[1],
Jörg Müller[2], Nemanja Memarovic[3]

[1]VIS, University of Stuttgart
{firstname.lastname}@vis.uni-stuttgart.de

[2]Telekom Innovation Laboratories, TU Berlin
joerg.mueller@tu-berlin.de

[3]University of Lugano
nemanja.memarovic@usi.ch

ABSTRACT

After years in the lab, interactive public displays are finding their way into public spaces, shop windows, and public institutions. They are equipped with a multitude of sensors as well as (multi-) touch surfaces allowing not only the audience to be sensed, but also their effectiveness to be measured. The lack of generally accepted design guidelines for public displays and the fact that there are many different objectives (e.g., increasing attention, optimizing interaction times, finding the best interaction technique) make it a challenging task to pick the most suitable evaluation method. Based on a literature survey and our own experiences, this paper provides an overview of study types, paradigms, and methods for evaluation both in the lab and in the real world. Following a discussion of design challenges, we provide a set of guidelines for researchers and practitioners alike to be applied when evaluating public displays.

Keywords

Public Displays, Evaluation, Digital Signage, Methods

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Evaluation/methodology*

General Terms

Experimentation, Human Factors

1. INTRODUCTION

Interactive public displays are leaving the labs and are being deployed in many places. Nowadays, they permeate public spaces, shop windows as well as malls, workspaces, and public institutions and are equipped with sensors, such as cameras, that enable presence and motion sensing. At the same time, new (consumer) devices and software enter the market (e.g., the Microsoft Kinect),

PerDis 2012, June 04 - 05 2012, Porto, Portugal.

Figure 1: Studying public displays in the lab (left) / field (right)

hence providing opportunities for researchers to create novel interaction techniques. Hence, there is an emerging need for both practitioners and researchers, to understand how to best evaluate public displays with regard to effectiveness, audience behavior, user experience and acceptance, and social as well as privacy impacts.

Since today no commonly accepted guidelines exists as to how (applications for) public displays should be designed, the evaluation of these is both crucial and challenging for several reasons. First, *display deployments are often opportunistic*. As new infrastructure or real estate is being created, the premises are often augmented with public displays, having only little knowledge of the audience. Second, simulations of the environment a display is deployed in are difficult, as there are *no (dynamic) models* yet (e.g., of the stream of visitors passing through a pedestrian area). As a result, evaluation has to be conducted in context (both in the real world and in the lab). Third, there is *not one single goal* that public displays (or their content) try to achieve. Ads most likely strive for maximizing attention, interactive games may want to create an engaging experience, informative applications (e.g., a public transport schedule) may aim at maximizing usability, and some displays may be deployed to show warnings to passers-by or support the fast evacuation of a building. Fourth, *measuring the effectiveness* of a display is difficult. Compared to the internet, it is often not possible to monitor user interaction, but sensors might in the future allow to extract richer information based on the interaction. This, however, might raise privacy concerns (e.g., when using a camera in public space), hence determining and restricting the means for evaluation.

In order to tackle these challenges, we set out to provide guidelines for evaluating public displays. Our work is grounded in a comprehensive literature survey, based on which we identified common study types, paradigms, and methods including their respective advantages and disadvantages. We discuss and validate them, ultimately deriving guidelines that can help researchers and practitioners to choose an evaluation method for their public display.

2. RELATED WORK

Our research is based on an extensive literature review with the goal to identify methods and tools that are used to evaluate public display. As of January 2012, 522 papers can be found in the ACM Digital Library that are concerned with public displays or digital signage. Most of these papers evaluate their concepts and deployments. Even nowadays, more than 30 years after *Hole-in-Space*[1], one of the first public display installation, neither design guidelines for public displays exists, that cover a broad spectrum of systems and applications, nor do generally accepted evaluation guidelines. However, several ideas have been around in recent years.

Cheverst *et al.* [11] reported on challenges of evaluating situated displays deployed in a community setting. Storz *et al.* published lessons learned from the deployment of the eCampus public display network [36] that provides useful information for informing the design of public display (networks) but only little information with regard to evaluation. Mankoff *et al.* [20] looked at the evaluation of ambient displays, focusing mainly on effectiveness and usability. Starting from Nielsen's usability heuristics, they created a modified set to be used for the evaluation. Finally, Matthews *et al.* [21] used activity theory to evaluate peripheral displays. They identified an initial set of evaluation metrics (appeal, learnability, awareness, effects of breakdowns, and distraction), that vary depending on the importance of the display, but do not focus on evaluation methods.

Though many research papers provide useful lessons learned or recommendations based on their findings (e.g., [15, 36]), most previous work either focusses on a rather specific application domain (community/situated displays, ambient displays), draws from findings of their deployment(s) only, or treats evaluation only on the side / on a high level. To overcome these limitations, we base our findings on a comprehensive literature review, identifying research questions, research types, and approaches to research and methods used in public display research.

3. RESEARCH QUESTIONS

In the following we describe the most popular questions researchers tried to answer during their evaluations. Note, that many projects tackled sub questions to these (e.g., numbers of glances or interactions as subquestions to display effectiveness).

- **Audience Behavior**: A major focus of research is how the audience behaves around a display. Prior work identified effects such as the honeypot [9, 25], where interacting users attract more users, the sweet spot [6], which is a preferred position in front of the screen, or the landing zone [25], where people only realize that a display is interactive after passing it. Though mostly conducted in the real world, we also found examples of evaluation in the lab [6]. Audience behavior can be assessed by observations [13] and log data [32].

- **User Experience**: User experience describes the overall experience of interacting with a display. This is important in public display research as high user experience may lead to a higher motivation to use the application and possibly take in the user for as long as possible. Researchers examined different interaction techniques and their effect on user experience based on (standardized) questionnaires, e.g., interactions mediated through a mobile device [4] or direct touch [31].

- **User Acceptance**: Often used in early stages of the development process, the user acceptance investigates users' motives and incentives to interact with a display. It can be assessed qualitatively based on subjective feedback, e.g., in focus groups to collect the target group's view and concerns [11] or quantitatively based on questionnaires [23].

- **User Performance**: Measuring effectiveness from a user's perspective is often done when evaluating novel interaction techniques, e.g., based on a camera, mobile phones [4] or direct touch [12]. User performance can be quantified by measuring task completion times and error rates [4, 12, 33].

- **Display Effectiveness**: Interesting especially from an economic perspective, several studies aimed at measuring the effectiveness of public displays (e.g., how many people passed-by a display [24], how many looked at it [15, 26] and how many started interacting with it [25].

- **Privacy**: Some projects aimed at understanding the users' concerns on privacy. Alt *et al.* look at how mobile phone can overcome privacy issues [3], whereas Shoemaker *et al.* explored an interaction technique that allows private information to be shown on a shared display [35].

- **Social Impact**: Finally, social impact has been subject to research. Evaluations looked at how display applications could foster social interaction [22], how users engage in social interaction [25], which types of communities form around public displays [1, 11], and which social effects occur [8].

Table 1 provides a summary of research projects classified by the tackled research questions. Furthermore, we distinguished whether the evaluation was conducted prior to creating a prototype by asking users or running an ethnographic study, or whether a prototype was evaluated in a lab study, field study, or in the context of a deployment (see section *4.3 Paradigms*).

4. APPROACHES TO RESEARCH

The following section provides an overview of different study types and paradigms we found throughout the literature review. They will be briefly explained, followed by a discussion of their advantages and disadvantages. As different notions are used throughout literature, we adhere to the notions by Nielsen [27] and Lazar [19].

4.1 Study Types

We categorized related work according to the three principle types of research: descriptive, relational, and experimental research.

4.1.1 Descriptive Studies

Descriptive research aims at merely describing what is going on in a certain situation. This description can be qualitative (e.g., observations [26], interviews [1], focus groups [2]) as well as quantitative (e.g., photo logs [1], etc.). Descriptive research is the only of the three types of research where variables do not need to variate, e.g., multiple prototypes to be compared are not needed. It is striking that the vast majority of public display research includes descriptive methods. Good examples for descriptive research are the CityWall [30] and Worlds of Information [17]. In both studies, a single prototype is deployed and user behavior around the display is being measured, analyzed, and described (e.g., observations, questionnaires, etc.). One major benefit of descriptive research is that no hypotheses to be tested are needed, and therefore no general theory is necessary that the hypotheses can be derived from. It is especially suited for a research field that is in an early phase and does not possess general theories yet, like public displays. However, descriptive studies of single prototypes create isolated spots

[1]Hole-in-Space: `http://www.ecafe.com/getty/HIS/`

	Asking Users	Ethnography	Lab Study	Field Study	Deployment-based Research
Audience Behavior	Display Blindness [26], Public Notice Areas [1]	Public Notice Areas [1]	Cylindrical Displays [6]	Looking Glass [25]	Hermes [10], eCampus [36], MobiDIC [26], UBI-Oulu [29]
User Experience			Sweep and Point & Shoot [4], RayCasting [38]	Digifieds [3]	Nnub [31]
User Acceptance	Campus Coffee [11]	Public Notice Areas [1]	Mobile Contextual Displays [2], Campus Coffee [11]	AmIQuin [23]	
User Performance			TouchProjector [7], Push-and-Pop [12], PhoneTouch [33], Spotlight [18]	LightSpace [39]	MobiDIC [26]
Effectiveness	Display Blindness [26]	When does the public look at displays? [15]	Looking Glass [25], Interactive Public Ambient Displays [37]	Reflective Signs [24], Looking Glass [25], Digifieds [3]	
Privacy			Digifieds [3]	Digifieds [3], Single Display Privacyware [35]	
Social Impact		Public Notice Areas [1]		FunSquare [22], CityWall [30], Worlds of Information [17]	Nnub [31]

Table 1: Selected examples for the evaluation of public display (applications) based on different objectives.

in the design space of public displays without relations to other studies. This makes it difficult to compare results and designs, and ultimately to understand the structure of the entire design space. Hence, in the long run the progress of public display research may be hindered if the pure focus on descriptive studies persists.

4.1.2 Relational Studies

Relational research aims at showing that two or more variables covariate, i.e., that there is a relation between two or more factors. In particular, by relational research, no causality can be attributed, i.e., it is unknown which of the variables causes the other to change, or whether both depend on a third, unknown variable. Relational studies are rare in public display research, in particular because not many relationships between different dependent variables are considered to be interesting. Exceptions include ReflectiveSigns [24], where it is shown that the time people spend looking at public display content does not correlate with people's stated interest.

4.1.3 Experimental Studies

Experimental research aims at determining causality, i.e., that one variable directly influences another variable. Experiments possess the following characteristics [19]: They are based on hypotheses, there are at least two conditions, the dependent variables are measured quantitatively and analyzed through statistical significance tests, they are designed to remove biases, and they are replicable. Experiments aim to refute (or fail to refute) hypotheses, and these hypotheses are usually derived from theories. Therefore, entire theories can be refuted by experiments. Experiments can be conducted in the laboratory (more control) or in the field (higher ecological validity). Whereas a lot of experimental studies conducted in the lab can be found (e.g., in order to evaluate user performance with regard to a novel interaction technique [4, 7, 12, 18]), real-world experiments are rare in public display research, partially because no coherent theories of public displays exist. Another reason is that for an experiment multiple variations of a prototype need to be developed, making such experiments in the real world particularly time-consuming. One example for a field experiment is Looking Glass [25], where the influence of different interactivity cues on how many people interact with the displays were tested.

4.2 Research Phases

Research methods can be used in different phases during a project. In the beginning, there is usually a phase of *requirements analysis*, and there is no prototype yet [1, 2]. Typical methods used during requirements analysis encompass ethnography and techniques to ask users, such as focus groups, interviews, or questionnaires. When the first prototypes are developed, usually a phase of *formative studies* follows. These are intended to give direction to the design process and find properties and problems of the current prototype. Formative techniques include deployment-based research, lab studies, and techniques to ask users [3] (see below). When the final prototype exists, *summative studies* come into play. They usually try to make some conclusion about the final prototype, e.g., by comparing it to a baseline in a lab [6, 25] or field study [25].

4.3 Paradigms

We identified 5 evaluation paradigms either used to inform the design of a prototype (ethnography, asking users) or to evaluate a prototype (lab study, field study, deployment-based research).

4.3.1 Ethnography

In ethnographic studies, usually certain (social) settings are being investigated without intervention, e.g., without deploying a prototype. Ethnographic studies have been used to inform the design of public display systems. Alt *et al.* conducted an ethnographic study to assess the motivation and intentions of stakeholders as well as social impact [1]. Huang *et al.* investigated current practices around public displays [15]. The advantage of ethnographic studies is, that they provide valuable information that could be used to enhance the design of a public display (system), but often require a lot of effort.

4.3.2 Asking Users

Similar to ethnography, users can be asked by interviews, questionnaires and focus groups. However, hypothetic questions usually lead to poor answers (e.g., *"What functionality would you expect from a future version of this app?"*). Therefore, users can also be prompted with a prototype, and these methods can be combined with lab or field studies, ethnographies, or longer deployments.

4.3.3 Lab Study

Lab studies are aimed at evaluating a system, e.g., a novel interaction techniques, within a controlled environment. Lab studies can be descriptive, relational, or experimental. During the lab study both qualitative data (e.g., interviews [25], observations [6]) and quantitative data (task completion time, error rates [4, 7]) can be collected. The advantage of lab studies is that external influences (other passers-by, environmental conditions) can be minimized and (sensitive) equipment for proper measurements (cameras, sensors) that would be difficult to deploy in public can be used [6]. The disadvantage of lab studies is, that they may provide only low ecological validity and that the dynamics of the real world are excluded.

4.3.4 Field Study

In contrast to lab studies, field studies aim at evaluating a system or technique in a (semi-) public setting. In contrast to deployment-based research, they are rather short (days to months) and focussed towards a single research question. Similar to the lab, they may be descriptive, relational, or experimental. Data collection in the field is often cumbersome and time-consuming, as automation may be difficult due to privacy issues (e.g., recording video in a public space). The advantage, however, is that a high ecologic validity of the data can be assumed. Furthermore, there are aspects, such as effectiveness [24], social effects [22], audience behavior [25], and privacy implications [3] that are almost impossible to measure in the lab. The disadvantage of field studies is, that they are usually complex due to the high number of potential influences, and require a tremendous effort in preparation (finding a suitable place, legal issues, etc.). Traditional methods are observations and logging.

4.3.5 Deployment-based Research

Deployment-based research is a kind of action research that introduces technology (e.g., public displays) into a social setting (e.g., a city), to address some research questions derived from theory [25] [3]. User feedback and involvement is then obtained, and in an iterative process, the deployment is improved. At the same time, this data is used to build and refine theory, which in turn generates new research questions that can be addressed through changes in the deployment. In contrast to field studies, deployments are really integrated into the everyday life of their users. The difference with ethnography is, that there the researchers do not intervene by deploying a prototype.

There is a continuum from cultural probes over technology probes to deployments [16]. *Cultural probes* support users with things like cameras to document their lives, while *technology probes* introduce small prototypes in order to understand a given domain, sometimes without the scientific rigor introduced in experiments. Only *deployments*, however, really become permanent useful artifacts in everyday life. Deployments enable researchers to investigate longitudinal effects of use that cannot be investigated with other means. They are also the only method which can really get rid of the novelty factor which influences other kinds of studies. On the other hand, the maintenance of such deployments binds considerable resources. Examples of deployment based research are Hermes [10], the Wray display [11], eCampus [36], and UBIOulu [29].

5. METHODS & TOOLS

Classical data collection methods are interviews, questionnaires, focus groups, observations, and logging.

Interviews are often semi-structured, i.e., the interviewer follows some pre-defined guidelines, but would dig deeper if he discovers interesting findings. Interviews can also be conducted in context (e.g., shortly after the subject used the system). Interviews are a powerful method, to assess the user's view (e.g., concerns, problems, opinions [3]).

Questionnaires are a useful method for the quantitative evaluation of public displays. Questionnaires can be *standardized*, hence allowing different system to be assessed and compared with regard to, e.g., usability (System Usability Scale [5]), user experience (AttrakDiff [2]), or task load (NasaTLX [14]). On the other hand, questionnaires can be *customized* and used to ask the user about his personal view. Questionnaires have been used in many of the reviewed projects in order to assess, e.g., user experience [6] or performance and the users' view [4]. Focus groups are used in early stages of the design process (usually as soon as an early prototype exists), to discuss it with people of the potential target group.

Focus groups are usually run with 5-8 people and sessions last about 1-2 hours, including a demonstration (and hands-on) of the system, followed by a discussion. The discussion is led by one of the researchers based on guidelines, trying to answer important research questions. In [2] a focus group was used to assess the user's view on contextual mobile displays and [11] used focus groups to discuss multiple system designs with different degrees of interaction. The advantage is that feedback (e.g., on potential issues) can be provided in very early stages of the design process. On the other hand, opinion leaders in the group may prevent some people from stating their (contrasting) view.

Observations are most powerful when it comes to (post-hoc) assessing audience behavior [6], but also effectiveness [25] and social impact [17]. In general, two forms can be distinguished. During *automated observations* users are observed by cameras, installed in a fix location (potentially filming both the screen and the viewer) [6]. The video footage can be analyzed post-hoc using methods from computer vision, such as shape or movement detection, eye recognition, or manual annotation and coding [6]. When conducting *manual observations*, data can be gathered, e.g., by taking field notes or pictures and videos from both the subject and the display [25]. In this case, the observers usually hide in a location from which both the screen and the interacting persons can be seen [3]. The advantage of the method is that users behave most natural as they are not aware of being under investigation, making the findings ecologically highly valid. On the downside, video-based observation may compromise the subjects' privacy and conclusions on why users behaved in a certain way may be very difficult. Because of this, observations are often combined with (post-hoc) interviews.

User interaction can also be *logged* (e.g., time to perform a task, number of clicks) during the study and be post-hoc analyzed. This method is especially helpful especially when conducted over a longer period of time. Means for logging include all types of (optical) sensors that allow motion [6], eye gaze [34], presence [25], or user interaction [3] to be assessed. Logging has been used in many observations, e.g., in order to assess trajectories [6], time of day [3] and type of content [34] with which interaction occurred. The advantage of log files is that a lot of data can be gathered with literally no effort. The disadvantage is that logging often concerns personal data, which may be an issue in a public setting.

[2] `http://www.attrakdiff.de/en/Home/`

	Field Study			Lab Study		
	avg.	med.	sd.	avg.	med.	sd.
Interview	26.9	15.0	29.5	16.0	12.0	2.9
Observation	35.5	36.0	12.7	14.3	6.0	5.9
Questionnaire	38.4	32.0	37.4	15.0	12.5	5.0

Table 2: Average number of participants for different methods. (Note that only within-subject designs were considered.)

6. ASSESSING VALIDITY

Our literature review revealed a broad set of study types and methods that have been used by researchers during their evaluations. We found that most of the projects used various methods in parallel showing the complex nature of public display evaluations. In the past 4 years we have carried out more than 30 studies including public displays in the context of various (large-scale) research projects, which allowed us to validate the methods described above. In the following we provide a high-level summary of both our experiences and findings from literature.

6.1 Internal, External & Ecological Validity

Most studies can be criticized as not exercising sufficient control over confounding variables (internal validity), not generalizing to other settings and situations (external validity), or not testing a realistic situation (ecological validity). Internal, external, and ecological validity can usually not be achieved at the same time. Rather, studies must often sacrifice one or two of them to improve the third. Often, in HCI, internal validity is prioritized, leading to highly controlled lab studies with rather low ecological validity. In contrast, public displays are, by their very nature, a very social phenomenon. Behavior in the public space may be very different than what is expected from lab studies [15]. Therefore, ecological validity is often prioritized over internal and external validity.

6.2 Study Size

An often occurring questions is that for sample size in order to be able to draw well-founded conclusions. Though there is certainly not an ultimate answer, we analyzed the number of participants from more than 80 studies. Table 2 provides a rule of thumb as to which sample sizes may be most appropriate for which method.

7. GUIDELINES

Obviously, researchers first of all need to be clear about their research questions and should then decide whether to run a descriptive, relational, or experimental study. Besides, we derived the following set of guidelines to help informing the design of public display studies.

Choose your focus on internal, external, or ecological validity: Often, control, generalizability, and realism may not all be achieved at the same time. It is important to make clear which kind of validity is in focus and which validity is partially sacrificed. In public displays, often ecological validity is valued high, but it has to be clear how internal and external validity are reduced and measures may be taken to improve these (e.g., randomization to decrease the influence of confounding variables).

Consider the impact of the content: Public display research is not possible without content, but the impact of content and other factors on usage is indistinguishable [36]. Thus, every study is at risk that the results are only valid for the particular kind of content tested. Testing different contents might help.

Understand the users: Public displays may have widely different users at different locations and different times. In [25], for example, school children in the morning behaved very differently from drunk people at night. Furthermore, it is generally advisable to practice triangulation and combine different methods towards a common research question. Experiments can often be accompanied by descriptive research, and quantitative by qualitative measures. The deployment in public space also introduces ethical issues, and anonymization of any required data will usually be necessary.

Check for common problems: One of the most common problems with public displays is that they do not receive a lot of attention. This happens as passers-by do not expect them to be interactive [25] nor they expect to find interesting content [26]. Engaging people that actively promote displays might help to raise audience awareness of a display's interactive capabilities [28]. On the other hand creating customized content that reflects the users' interests is expensive [36]. A possible solution can be seen in autopoiesic content, i.e., self-generated content [22].

However, the effect of display blindness might be caused by other factors imposed by the environment. For example, in the Looking Glass study [25] at one location very few people looked at the display because they turned their head in the opposite direction looking down a road as they approached. In another example Storz *et al.* [36] point out technical challenges. Their initial deployment was much shorter than envisioned because their hardware was not built for the specific setting that had a high volume of diesel fumes, which in turn caused the projectors to shut down automatically. Besides influencing the hardware, the setting can also influence the user acceptance: previous research showed that placing a display in a leisure-oriented environment, e.g., in a swimming hall, can attract more attention than in a business-oriented, e.g., municipal service center [28]. Hence, understanding the environment and its challenges before the deployment is crucial.

8. CONCLUSION

In this paper we presented findings from an extensive literature and from our own experience in evaluating public displays. We extracted the most common research question, research types, and methods and presented selected examples for the reader's reference. Based on our findings, we discussed common design challenges and extracted a set of universally applicable guidelines, hence accounting for the need of many researchers to design studies for evaluating public displays.

9. ACKNOWLEDGEMENTS

We thank Nigel Davies for his useful comments on this research. This work was supported by the European Union 7th framework programme under grant agreements no. 215893 and 244011.

10. REFERENCES

[1] Alt, F., Memarovic, N., Elhart, I., Bial, D., Schmidt, A., Langheinrich, M., Harboe, G., and Huang, E. Designing shared public display networks – implications from today's paper-based notice areas. In *Proc. of Pervasive 2011* (2011).

[2] Alt, F., Schmidt, A., and Evers, C. Mobile contextual display systems. In *Adj. Proc. of HotMobile'09* (2009).

[3] Alt, F., Shirazi, A. S., Kubitza, T., Bial, D., Zaidan, F., Ortel, M., Zurmaar, B., Lewen, T., and Schmidt, A. Digifieds: Insights into deploying digital public notice areas in the wild. In *Proc. of MUM'11* (Dec. 2011).

[4] Ballagas, R., Rohs, M., and Sheridan, J. G. Sweep and point and shoot: phonecam-based interactions for large public displays. In *CHI '05 EA*, ACM (NY, USA, 2005).

[5] Bangor, A., Kortum, P., and Miller, J. An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction 24*, 6 (2008).

[6] Beyer, G., Alt, F., Müller, J., Schmidt, A., Isakovic, K., Klose, S., Schiewe, M., and Haulsen, I. Audience behavior around large interactive cylindrical screens. In *Proc. of CHI'11*, ACM (New York, USA, 2011), 1021–1030.

[7] Boring, S., Baur, D., Butz, A., Gustafson, S., and Baudisch, P. Touch projector: mobile interaction through video. In *Proc. of CHI'10*, ACM (New York, USA, 2010), 2287–2296.

[8] Brignull, H., Izadi, S., Fitzpatrick, G., Rogers, Y., and Rodden, T. The introduction of a shared interactive surface into a communal space. In *Proc. of CSCW '04*, ACM (New York, NY, USA, 2004), 49–58.

[9] Brignull, H., and Rogers, Y. Enticing people to interact with large public displays in public spaces. In *Proc. of INTERACT'03*, vol. 3 (2003), 17–24.

[10] Cheverst, K., Fitton, D., and Dix, A. Exploring the evolution of office door displays. *Public and Situated Displays* (2003).

[11] Cheverst, K., Taylor, N., Rouncefield, M., Galani, A., and Kray, C. The challenge of evaluating situated display based technology interventions designed to foster a sense of community. *Proc. of USE'08* (2008), 1–5.

[12] Collomb, M., Hascoët, M., Baudisch, P., and Lee, B. Improving drag-and-drop on wall-size displays. In *Proc. of GI '05* (2005), 25–32.

[13] Fatah gen Schieck, A., Kostakos, V., and Penn, A. The urban screen as a socialising platform: exploring the role of place within the urban space.

[14] Hart, S., and Staveland, L. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Human mental workload 1* (1988), 139–183.

[15] Huang, E. M., Koster, A., and Borchers, J. Overcoming assumptions and uncovering practices: When does the public really look at public displays? In *Proc. of Pervasive'08*, Springer-Verlag (Berlin, Heidelberg, 2008), 228–243.

[16] Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B. B., Druin, A., Plaisant, C., Beaudouin-Lafon, M., Conversy, S., Evans, H., Hansen, H., Roussel, N., and Eiderbäck, B. Technology probes: inspiring design for and with families. In *Proc. of CHI'03*, ACM (NY, USA, 2003), 17–24.

[17] Jacucci, G., Morrison, A., Richard, G. T., Kleimola, J., Peltonen, P., Parisi, L., and Laitinen, T. Worlds of information: designing for engagement at a public multi-touch display. In *Proc. of CHI '10*, ACM (2010).

[18] Khan, A., Matejka, J., Fitzmaurice, G., and Kurtenbach, G. Spotlight: directing users' attention on large displays. In *Proc. of CHI '05*, ACM (New York, USA, 2005), 791–798.

[19] Lazar, J., Feng, J., and Hochheiser, H. *Research methods in human-computer interaction*. John Wiley & Sons Inc, 2009.

[20] Mankoff, J., Dey, A. K., Hsieh, G., Kientz, J., Lederer, S., and Ames, M. Heuristic evaluation of ambient displays. In *Proc. of CHI '03*, ACM (New York, USA, 2003), 169–176.

[21] Matthews, T. L., Rattenbury, T. L., and Carte, S. A. Defining, designing, and evaluating peripheral displays: An analysis using activity theory. *HCI 22*, 1-2 (2007), 221–261.

[22] Memarovic, N., Elhart, I., and Langheinrich, M. Funsquare: First experiences with autopoiesic content. In *Proc. of MUM'11* (Dec. 2011).

[23] Meschtscherjakov, A., Reitberger, W., Mirlacher, T., Huber, H., and Tscheligi, M. Amiquin - an ambient mannequin for the shopping environment ambient intelligence. In *Proc. of AmI'09*. Springer Berlin / Heidelberg, 2009, 206–214.

[24] Müller, J., Exeler, J., Buzeck, M., and Krüger, A. Reflectivesigns: Digital signs that adapt to audience attention. In *Proc. of Pervasive'09*, Springer (Berlin, 2009), 17–24.

[25] Müller, J., Walter, R., Bailly, G., Nischt, M., and Alt, F. Looking glass: A field study on noticing interactivity of a shop window. In *Proc. of CHI'12* (May 2012 (to appear)).

[26] Müller, J., Wilmsmann, D., Exeler, J., Buzeck, M., Schmidt, A., Jay, T., and Krüger, A. Display blindness: The effect of expectations on attention towards digital signage. In *Proc. of Pervasive'09*, Springer-Verlag (2009), 1–8.

[27] Nielsen, J. *Usability engineering*. Morgan Kaufmann, 1994.

[28] Ojala, T., Kostakos, V., and Kukka, H. It's a jungle out there: Fantasy and reality of evaluating public displays in the wild. *Large Displays in Urban Life 4* (2011), 1–4.

[29] Ojala, T., Kukka, H., Lindén, T., Heikkinen, T., Jurmu, M., Hosio, S., and Kruger, F. Ubi-hotspot 1.0: Large-scale long-term deployment of interactive public displays in a city center. In *Proc. of ICIW'10*, Ieee (2010), 285–294.

[30] Peltonen, P., Kurvinen, E., Salovaara, A., Jacucci, G., Ilmonen, T., Evans, J., Oulasvirta, A., and Saarikko, P. It's mine, don't touch!: interactions at a large multi-touch display in a city centre. In *Proc. of CHI'08*, ACM (2008).

[31] Redhead, F., and Brereton, M. Designing interaction for local communications: An urban screen study. In *Proc. of INTERACT '09*, Springer (Heidelberg, 2009), 457–460.

[32] Rogers, Y., Hazlewood, W. R., Marshall, P., Dalton, N., and Hertrich, S. Ambient influence: can twinkly lights lure and abstract representations trigger behavioral change? In *Proc. of Ubicomp '10*, ACM (NY, USA, 2010), 261–270.

[33] Schmidt, D., Chehimi, F., Rukzio, E., and Gellersen, H. Phonetouch: a technique for direct phone interaction on surfaces. In *Proc. of UIST '10* (2010), 13–16.

[34] Schrammel, J., Mattheiss, E., Döbelt, S., Paletta, L., Almer, A., and Tscheligi, M. Attentional behavior of users on the move towards pervasive advertising media. In *Pervasive Advertising*, J. Müller, F. Alt, and D. Michelis, Eds., Springer Limited London (2011).

[35] Shoemaker, G. B. D., and Inkpen, K. M. Single display privacyware: augmenting public displays with private information. In *Proc. of CHI '01*, ACM (2001).

[36] Storz, O., Friday, A., Davies, N., Finney, J., Sas, C., and Sheridan, J. Public ubiquitous computing systems: Lessons from the e-campus display deployments. *Pervasive Computing, IEEE 5*, 3 (2006), 40–47.

[37] Vogel, D., and Balakrishnan, R. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proc. of UIST'04*, UIST '04, ACM (New York, NY, USA, 2004), 137–146.

[38] Vogel, D., and Balakrishnan, R. Distant freehand pointing and clicking on very large, high resolution displays. In *Proc. of UIST '05*, ACM (New York, NY, USA, 2005), 33–42.

[39] Wilson, A. D., and Benko, H. Combining multiple depth cameras and projectors for interactions on, above and between surfaces. In *Proc. of UIST '10*, ACM (NY, 2010).