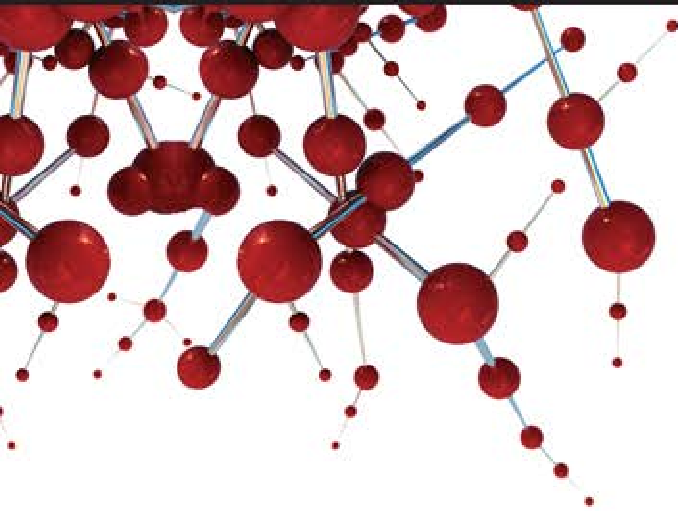
Interactive Technology · Faculty of Communication Sciences · University of Tampere



Ville Mäkelä

Design, Deployment, and Evaluation of Gesture-Controlled Displays in Ubiquitous Environments

Ville Mäkelä

Design, Deployment, and Evaluation of Gesture-Controlled Displays in Ubiquitous Environments

ACADEMIC DISSERTATION
To be presented with the permission of the Faculty of Communication Sciences of the University of Tampere, for public discussion in the Pinni auditorium B1097 on May 26th, 2018, at noon.

Faculty of Communication Sciences
University of Tampere

Dissertations in Interactive Technology, Number 27
Tampere 2018

ACADEMIC DISSERTATION IN INTERACTIVE TECHNOLOGY

Supervisor:	Professor Markku Turunen, Ph.D. Faculty of Communication Sciences, University of Tampere, Finland
Opponent:	Professor Nigel Davies, Ph.D. School of Computing and Communications, Lancaster University, United Kingdom
Reviewers:	Associate Professor Edward Lank, Ph.D. Cheriton School of Computer Science, University of Waterloo, Canada Dr. Sarah Clinch, Lecturer School of Computer Science, University of Manchester, United Kingdom

Acta Electronica Universitatis Tamperensis 1886
ISBN 978-952-03-0743-1 (pdf)
ISSN 1456-954X
http://tampub.uta.fi

Dissertations in Interactive Technology, Number 27

Faculty of Communication Sciences
FIN-33014 University of Tampere
FINLAND

ISBN 978-952-03-0741-7
ISSN 1795-9489

Juvenes Print – Suomen Yliopistopaino Oy
Tampere 2018

Abstract

Interactive displays are increasingly deployed in various locations, such as shopping malls, train stations, and universities. However, the complex and unpredictable nature of public and semi-public spaces introduces many challenges and limitations for the use and evaluation of interactive displays. Interactive displays are at the mercy of a range of external forces, such as the weather, people, and other technology that can interfere with the displays in different ways. Moreover, the evaluation of interactive displays is difficult and time-consuming, which calls for the investigation of new, streamlined evaluation methods.

At the same time, mid-air gestures have recently surfaced as a new interaction modality, but they have not been widely used, for instance, to interact with information displays. Even more so, the unique advantage of gestures—being able to interact from a distance—has not been fully utilized. This ubiquity of gestures enables scenarios that are not possible or require considerably more effort using other interaction modalities. An ideal use case for this is the transferring of content from a display to a personal mobile device using gestures, as users do not need to walk up to the display.

The overall aim of this dissertation is *to help design meaningful and streamlined gestural interactions and to assist with the deployment and evaluation of interactive systems*. Because this objective revolves around several areas within the field of human-computer interaction (HCI), I have divided this work into three themes, each with a distinct research question. First, I study issues regarding the deployment and evaluation of interactive displays. Although this theme is not limited to any particular modality, the results are especially useful for gesture-controlled systems. Second, I focus on the design and usability of gesture-controlled interfaces. Finally, I study the design and user experience of mid-air gestures for cross-device interactions, in particular, for content transfer between displays and personal mobile devices. Hence, the research questions are as follows:

- **RQ1:** *What external factors affect interactive displays, and how can long-term evaluations of interactive displays be streamlined without considerable use of resources?*
- **RQ2:** *What techniques and design principles should be used for gesture-controlled information displays to provide an efficient and usable experience?*

- **RQ3:** *Can mid-air gestures be employed to enable seamless, low-effort exchanging of information between displays and mobile devices, and what is the user experience of such interactions?*

I developed four interactive display applications to carry out the studies presented in this thesis. The first two were utilized to address **RQ1** and **RQ2**, and were deployed in both laboratory and public settings. The remaining two applications primarily contributed to **RQ3** by allowing the transfer of content from the display to a mobile device using mid-air gestures.

In addressing **RQ1**, I first investigated what kinds of external factors affect public display deployments. I discovered six types of external factors and found that their effects are most often negative. Furthermore, I produced guidelines to help other researchers recognize and deal with external factors in their public deployments. Second, I investigated whether evaluations of large-scale public deployments could be automated, to some extent. Through an analysis of extensive interaction and skeletal data collected during a public display deployment containing the data of more than 100,000 passersby, I showed that meaningful findings can be reached by programmatically analyzing logged data. This process is useful especially for evaluating the long-term usage of public displays utilizing gestural or proxemic interactions.

To address **RQ2**, I conducted two studies. First, I investigated techniques to improve target selection in freehand pointing interfaces. Through a rigorous review of existing work, I designed and evaluated a technique called the magnetic cursor. The magnetic cursor significantly improves pointing speed and accuracy, and it alleviates physical fatigue during target selection. Second, I evaluated the use of a gesture-controlled application in both laboratory and public settings, and provided guidelines for designing similar applications.

Finally, three primary findings provide answers to **RQ3**. First, I propose a novel location-based solution to automatically recognize which mobile devices in the space belong to each person. Consequently, users do not need to set up a connection or even interact with the mobile device to transfer content to it from the large display. Such novel pairing methods are a fundamental part of enabling seamless and natural cross-device interactions. Second, I show that using mid-air gestures in combination with the pairing method for content transfer scenarios provides an impressive and promising user experience. Third, I successfully combine different gestural interaction techniques within the same system, which suggests that mid-air gestures can be used for more advanced interactions than what existing research has suggested.

Acknowledgements

Even though there is only one name on the cover, this thesis would not have been possible without the contributions of many.

First of all, I wish to thank my supervisor, Professor Markku Turunen, for putting his faith in me, for providing me with a flexible work environment, and for allowing me to work on many interesting projects. I also wish to thank Associate Professor Edward Lank and Dr. Sarah Clinch for reviewing my thesis. Professor Nigel Davies, thank you for agreeing to act as my opponent in the public defense—it is an honor.

I wish to express my sincere gratitude to Assistant Professor Tomi Heimonen, who mentored me on a great many occasions during the early years of my research career. It was a thankless job I'm sure, but it was immensely helpful.

I also want to thank everyone in my research group at the Tampere-Unit for Computer-Human Interaction (TAUCHI) as well as all my co-authors and everyone I've worked with. There are too many names to list here, but I want to give a special shoutout to my fellow PhD students, in particular Sumita Sharma, for the bi-weekly venting sessions, and for actively sharing ideas and giving feedback. I also wish to thank Jobin James, whose high-quality contributions in the later stages of this research were invaluable. In addition, I want to thank Tuuli Keskinen for her assistance with many practical matters related to putting this thesis together.

I wish to thank my family, especially my mom and dad, for providing me with an environment to grow in which I could thrive and pursue my own interests. I was never forced nor expected to fill a predetermined role, and for that I'm grateful. I believe this has had a great impact on my adult life, on my career choice, and therefore, in part, on making this thesis a reality.

Finally, my dear Iina, thank you for believing in me, even during the times when I didn't. Thank you for being there and for just listening. It was all I ever needed, but I only needed it from you. I know it wasn't always pleasant. I'm doing my best every day to repay you.

Tampere, April 17, 2018

Ville Mäkelä

Contents

List of Publications

This dissertation is composed of a summary and the following original publications, reproduced here by permission.

The Author's Contribution to the Publications

This work would not have been possible without the assistance of many colleagues. All the papers included in this thesis were co-authored. Unless otherwise stated, I was the main contributor in designing and implementing the prototypes needed for the studies, designing and conducting user studies, and writing the papers. All co-authors, at the very least, contributed by giving feedback on the study designs or papers. More information on the contributions for each paper is provided below.

Paper I: "Challenges in Public Display Deployments: A Taxonomy of External Factors"

> This paper presents a literature review and taxonomy of external factors that result in issues with public display deployments. Sumita Sharma and I conducted the literature review, designed the methodology, and wrote the paper.

Paper II: "Semi-Automated, Large-Scale Evaluation of Public Displays"

> This paper presents a method for evaluating public displays by utilizing extensive data sets collected during long-term deployments. We apply this method to a large data set, collected during a one-year deployment of the Information Wall first presented in paper IV. I was the main contributor in developing the system, maintaining the deployment, conducting the data analysis, and writing the paper.

Paper III: "Magnetic Cursor: Improving Target Selection in Freehand Pointing Interfaces"

> This paper evaluates a targeting assistance technique designed specifically for mid-air pointing interactions. I developed the study application, designed the study with the other co-authors, and wrote most of the paper.

Paper IV: "Information Wall: Evaluation of a Gesture-Controlled Public Display"

> This paper presents an evaluation of a gesture-controlled public information display, the Information Wall. I was the main

contributor to the design and the development of the Information Wall, as well as to writing the paper.

Paper V: "Investigating Mid-Air Gestures and Handhelds in Motion Tracked Environments"

This paper presents an evaluation of a large display application, the NFC Information Wall, in which content can be transferred between the display and personal mobile devices by using mid-air gestures and near-field communication (NFC) technology. The development of the system was split between Antti Järvi and me. I was the main contributor to the data analysis and writing of the paper.

Paper VI: "It's Natural to Grab and Pull": Retrieving Content from Large Displays Using Mid-Air Gestures"

This paper presents an evaluation of two-mid air gestures for transferring content from a public display by utilizing the last prototype presented in this thesis, SimSense. SimSense automatically pairs users with their mobile devices based on location data and therefore allows the immediate transfer of content from a distance. I presented the original idea and took the primary role of supervising the project, assisted with the development, and wrote the paper.

1 Introduction

Interactive displays are increasingly deployed in various locations, such as shopping malls, train stations, and conference rooms. Interactive displays provide information, services, and entertainment, and enable users to complete a variety of tasks.

However, it is difficult to successfully deploy and research interactive displays in real-world settings. In particular, two problem areas can be identified that require more attention from the research community. First, various properties of and phenomena in public and semi-public spaces often introduce many surprising issues. For instance, deleterious weather may destroy hardware, and events held at the location may result in stands and crowds that hinder the visibility of, or block access to, the display. It remains unclear exactly what kinds of external forces there are, what the extent of their effects is, and how they could be effectively mitigated. Second, it is demanding to evaluate interactive displays in the wild. Common evaluation methods, such as on-site observations, interviews, and video logs, are time-consuming and often unrealistic for long-term evaluations. This limitation calls for more advanced and less resource-intensive evaluation methods.

Meanwhile, mid-air gestures have surfaced as a new interaction modality. However, mid-air gestures have largely remained a platform for games and playful applications, and have not been widely used in more serious contexts, for instance, to interact with information displays. Therefore, it remains unclear how such interactions are experienced and perceived, and how gesture-controlled interfaces should be designed for serious and immediate use.

Mid-air gestures allow interaction to begin seamlessly without the need for approaching the display, which enables scenarios that are not possible, or require considerably more effort, when using other interaction modalities. One such case is the transferring of content from an interactive display to a personal mobile device. Imagine, for instance, walking past a display some distance away that presents interesting content. Without ever even stopping, with a simple hand gesture, that content would be transferred to your personal device for later consumption. As such, this dissertation is partly built around the vision of truly ubiquitous spaces, where devices, such as displays, can be interacted with at will from anywhere in the space to enable seamless, efficient (cross-device) interactions with technology. However, as pointed out earlier, it remains unclear how mid-air gestures perform and how they should be designed in serious contexts. With these advanced and seamless use cases in mind, it further remains unclear how mid-air gestures actually perform in such scenarios and how users perceive such interactions. Another notable point here is that enabling such seamless scenarios also requires innovative solutions on a system level, which I will delve into towards the end of this thesis.

In this dissertation, I provide an understanding of and solutions to the research problems described above. Understanding the effects of the deployment space on newly deployed technology, for example, interactive displays, and providing tools for evaluating such displays on a large scale are crucial to creating interactive systems that consistently find their users and remain accessible and usable throughout their life cycle. Moreover, understanding how people perceive and engage with gesture-controlled systems and researching various gestural interaction techniques provide more opportunities for seamless interactions with technology.

1.1 OBJECTIVE

The overall aim of this dissertation is *to help design meaningful and streamlined gestural interactions and to assist with the deployment and evaluation of interactive systems*. This goal encompasses several research problems revolving around interactive displays, mid-air gestures, and cross-device interaction. Therefore, I have divided this work into three distinct themes, which I will present next.

Theme 1: Deployment and Evaluation of Interactive Displays

The research community has actively studied interactive displays. One of the main focus points for current research has been how users (passersby) behave around and engage with interactive displays. However, given the sheer complexity of this field regarding, for example, the deployment of interactive displays, usage patterns, and evaluation methods, much remains to be researched. I have identified two distinct gaps in the current research, which I aim to address through this theme.

First, it is unclear how different properties and phenomena within and around the deployment space affect the display and its use. This gap was identified through several public display deployments, all of which suffered from a range of surprising issues that were caused by external factors. For instance, motion tracking sensors that enable interaction with the display can be blocked by other objects, and crowds can prevent access to the display. While some existing research deals with challenges in deploying interactive displays in the field (Harris et al., 2004; Storz et al., 2006; Dalsgaard & Halskov, 2010), none focuses on the *source* of the issues, which calls for a systematic investigation of external factors.

Another gap can be found in research methods that are prominently used to evaluate interactive displays. A common approach is to deploy researchers in the field, who then observe the usage of, and behavior around, the display. While this method can provide deep insight into the usage of the display, it is limited by human capabilities and is time-consuming, making it particularly unsuitable for studying the long-term usage of a display. At the same time, some display installations utilize interaction logs; however, the logs are mainly used to complement other research methods and are often manually analyzed (e.g., Schmidt et al., 2013; Müller et al., 2014a; Walter et al., 2014; Ackad et al., 2015). It is worthwhile, then, to investigate whether such logged data could be extended and used more systematically to alleviate the need to deploy researchers in the field.

Gathering these two gaps in this widely researched area, then, leaves the following research question unanswered:

RQ1: *What external factors affect interactive displays, and how can long-term evaluations of interactive displays be streamlined without considerable use of resources?*

Theme 2: Interaction with Mid-Air Gestures

Interacting with mid-air gestures has long belonged to the realm of sci-fi movies. For some time, such gestures materialized in crude, equipment-intensive prototypes. However, the rise of inexpensive motion recognition hardware, primarily the Microsoft Kinect, made way for gestural interaction in gaming and other experiential contexts wherein the human body and its movements serve as input with no external controllers.

Mid-air gestures have not been extensively studied in terms of usability, user experience (UX), and design. In particular, mid-air gestures have been largely overlooked in non-playful contexts, such as interaction with public information displays. Therefore, within this theme, I aim to answer the following research question:

RQ2: *What techniques and design principles should be used for gesture-controlled information displays to provide an efficient and usable experience?*

The goal is to provide usable gestural interactions to non-expert users. Immediate usability and learnability are key factors in this area, as people in public and semi-public spaces tend to interact with displays for only a short while.

Theme 3: Content Transfer between Interactive Displays and Mobile Devices Using Mid-air Gestures

Different methods for transferring content between devices, namely, between public displays and personal mobile devices, have been actively evaluated and proposed. However, to the best of my knowledge, mid-air gestures have not been previously utilized for content transfer. Rather, current methods employ touch interactions, and some setups exhibit gaze, for example, although these proposed methods are experimental in nature and require additional hardware.

Content transfer can greatly benefit from mid-air gestures, as they allow interaction from a distance, with no need to approach the display from which the content is being taken. Ultimately, I envision ubiquitous spaces wherein people can transfer information at will from far away, without needing to deviate from their path at all, or even without the need to stop walking.

In investigating the previous theme (and addressing RQ2), I received a fundamental understanding of how interaction with mid-air gestures should be designed. Nevertheless, it is largely unclear how mid-air gestures perform in content transfer scenarios, and even more importantly, what the user experience and acceptance of such interactions are. Furthermore, it is unclear whether these unique use cases affect gestures and their design, as well as what kinds of new challenges may arise. Therefore, in this theme, I aim to answer the following research question:

RQ3: *Can mid-air gestures be employed to enable seamless, low-effort exchanging of information between displays and mobile devices, and what is the user experience of such interactions?*

1.2 Context of Research

The work presented in this thesis lies in the field of human-computer interaction (HCI). The main focus of this research is the design, performance, and user experience of the gesture-controlled systems developed and their inherent interaction techniques.

Much of the research presented in this thesis is conducted in *public* or *semi-public spaces*. *Public spaces* generally refer to high-traffic locations that people can enter and exit freely, such as city centers and shopping malls. *Semi-public spaces* generally refer to locations with a more focused demographic, such as universities. However, it is notable that such locations rarely conduct any explicit control over who spends time in that space. In this regard, in the context of this dissertation, there is little difference between public and semi-public spaces. I use these terms rather vaguely, mainly to separate such locations from controlled research settings like laboratories, as well as from those where only select individuals are allowed entry, such as certain work environments.

Interaction technique refers to a specific design and implementation of a mechanism that can be used to interact with a system to carry out a specific task. In research, for instance, we might design and implement several interaction techniques that all achieve the same result. This gives us the opportunity to evaluate and compare the techniques to assess the strengths and weaknesses of each.

Mid-air gestures (or *in-air gestures*) generally refer to the use of physical movement as input. In this thesis, I limit this term to cover vision-based, at-a-distance interaction without external hardware, such as handheld controllers. For instance, from a distance, a user can point with her hand, in mid-air, at elements on a display to select an element or trigger an element-specific action. Note that *gestures* exist in other modalities as well; however, they are outside the scope of this dissertation. Therefore, in this thesis, I use *gestures* and *mid-air gestures* as synonyms, and I employ the term *gesture-controlled interfaces* to refer to interfaces with mid-air gestures as input. I present an overview of gesture interface types in Chapter 3.1, wherein I discuss this topic in more detail.

User experience (UX), generally, refers to a user's subjective experience with and feelings about a system, product, or interaction. I use this high-level definition to encompass most existing definitions within it (see, e.g., All about UX, 2018). I approach the measurement of user experience from assorted perspectives based on the system and interaction being studied as well as the research questions I aim to answer.

1.3 METHODOLOGY

This work consists of six studies, employing a variety of research methods and approaches.

In Publication I (Mäkelä et al., 2017a), we conducted an extensive literature review of 71 papers to collect issues with public display deployments. We then used affinity diagramming to categorize and further analyze the collected issues. In Publication II (Mäkelä et al., 2017c), we investigated the benefits and limitations of a research method that involves the automated analysis of logged interaction, skeletal, and location data of passersby during a long-term public display deployment. We deployed a gesture-controlled display at a university campus for one year, during which time logged data was collected. We then applied our proposed evaluation process to this data to assess whether valid findings could be reached.

For the four remaining publications, the main contribution was a user study and a system prototype. We first designed and developed a functional large display application prototype, the design of which was based on existing research, whenever possible. Subsequently, the prototype was set up either in a public or a laboratory setting (or both), depending on the aim of the study. In the studies, participants carried out specific tasks representing realistic use cases with the developed prototype.

In the user studies, mixed-method approaches were used to gather a set of both qualitative and quantitative data. These consisted of, for example, interaction logs, video and audio recordings, questionnaires, and interviews. The key factor was to overlap these methods somewhat for them to support each other's findings and to gain further insight on the researched phenomena. For instance, interaction logs and questionnaires could provide information on the user's performance and experience with certain interaction techniques, and interviews could further explain *why* the user performed and felt like he or she did.

The research methods employed will be presented in more detail in Chapter 6, when I discuss the six studies that form this dissertation.

1.4 RESULTS

Regarding the first theme and research question, I first set forth a taxonomy of external factors (Mäkelä et al., 2017a) that affect interactive systems: *weather*, *events*, *surroundings*, *space*, *inhabitants*, and *vandalism*. Furthermore, I present examples of each external factor and considerations to prepare for and adapt to these factors. Second, I propose and examine a semi-automated process (Mäkelä et al., 2017c) to evaluate interactive displays and their usage.

For the second theme, wherein I move towards gestural interaction, I first propose and evaluate a technique that aims to address some of the problems with physical freehand pointing interfaces by making the acquisition of, and interaction with, interactive onscreen elements easier (Mäkelä et al., 2014a). Second, I evaluate the usability and user experience of a gesture-controlled information display (Mäkelä et al., 2014b), through which I present recommendations to guide the future designs of gesture-controlled interfaces.

In the third theme, I utilize mid-air gestures specifically for transferring content from situated displays to personal mobile devices. Mid-air gestures have not been previously used for such a purpose. Hence, it is unclear how mid-air gestures in such scenarios should be designed, what the user experience and efficiency of such interactions are, and what challenges arise from the use of both public and personal devices during gestural interaction.

First, I investigate how interaction with a mobile device and a proxemic device (an NFC reader) affects gestural interaction when integrated into the same infrastructure (Mäkelä et al., 2016). Our results show that such integrations are problematic, especially since interacting with a mobile device easily interferes with gestures. We solve this and several problems with the SimSense system, with which users can transfer content to their mobile device using mid-air gestures while keeping the mobile device in their pockets (Mäkelä et al., 2017b). In this final study, we show that mid-air gestures are well suited for seamless content transfer interactions and can offer an impressive user experience.

The results are useful to researchers, practitioners, and designers in many ways. First, the results assist with the successful deployment of interactive systems in the wild and help evaluate their usage. While these results are not limited to specific ways of interaction, the results are particularly practical for interactive displays with motion-tracking capabilities and, therefore, displays that utilize mid-air gestures as input. Second, the results give insight into how people engage and perceive gesture-controlled systems, and help with the design of usable gesture-controlled systems. Third, the results show the promise and positive experience of mid-air gestures for advanced content transfer scenarios, which in turn offer a way forward for new research in this domain.

1.5 STRUCTURE

This dissertation is a compound thesis consisting of six original publications and their summary. In the next three chapters, I present the backgrounds of the three themes and research questions presented above. Chapter 2 deals with the deployment and evaluation of interactive displays, as well as the user behavior and phenomena around interactive displays. Chapter 3 discusses existing work concerning interactions with mid-air gestures. Chapter 4 presents existing methods for cross-device interaction, primarily, content transfer between public and personal devices. In Chapter 5, I present the four interactive display applications that were created to carry out this work. In Chapter 6, I introduce and summarize the six publications contributing to this thesis. Finally, I discuss the results in Chapter 7.

Note that the three themes are not a chronological representation of the work done in this thesis. Work on the themes was partly done in parallel, and the first theme in particular (the deployment and evaluation of interactive displays) slowly materialized during my work on themes 2 and 3 as I gained more experience with interactive displays. Therefore, I have organized this categorization for the sake of clarity, to help focus on one aspect of this research at a time. Another way to perceive this categorization is that I employ a top-down approach, wherein I first research interactive displays on a higher level, then move down to the more focused topic of interaction with mid-air gestures, and then further down into seamless cross-device interaction.

2 Deployment and Evaluation of Interactive Displays

Interactive displays have surfaced in various environments in recent years. For instance, shopping malls deploy interactive displays near entrances through which visitors can access relevant information in the form of maps, opening hours, and discounts. As another example, train and bus stations commonly offer interactive displays for purchasing tickets.

The research community has been increasingly interested in interactive displays, namely, interactions with and behavior around such displays. Hence, a substantial number of interactive display prototypes with a large degree of variance in design, purpose, and research goals have been deployed and studied in different settings. A highly successful example of such research-driven deployments is the UBI-hotspot project (Ojala et al., 2010). UBI-hotspots are interactive touchscreens installed in diverse indoor and outdoor locations around the city of Oulu, Finland (Figure 1). Another good example of large, long-term deployments is the e-Campus system, a network of interactive screens at Lancaster University (e.g., Davies et al., 2009, 2014).

Despite the research community's interest and work concerning interactive displays, I have identified two prominent gaps in the current research. As such, this section shows the research gaps that led to the first research question of this dissertation, the answers to which I provide in Publications I and II:

RQ1: *What external factors affect interactive displays, and how can long-term evaluations of interactive displays be streamlined without considerable use of resources?*

This dissertation concerns interactive displays generally and is not limited to specific types of settings. However, in this section, and when I address **RQ1**, I primarily investigate interactive displays in *public* spaces. This makes sense because public spaces are rarely controlled by those conducting research and deploying interactive displays, which may bring forth many challenges and aspects that would not exist in other settings. In addition, public spaces are inhabited by a great many non-expert users, offering a chance to collect more data from authentic users. Therefore, I argue that addressing **RQ1** in the public context leads to more comprehensive results.

Figure 1. UBI-hotspot, an interactive public display, in downtown Oulu, Finland (Ojala et al., 2010, Figure 1, © 2010 IEEE).

The remainder of this chapter is divided into three subsections. Section 2.1 presents an overview of how people act around and interact with interactive displays. The literature presented in this section partly drives the design of the systems developed to carry out this research; the designs of the studies, including the methodologies used; and the low-level research questions.

Section 2.2 discusses the complexity and challenges of public spaces and the deployment of interactive displays, and shows there is room in current research for a systematic investigation of external factors within public spaces. This gap is closed with Publication I (Mäkelä et al., 2017a), which is presented later in Chapter 6.

Section 2.3 presents the prevailing research methods in public display research, and shows that the methods used especially within in-the-wild studies, while usually resulting in valid findings, are often resource-intensive and time-consuming. As a solution, I present a semi-automated evaluation process in Publication II (Mäkelä et al., 2017c), which I argue is especially useful for long-term deployments.

2.1 Usage Patterns with and Behavior around Interactive Displays

The behavior of passersby—and potential users—around interactive displays is a largely researched, yet complicated topic. Interactive public displays must overcome a wide variety of challenges. For example, they must attract the attention of passersby, motivate them to interact with the display, offer interesting and meaningful content, utilize usable and intuitive interaction mechanics, and offer new and exciting experiences. A multitude of frameworks and guidelines have been proposed aimed at assisting the design and deployment of public interactive displays. Moreover, a wide range of different behavioral phenomena around public displays have been identified.

Michelis and Müller (2011) present the Audience Funnel framework that depicts six phases of public display interaction: passing by, viewing and reacting, subtle interaction, direct interaction, multiple interactions, and follow-up actions (Figure 2). The final stage, follow-up actions, refers to any activity related to the display that takes place after the interaction, for instance, taking a picture of the installation. One of the important realizations in the Audience Funnel framework is that, unlike many other computing technologies, such as desktop computers, interaction with public displays begins before the user has decided to interact with the display or before the user is even aware of the display. As such, the display needs to first draw the *attention* of passersby and then *motivate* them to interact (Müller et al., 2010).

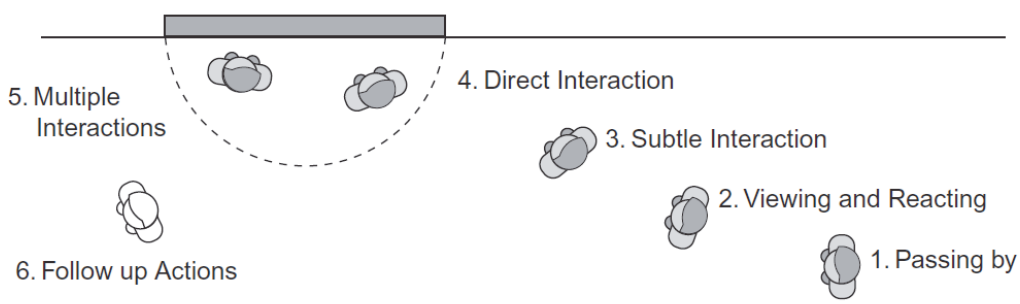


Figure 2. The Audience Funnel (Michelis & Müller, 2011, Figure 6, © 2011 Taylor & Francis Group, LLC).

Vogel and Balakrishnan (2004) presented four phases of interaction with public displays: ambient display, implicit interaction, subtle interaction, and personal interaction. Although they constructed these phases for particular types of displays, that is, displays that change to show more personal content when the user nears and begins to directly interact, some key similarities with the Audience Funnel can be recognized. First, both frameworks agree that displays should acknowledge and react to passersby to communicate their interactivity and to entice interaction. Second, for new users, interaction comes in different phases when they first, for example, try to wave their hand to see what happens before fully engaging with the display. Similarly, interactive displays should respond to this behavior to provide cues about how the interaction works.

Parra et al. (2014) complement the process of drawing attention and motivating users by adding that some installations may also contain an additional step, the “final goal,” after the interaction. In their awareness campaign, the goal was to entice a public display’s users to visit a website after the interaction; however, only a very small fragment of users (0.10%) did so.

As pointed out by the frameworks above, it is crucial to attract the attention of passersby and motivate them to interact. However, achieving these objectives may be surprisingly difficult, as many factors and phenomena are involved. People do not typically go out and *look* for a public display. Rather, they stumble upon it, which highlights the need for adequate motivation (Müller et al., 2010). Additionally, the nature of the location also plays a role, as people in relaxed settings like swimming halls, tend to be more inclined to interact with a display than people in busier, more hectic locations (Ojala et al., 2012). Ojala et al. (2012) further note that people are hesitant to use technology in public. People are afraid that they might break something, disrupt the installation, or bother other people in the space. Moreover, not wanting to interact in public may also be due to the fear of embarrassment (Deterding et al., 2015). Especially for shy people, it might be embarrassing to gain the attention of others while interacting with displays, that is, possibly acting in unexpected ways. On a related note, recent work by Gentile et al. (2017) reports that the presence of unrelated people (a “passive audience”) is connected not only to users’ willingness to interact, but also to how much distance users prefer to keep from the display. For instance, if there are people sitting close to the display, users prefer to interact from further away.

The above-mentioned issues all point to the fact that there is often a high initial threshold for interacting with displays. However, when this initial barrier is overcome, people tend to be more willing to interact. This is due to the *honeypot effect* (Brignull & Rogers, 2003), which refers to increased attention through existing users: passersby pay more attention to a display and are more willing to interact with it if they see other people using it. Afterwards, the honeypot effect has been observed in many studies (e.g., Hardy et al., 2011; Müller et al., 2012). People interacting in groups also tend to interact longer, as witnessed by Perry et al. (2010). This can be due to many reasons. For instance, group members may instruct and encourage each other to interact. Alternatively, they may take turns or discuss the system.

However, group dynamics are complex, and while they generally lead to positive effects regarding public displays, they can also lead to a form of social pressure. For instance, Hardy et al. (2011) noticed that if one member of a group was interacting, others in the group had a tendency to lose interest more quickly, which resulted in them leaving and forcing the one interacting to halt the interaction and rejoin the group.

Ten Koppel et al. (2012) studied the effect of different form factors for several adjacent public displays and found that juxtaposing screens in a line led to the strongest honeypot effect. Müller et al. (2014a) further quantified the honeypot effect in a study in which several interconnected interactive displays were deployed around a university campus. They not only observed an increase in users and interaction times through the honeypot effect, but they also discovered a remote honeypot effect. That is, seeing the representations of users interacting in a different location on the screen made passersby more likely to start interacting.

A frequently observed issue with public displays is known as *display blindness* (Müller et al., 2009): passersby generally only glance at or completely ignore public displays, assuming that displays' content is neither useful nor interesting. Huang et al. (2008) noticed that other attributes, such as the display's elevation, have an effect, suggesting that displays at eye level receive the most attention. However, more recent research by Dalton et al. (2015), wherein they used mobile eye tracking to measure attention to displays, suggested that people may look at displays more than previously assumed. This is likely because they also measured glances from faraway passersby, which earlier observational studies have likely overlooked. However, it remains unclear how distant glances affect the use of interactive displays and to what extent such passersby process the information in the displays. In general, comprehensively measuring and evaluating phenomena such as display blindness is difficult, as studies tend to be conducted in different settings with different factors, using different study methods (Memarovic et al., 2015).

Beyond display blindness, there are also other phenomena that occur between noticing a display and beginning to interact with it. Ojala et al. (2012) identified an issue called *interaction blindness*: even when passersby pay attention to a display, they do not necessarily realize that it is interactive. To combat interaction blindness, Ojala et al. suggested proxemic interactions, wherein the interface would react to, for example, the user's body position to better communicate its interactivity. Interestingly, Müller et al. (2012) conducted a study that fulfills this suggestion, and they observed a phenomenon called the *landing effect*. In their field study, their interactive display reacted to the movement of passersby. Visualizing people with mirrored silhouettes, they found that people would often notice the display's interactivity after they had already passed by it. Therefore, passersby interested in the display had to turn and walk back to it.

In summary, interactive displays need to overcome a variety of challenges to achieve success. Fortunately, these challenges can be in large parts overcome with good design and creative solutions. For instance, displays with motion-tracking capabilities can exhibit visual and auditory cues to draw the attention of passersby as well as to communicate that the display is interactive (Hardy et al., 2011). I will return to this discussion in Chapter 5, wherein I present the four gesture-controlled systems developed to carry out this research and discuss how the designs of these systems aim to address some of the issues presented in this section.

2.2 Challenges in Public Display Deployments

Deploying interactive displays in public spaces may be a seemingly simple task. However, a multitude of challenges are present, and the whole process, from an idea to a fully working—and successful—public display application and its deployment can be long and complicated.

While the design and development of an interactive display is already challenging, it is, in fact, the deployment space and all its functions that bring forth many unique opportunities and challenges. In short, public spaces and the people within them are messy and complicated. It is difficult to predict how an interactive display will affect the space and how the display will be perceived and used—if it will be used at all.

The process of designing and deploying an interactive display often involves a multitude of people and stakeholders. While stakeholders are out of the scope of the research problem presented here—and are often considered an internal part of a public display's deployment—it is important to acknowledge their existence and effect on the deployments. Stakeholders differ between projects, but may include parties such as location owners, display owners, administrators, researchers, users, and content providers (Valkama & Ojala, 2011; Clinch et al., 2012; Hosio et al., 2014; Elhart et al., 2016). Generally, stakeholders come with individual needs, expectations, and requirements, and a successful deployment should provide sufficient value regarding its cost for each stakeholder. For instance, researchers conducting the deployment are often different from those who own and manage the deployment space. Consequently, the location's managers must also see value in the deployment to go through the trouble of maintaining it (Hosio et al., 2014). Similarly, Elhart et al. (2016) discuss their experiences from deploying an interactive system at a university campus, in which they had to convince the university administration of the deployment's benefits as well as provide satisfactory tools for the administration to manage and control the system.

After conducting several display deployments in the field, my colleagues and I came to realize that our deployments often suffered from a variety of surprising issues that were not necessarily due to the characteristics of the deployment itself, but the various functions and phenomena in and around the deployment space. We call these *external factors*—properties and phenomena that are unrelated to the deployment, which still affect the deployment one way or another. For instance, rain can destroy hardware, and sunlight can hinder the visibility of the display's content or affect the performance of motion-tracking sensors—even indoors. As another example, people working at the location, such as janitors, may unintentionally impact the deployment by, for instance, leaving their equipment in the way or moving parts of the system. Such issues have not received much attention in the past, as current research largely focuses on *users*, for example, how people interact with the display, and how the barriers preventing casual passersby from becoming users can be overcome – as shown in the previous section.

Some existing work has presented challenges and lessons learned from public display deployments based on their own experiences (Harris et al., 2004; Storz et al., 2006; Dalsgaard & Halskov, 2010). Harris et al. (2004) discuss their experiences from two public deployments. One deployment was based indoors, and one, outdoors in a forest-like environment. Of particular interest in the work by Harris et al. (2004) is that they discuss the key differences between two very dissimilar deployment environments and further elaborate on how they needed to adapt to environmental changes in their outdoor deployments. Factors often taken for granted may not work as well when deploying technology outdoors. For instance, the ground in outdoor settings may be unstable and change from hard to soft, requiring base plates and other workarounds to offer stable attachment points for technology.

Storz et al. (2006) present 13 lessons they learned after facing issues while deploying interactive displays in different environments. They cover a variety of topics related to, for example, technology, monitoring, content creation, and user expectations. For instance, they discuss environmental effects on their deployments: in one of their deployment spaces, fire alarms were tested weekly, during which time power was completely cut off. Another deployment was set up in a tunnel under a heavily trafficked road: diesel fumes from the traffic clogged the projector, despite custom-built protective cases.

Dalsgaard and Halskov (2010) draw on their experiences from five different deployments, identifying and discussing eight challenges in the context of deploying urban media façades. Although media façades are often unique, many of the presented challenges are applicable to interactive displays in general. After all, media façades *are* public displays, albeit they can greatly vary in function, size, and shape. As an example, one of the challenges they discuss are the demands for robustness and stability. For instance, in one of their installations, reflections from puddles on the ground interfered with the sensors, and other installations were subjected to vandalism.

While the advice and guidelines presented above (Harris et al., 2004; Storz et al., 2006; Dalsgaard & Halskov, 2010) are certainly useful, two limitations can be recognized in the existing research. First, the existing research does not explicitly target or investigate the *source* of the encountered issues, that is, the external factors. Second, the existing research presents only the authors' own experiences.

Therefore, the research community is missing work that a) focuses on *external factors* and b) draws its results from the whole research community. I believe that systematic investigation into external factors will illuminate what kinds of external factors exist, to what extent they affect public displays, and how their effects can be tackled. I close this gap with the work presented in Publication I (Mäkelä et al., 2017a), presented later in Chapter 6.

2.3 EVALUATING PUBLIC DISPLAYS

Most public display research is conducted in the field, in a realistic setting. This is often mandatory to gain a deeper understanding of the display's usage. Public display research is rarely primarily concerned with the display's objective performance, but rather, with how the display is perceived and used in the real world, which is often impossible in a laboratory environment (Alt et al., 2012). For instance, Marshall et al. (2011) found that their collaborative tabletop display was used very differently during a public deployment than what was suggested by lab studies.

Alt et al. (2012) present an overview of study types and methods for the evaluation of public displays based on a literature survey. One of the several important points they make is that public displays have differing purposes, goals, and research questions. They recognize seven general areas of interest in which researchers often invest: audience behavior, user experience, user acceptance, user performance, display effectiveness, privacy, and social impact. Alt et al. (2012) further divide public display studies into three types: descriptive studies, relational studies, and experimental studies. Of these, however, descriptive field studies are the most common (Müller et al., 2014b), wherein a prototype is deployed in the wild without variations.

Alt et al. (2012) point out five primary data collection methods: interviews, questionnaires, focus groups, observations, and logging. Public display research in the wild often combines several of these. Observation is often the primary research method, which usually involves deploying one or more researchers in the field, who then observe passersby and users of the display while taking notes or using other means of recording. Observations are often supplemented by, for example, interaction logs as well as interviews conducted when the opportunity arises.

However, while the approach described above can reach high ecological validity, the obvious downside is that conducting research by deploying personnel in the field is time-consuming and resource-intensive. Especially when studying a public display's long-term usage on a large scale, observations are unrealistic.

Simultaneously, logged data can be easily gathered during a public deployment. However, in the past, logged data has been mainly used to support findings from observations, and logged data has been analyzed manually (e.g., Schmidt et al., 2013; Müller et al., 2014a; Walter et al., 2014; Ackad et al., 2015). Ackad et al. (2015) supported their findings by manually analyzing the depth data collected by a motion tracking sensor to deduce basic usage information such as session duration and the number of people in front of the display. Similarly, Müller et al. (2014a) collected depth data from six deployed public screens and manually analyzed the data to classify passersby into categories. Walter et al. (2014) also supplemented their observations by manually analyzing the recorded depth data from their deployment, as did Schmidt et al. (2013) with their Screenfinity system.

The important questions here, then, are a) whether logged data could be used more extensively and not just as supporting data for other research methods, and b) whether results from logged data could be extracted in a less resource-intensive manner, that is, whether the process could be automated partially or completely.

Some lightweight automation of data analysis was conducted by Müller et al. (2012), who wrote software to search for sessions from their data in which a user was detected for more than four seconds. However, the sessions were then reviewed and annotated manually. Therefore, automation was primarily used for filtering the data, rather than analysis itself.

Prior work has presented tools for tracking and analyzing audience behavior (Williamson & Williamson, 2014; Elhart et al., 2017). Williamson and Williamson (2014) presented the Pedestrian Tracker tool, which collects location data of people through a sensor with a bird's-eye view of the space and automatically generates movement paths from the data (Figure 3). Hence, the tool enables an analysis of how passersby approach or avoid public displays, and as such, how technology can change traffic in a space. The strengths of the tool are that it scales well and works independently of the deployment itself.

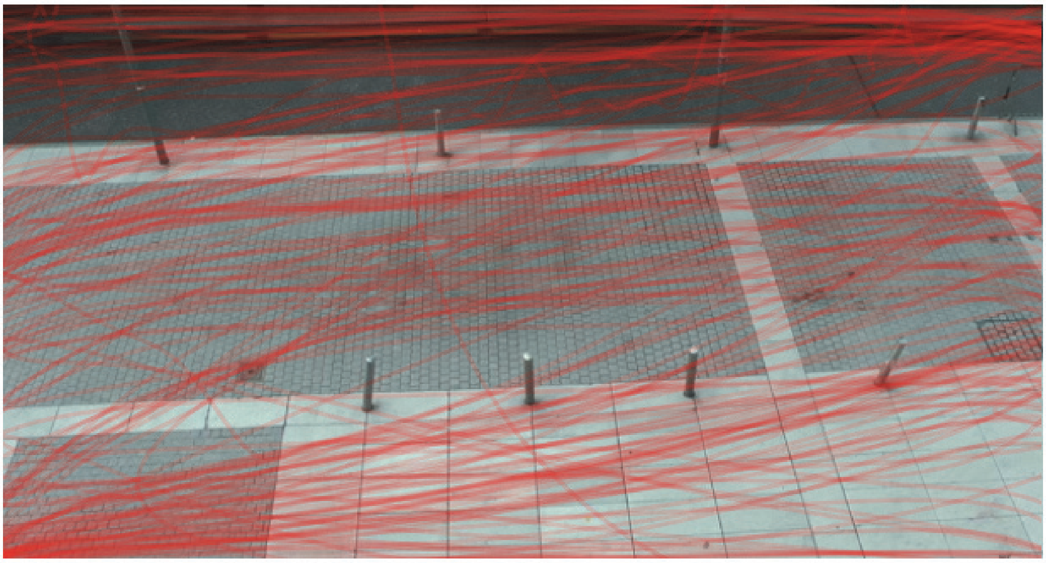


Figure 3. Visualization of pedestrian traffic using the Pedestrian Tool (Williamson & Williamson, 2014, Figure 1, © 2014 ACM).

Elhart et al. (2017) recently presented the Audience Monitor tool, which is relatively similar to the Pedestrian Tracker. The Audience Monitor, however, includes a rich set of analytics and visualizations, enabling researchers to track audience behavior in front of the display regarding, for example, walking speed and distance, and the number of passersby at certain times of day.

The Pedestrian Tracker and the Audience Monitor present similar limitations. The tools primarily focus on the presence and movement of people in the tracked space. They do not directly connect this data with display interactions; nor do they present ways to identify whether a user's actions relate to the display. Therefore, there is room for investigating methods that collect both the skeletal (movement) data of passersby as well as interaction data, allowing a more in-depth analysis of, for example, groups where some members interact with the display, while others remain passive observers. I aim to close this gap with Publication II (Mäkelä et al., 2017c), presented later in Chapter 6.

Mikusz et al. (2017) discuss the potential benefits and challenges of "analytics synthesis": the combination of analytics data from multiple stakeholders to reach new insights on, for example, how displays are used and how passersby move around the space. For instance, adjacent shops at a shopping mall might each have their own digital signage outside the stores. Combining their analytics data might lead to further insights and mutual benefit. While combining data from multiple stakeholders is out of the scope of the research problem being addressed here, the underlying concept is somewhat similar, that is, the combination of data from divergent sources and viewpoints to reach new insights. Nonetheless, the discussion provided by Mikusz et al. (2017) in part underscores the importance of, and possible future advancements in, this field.

As a final remark for this chapter, it is worthwhile to clarify that not *all* studies need to be conducted in the field. Even though this thesis revolves around public spaces and the usage of public displays, this thesis also concerns, for example, different low-level design decisions and interaction techniques for such displays. The most notable example of this is Publication III (Mäkelä et al., 2014a), wherein I do not investigate the *use* of a public display but, rather, the efficiency of techniques aimed at improving interactions with a display. Therefore, in some publications I will present a lab study instead of an in-the-wild deployment.

3 Interaction with Mid-Air Gestures

Mid-air gestures are a relatively new way of interaction that exhibit a variety of unique characteristics. Two of the most obvious benefits of gestures are that interaction can happen from a distance, and interaction can be device-free, that is, users do not need to hold or touch external hardware. These two characteristics alone bring forth new possibilities and advantages. Consider, for instance, the possibility of interacting with a faraway device without the need to walk up to it and touch it, making the information—and technology in general—around us more readily available.

The widely adopted use of mid-air gestures, however, has long been a possibility only envisioned in sci-fi movies. Even though gestural interface prototypes were already present nearly four decades ago (Bolt, 1980), it was not until the release of the Microsoft Kinect sensor in 2010 that gestural interactions were really introduced to the public. However, for most consumers, mid-air gestures were mainly a new way to engage with playful, lightweight games at home. Accordingly, it remained unclear how gestural interaction would perform and how it would be experienced in other contexts.

In research, some gesture-controlled installations have also surfaced. For instance, Müller et al. (2012) deployed a physics-based game that users would control with their bodies to study how well passersby notice the interactivity of a public display (Figure 4). As a more serious example, the Media Ribbon (Tomitsch et al., 2014) is an interactive information display providing relevant information to users at a university campus. The Media Ribbon is controlled by swiping or pointing the hands in different directions (Figure 5). These two examples provide a good overview of the versatility of gestural interfaces that caters to a variety of use cases. In this dissertation, however, I will focus more on serious interfaces, much like the Media Ribbon mentioned above.

Figure 4. An interactive display on a shop window with which users can interact using body movements. Users are visualized on the screen as silhouettes (Müller et al., 2012, Figure 1, © 2012 ACM).

Figure 5. The Media Ribbon. *Top*: The system is installed on the wall of a university building. *Bottom*: the application screen displays university-related information, a skeletal representation of the user, as well as instructions for interaction (Tomitsch et al., 2014, Figure 1, © 2014 ACM).

Despite researchers' interest in gestural interactions, much remains to be researched. As evidenced by the two examples above, gestural interfaces can vary significantly respecting, for example, the style of interaction, design, and purpose, and therefore will need to incorporate a range of guidelines and principles. Meanwhile, it is unclear what kind of user experience gestural interfaces offer and how such interactions are perceived in a serious context. With these gaps in mind, I formulate the second research question that I address in Publications III and IV:

RQ2: *What techniques and design principles should be used for gesture-controlled information displays to provide an efficient and usable experience?*

In the remainder of this chapter, I provide an overview of the characteristics, challenges, and existing design principles of interaction with mid-air gestures.

3.1 Types of Gestural Interfaces

As already outlined in the beginning of this chapter, there is considerable variation in the types of gestural interfaces and how they are interacted with. I roughly categorize gestural interfaces into three types:

- *Freehand pointing interfaces.* The primary way of interacting with freehand pointing interfaces is to face the display and point towards it with either hand. The pointing hand is often visualized as an onscreen cursor. As such, freehand pointing does not necessarily contain distinct "gestures," but rather, it is the pointing itself that conveys meaning. Note, however, that pointing can be combined with other gestures. For instance, when the user is pointing at an element, the interface can do different things when the user forms a fist (that is, "grabs" the element) and when the user makes a pushing motion.

- *Gesture set interfaces.* These interfaces are interacted with through a group of predefined gestures or poses. For instance, swiping to the side with either hand might change to the next and previous content item, and raising a hand might display more information about the current item, as demonstrated by the Media Ribbon in Figure 5 (Tomitsch et al., 2014).

- *Full-body interfaces.* In full-body interfaces, users interact with their body without predefined gestures or poses. A good example is the interactive shop window in Figure 4 (Müller et al., 2012), in which users can use their onscreen silhouettes to interact with objects falling from the top edge of the screen in the way they best see fit. For instance, users can hit or kick the objects and watch them fly around the screen. Note, however, that simply displaying a silhouette or some other representation of the user does not necessarily mean that the interface can be called *full-body*. In fact, such representations are often used to provide feedback to the user and to communicate that the display is interactive (e.g., Ackad et al., 2016).

Note that this separation into three types is not exhaustive; nor is this categorization commonly presented in the literature. I have created such a categorization for the sake of clarity regarding the remainder of this thesis. I argue that it is clearer to divide gestural interfaces into three generic types to avoid having to delve into the nuances and distinctions of different gestures, such as symbolic, deictic, and iconic gestures (Karam & schraefel, 2005).

Regarding the interface types above, in this dissertation I will mainly focus on the first two types, that is, freehand pointing and gesture set interfaces. The main reasons for this focus are the social implications of gestures (Perry et al., 2010) and fear of embarrassment (Deterding et al., 2015) during public interaction, which were discussed in the previous sections. This thesis revolves around serious information displays, and therefore I focus more on pointing and gesturing with the hands, rather than full-body gestures that often draw attention and are mainly used for playful interactions.

3.2 Challenges of Mid-Air Gestures

Despite offering many unique benefits and opportunities, mid-air gestures also present a great diversity of challenges. Many can be tackled or alleviated with good design; however, some remain unsolved.

Mid-air gestures suffer from an issue commonly known as the *Midas touch* (Kjeldsen & Hartman, 2001). The Midas touch, at its core, refers to unintentional interaction with a system. This is because gestures are continuous in their input in that the user is constantly tracked by a motion-tracking sensor and it is the system's responsibility to deduce meaningful information from the tracking data. Thus, a system may erroneously interpret the user's movement as an intent to interact. With mid-air gestures, this can easily happen when, for example, users are talking to another person and (subconsciously) moving their hands, or when the same interaction space is used to carry out other, unrelated tasks that require physical movement (Walter et al., 2014). The same issue is prominent in other modalities that are likewise based on continuous tracking, such as eye gaze (Vrzakova & Bednarik, 2013).

The Midas touch issue can be combated with so-called *clutch actions*, wherein users perform an action to communicate to the system that they want to interact. For instance, Walter et al. (2013) used a "teapot" body pose, that is, users put their hands on their hips, to activate interaction with a display. Schwarz et al. (2014), conversely, analyzed the user's body, gaze, and gestures to predict whether they wanted to engage the system. Freeman et al. (2016) approached this from another perspective, for example, using light and audio to communicate the interaction area of different devices, preventing users from accidentally interacting with the wrong device.

Partly linked to the Midas touch, another challenge with mid-air gestures is the lack of tactile feedback. In traditional input methods, there is a clear start and end for an interaction, for instance, when pressing a button or touching a screen. Moreover, when interacting with physical input devices, their form factors force the user to interact correctly. For example, a knob can only be turned in two meaningful directions, and a button can only be pressed down or released (Nancel et al., 2011). In freehand gesturing, this feedback does not exist. Therefore, other types of feedback (namely, visual and auditory) are emphasized even further.

Another issue with gestural interaction is known as the *Gorilla arm*, which refers to physical fatigue due to prolonged gesturing and body movements. While the term was originally introduced in the context of vertical touchscreens, it is even more prominent in gesturing. Researchers have proposed models to evaluate fatigue levels (Hincapié-Ramos et al., 2014; Jang et al., 2017) to help with the design and evaluation of cost-effective gestural interactions.

In addition, Perry et al. (2010) point out that gestures contain a highly performative social aspect. They note that expressive gestures can help raise interest in the system and foster learning and collaborative tasks. However, they admit that gestures may simultaneously discourage use due to their expressiveness and visibility. As noted in the previous chapter, people fearing embarrassment may hesitate to interact with a display around other people (Deterding et al., 2015). Mid-air gestures might be particularly susceptible to this problem, but the extent of this challenge remains unclear.

In summary, interaction with mid-air gestures presents many unique, interesting challenges. While some stem from the limitations of current technology, many issues can also be addressed from the design perspective. I will discuss this further in Chapter 5, in which I present the four gesture-controlled applications developed to conduct this research. Of the challenges presented here, I particularly focus on the Gorilla arm issue by proposing a technique to alleviate the physical demands required to point to and interact with onscreen elements in gestural interfaces (Mäkelä et al., 2014a).

3.3 Designing Mid-Air Gestures and Interfaces

Various techniques for selecting items on a large screen using mid-air gestures have been proposed and studied. Hespanhol et al. (2012) assessed the intuitiveness of several mid-air gestures for said purpose (Figure 6). In their studies, the *dwelling* technique (also called *point-and-dwell*) was found to be the most intuitive. *Dwelling* refers to selections that trigger over time: users point at a desired item on the screen and keep their hand still until the item is selected. Usually, when an item is being pointed at, an animation is shown atop the corresponding item to visualize that it is about to be selected. Support for dwelling as an intuitive selection technique was also presented by Walter et al. (2014), as they found that people would commonly point and dwell at interactive elements if no explicit onscreen instructions were provided.

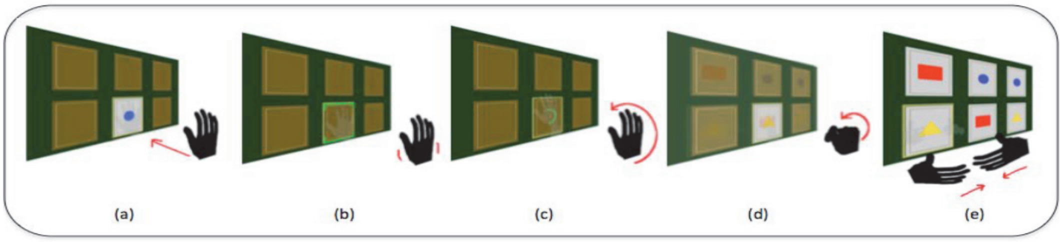


Figure 6. Gestures for selecting targets in mid-air: (a) pushing, (b) dwelling, (c) lassoing, (d) grabbing, and (e) enclosing (Hespanhol et al., 2012, Figure 1, © 2012 ACM).

Yoo et al. (2015) found somewhat contradictory results, as in their study, participants preferred the push gesture over dwelling for selecting items. Yoo et al. (2015) explain this by the perceived delay that occurs during the dwell time. It should be noted, however, that Yoo et al. (2015) used a relatively long dwell time of two seconds as opposed to just the one second employed by Hespanhol et al. (2012). A relevant implication here is that the preferred dwell time can vary significantly among users, namely, based on the amount of experience with the interface. For instance, in the study by Yoo et al. (2015), some participants complained that the dwell time was too fast; others reported it to be too slow.

Moreover, gestural interfaces should support one-handed interactions, as people (especially in public spaces) often carry items, such as bags and coffee cups (Hardy et al., 2011; Walter et al., 2014; Ackad et al., 2015). Walter et al. (2014) further found that, even if both hands are free, people still tend to use only one hand for interaction.

Ackad et al. (2015) supply guidelines for designing usable, easy-to-learn gesture set interfaces. They recommend that gesture sets be kept small, and one practical way to do so is to *overload* gestures; that is, the same gesture can have multiple functions depending on the current state of the interface. Second, they recommend designing *symmetric* gestures, such as similar swipe gestures in both directions (left, right) and similar pointing gestures (up, down). They then recommend visual feedback in the forms of (1) always visible dynamic instructions for the core gestures and (2) visualization of the user in some form, such as a skeleton, as it attracts attention and helps the user understand the interaction.

Ackad et al. (2016) further investigated the effect of two different onscreen user representations on how well the display draws interest and how it is interacted with. They found that a silhouette draws more users and facilitates serious interaction with the display. The representation of a skeleton facilitated more playful interaction, which was primarily described as a kind of puppeteering, in attempts to manipulate the skeleton in ways that appear funny. The skeleton was also found to result in longer interaction sessions, as opposed to the silhouette.

In summary, current researchers have done a significant amount of work in evaluating techniques and designs to support gestural interaction. As outlined earlier, however, gestural interfaces can vary greatly not only in design, but also in what kinds of gestures they incorporate and how exactly they are used. Therefore, the existing guidelines and findings are far from exhaustive, although they do provide valuable guidance that can be applied in future interfaces.

Putting together the findings from current studies as well as the identified challenges regarding gestural interaction, two main implications can be drawn. First, the consensus seems to be that gestural interactions should be simple, for instance, by utilizing gesture sets that employ symmetric and consistent gestures, and by designing for one-handed interaction. Second, especially considering the lack of tactile feedback as well as the fact that gestures are a novel modality to many, there should be special emphasis on clear, dynamic visual feedback. For instance, users and passersby should be visualized in some form (e.g., skeletal/silhouette representation), and instructions should be provided based on the state of the system.

4 Content Transfer between Displays and Mobile Devices

In many public and semi-public spaces, there are situations wherein the user wants to transfer content between devices, especially from a public display onto a personal mobile device. Imagine, for instance, a person hurrying from an office to catch a bus who would like to transfer news articles from a display to their smartphone to read later on the bus. Especially now that such spaces are increasingly being inhabited by interactive displays and other technology, more and more opportunities are arising wherein it is useful to transfer content between devices.

In this section, I present an overview of existing methods for transferring content between public and mobile devices, and show the limitations in current work that led to the third research question of this dissertation, addressed in Publications V and VI:

RQ3: *Can mid-air gestures be employed to enable seamless, low-effort exchanging of information between displays and mobile devices, and what is the user experience of such interactions?*

Current research has been actively exploring new and creative ways to transfer content between devices. Current solutions utilize a variety of technology and methods, for example, quick response (QR) codes, near-field communication (NFC) technology, radio frequency identifier (RFID) tags, Bluetooth pairing, and Wi-Fi connections. In this section, I will present some of these solutions, focusing on content transfer and other interactions especially between situated displays and personal mobile devices, such as smartphones.

Many solutions utilize spatial interactions, for example, touching the display or some attached device with the mobile device to transfer content. One such approach is to use near-field communication (NFC), wherein an NFC-enabled device, such as a smartphone, can be situated next to an NFC reader or a tag to transfer information between the two. Hardy and Rukzio (2008) assembled a mesh of NFC tags on a wall, on top of which a display was projected. Each NFC tag would then represent an action or an element on the screen. Therefore, users could, for example, transfer pictures from the display to their mobile devices by simply touching the corresponding picture (i.e., the NFC tag behind it) with the mobile device. In their studies, Hardy and Rukzio found that their technique was comparable to conventional cross-device methods in terms of speed, but was seen as more intuitive and enjoyable as well as easier to use.

Broll et al. (2011) utilized a similar mesh of NFC tags for a variety of interactions. Besides moving content from a screen to a mobile device, users could, for example, upload content from the mobile onto the screen and preview tooltips of onscreen items on their mobile device. Seewoonauth et al. (2009) used a similar NFC approach, focusing on content transfer between a mobile device and a laptop, allowing for quick transferring of files between the devices.

Spatial interactions have also been utilized in tabletop installations. PhoneTouch takes a different approach to manipulate content in which users touch the display's surface (Schmidt et al., 2010). Users could pair their devices with the horizontal public display by tapping it three times. These touch events are matched with the touch events recognized by the display to establish a connection. Through this pairing method, users can, for example, display photos from their smartphones on the screen or transfer objects to their mobile device (Figure 7). Wilson and Sarin (2007) utilized a combination of Bluetooth and computer vision, allowing users to pair their mobile devices with a tabletop display simply by placing the device on the table's surface. While this enabled content transfer between the two devices, it also allowed more advanced scenarios, such as interaction between two mobile devices placed on the table surface. Bazo and Echtler (2014) implemented a similar method wherein devices can be placed on the table's surface to enable content exchange using a combination of NFC tags and optical markers.

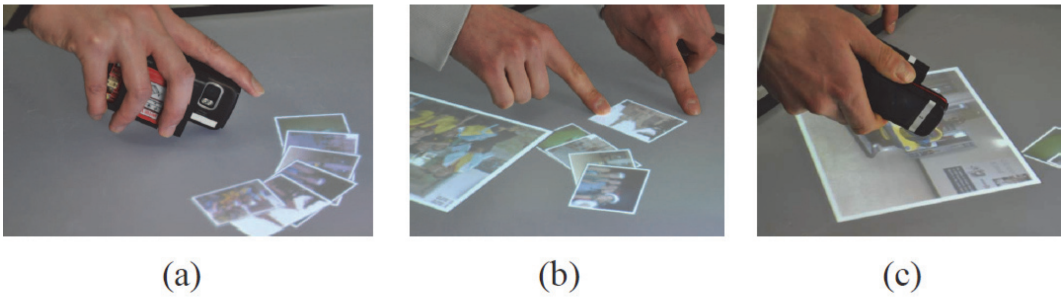


Figure 7. PhoneTouch: (a) Users can share photos from their smartphone on the display's surface. (b) Users can manipulate the photos on the display's surface. (c) Other users can transfer the photos to their smartphones (Schmidt et al., 2010, Figure 2, © 2010 ACM).

QR codes have been utilized in the real world somewhat actively. QR codes are visual markers that can be scanned with a mobile device's camera using a QR-enabled application. QR codes can contain a range of content, such as a URL or identification-related information. For instance, a movie poster at a bus stop might contain a QR code that a person can scan to access the movie's website for more information, or a conductor might scan a train ticket via the QR code on the ticket. In research, Karaman et al. (2016) utilized QR codes in a museum to offer visitors more information about the artwork in the museum through their personal mobile devices.

Alt et al. (2013) studied many techniques for exchanging content between a public display and a mobile device (Figure 8). Among the techniques were QR codes, an NFC-based approach similar to that of Schmidt et al. (2010) presented earlier, as well as more traditional approaches such as sending content by email and printing the content on paper. An interesting result by Alt et al. (2013) is that there was no single best technique for exchanging content. Rather, the preference and efficiency of the techniques is dependent on, say, the user and the current situation.

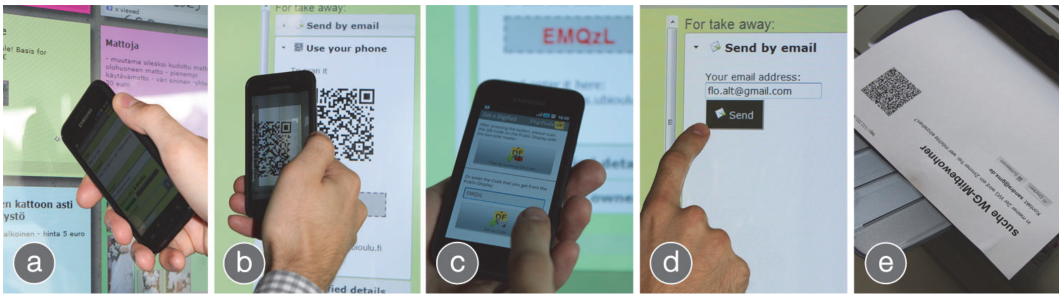


Figure 8. Interaction techniques for taking away content from a public display. a) Touching the display with the smartphone. b) Scanning a QR code. c) Entering an alphanumeric code. d) Sending the content to an email address. e) Paper printout (Alt et al., 2013, Figure 4, © 2013 ACM).

A survey of device pairing methods between smart devices and pervasive screens was presented by Ng et al. (2017). While device pairing can be utilized for many tasks and scenarios, the transfer of content between public and private devices is nonetheless one of the popular areas in this domain. Ng et al. (2017) also discuss the prominence of radio-based pairing methods, such as NFC tags, and vision-based pairing methods like QR codes.

Turner et al. supply an assortment of interesting techniques for transferring content between a display and a mobile device wherein they combine gaze and touch interactions (2013a, 2013b, 2014). In their *Eye Pull, Eye Push* concept (Turner et al., 2013a), users can look at content on a public display and swipe down on the touchscreen of their mobile device to transfer the content to the mobile device. With their *Eye Drop* concept (Turner et al., 2013b, 2014), content can be transferred with a similar method; however, the exact location of where the content will be positioned on the mobile device can be decided by dragging the finger across the screen before releasing it.

Some solutions expand the interaction space by allowing the transfer of content without being in the immediate proximity of the display. Boring et al. (2007) used a dedicated mobile application with which users could take a picture of part of the display. A remote server would then match the picture with the display's content and provide the recognized screen elements in a more accessible form. Baldauf et al. (2012) employed a mobile application and the device's camera: users could point at the screen with the device's camera and interact with the remote screen's elements on the mobile screen, for example, by dragging elements into their own local gallery to transfer them.

She et al. (2013) showcased a system wherein users can point at one of multiple screens with a smartphone and perform a dragging gesture, that is, lift the smartphone towards their face. Content from the screen pointed at before the gesture is then transferred to the smartphone. The focus in their study was to differentiate between multiple screens—multiple items on the same screen are not differentiated (in their study, they used one full-screen picture per screen).

Langner et al. (2016) provide many techniques for transferring content from a large display to a mobile device. For instance, single content items can be transferred by pointing at an item with the mobile device and swiping down on the mobile's touchscreen. In their solutions, Langner et al. (2016) attached external sensors to the mobile device to deduce the pointing direction.

In this section, the most important takeaway is that current solutions for content transfer between interactive displays and mobile devices require users to hold and manipulate the mobile device to transfer content. Therefore, users must undergo the effort of locating their mobile device, for example, in a pocket or bag before interaction can begin. This can be slow and tedious, especially for quick, one-way content transfers.

In addition, most current solutions require users to approach the display and touch it with their hand or with the mobile device. This limits the interaction space and imposes restrictions on where the display can be positioned and how large it can be. Some solutions do allow interaction from a distance (e.g., Boring et al., 2007; Baldauf et al., 2012; She et al., 2013; Langner et al., 2016), although it remains unclear exactly how far away users can interact from and how much of an effect distance has on performance. Furthermore, only Boring et al. (2007) report the speed of their technique, which was almost ten seconds for a single transfer.

To the best of my knowledge, mid-air gestures have not been used in content transfer before. This thesis largely centers on the concept of utilizing mid-air gestures for content transfer interactions, not only solving some of the issues inherent in current content transfer solutions, but also offering a way forward towards more advanced and seamless cross-device interactions. I envision ubiquitous spaces wherein people can interact with distant devices at will from anywhere in space and transfer information between devices, both public and personal.

It should be noted that, in many existing solutions, the manipulation of the mobile device makes some interactions possible which mid-air gestures necessarily do not. For instance, some solutions presented above enable the exchange of content between not only a display and a mobile device, but also between multiple mobile devices connected to the display (Wilson & Sarin, 2007; Schmidt et al., 2010; Bazo & Echtler, 2014). Therefore, I do not claim that the methods I propose in this dissertation are comprehensively better. Rather, the methods I propose focus on quick, one-way content transfer scenarios (from a display to a mobile device) and on scenarios that are not possible or require considerable effort using other modalities. One such scenario is enabling the transferring of content from a distance while on the move.

5 Gesture-Controlled Large Display Applications

In this chapter, I present four gesture-controlled large display applications built to carry out the research presented in this thesis: Information Wall 1, Information Wall 2, NFC Information Wall, and SimSense. The first two applications, Information Wall 1 and 2, primarily explore gestural interaction in the context of large information displays.

The NFC Information Wall extends the previous systems by allowing users to transfer content from the display to their mobile devices by using an NFC reader. Moreover, users can upload their own content on the screen by creating notes using a mobile application and finalizing the upload using the NFC reader.

Finally, SimSense takes content transfer between large displays and mobile devices in a more sophisticated and streamlined direction by automatically recognizing the owners of mobile devices in the space. Therefore, users can transfer content to their mobile devices directly using only mid-air gestures, with no need to manipulate the mobile device.

5.1 INFORMATION WALL 1

Information Wall 1 is a gesture-controlled large information display application that allows users to access location-relevant content from a distance.

Passersby are visualized on the screen as rectangular shapes that follow the movements of the corresponding passersby and change their size based on how far away they are (Figure 9A). At the same time, a dialog is displayed encouraging passersby to interact. The visualization is primarily meant to draw the attention of passersby as well as to communicate the interactivity of the system. This relates to existing interaction models (Vogel & Balakrishnan, 2004; Michelis & Müller, 2011) as well as to issues such as display blindness (Müller et al., 2009) and interaction blindness (Ojala et al., 2012), as discussed in Chapter 2.

When passersby step close enough to the display, a three-dimensional information cube is opened for them to interact with (Figure 9B). The system supports two simultaneous users, in which case the user interface is scaled and another information cube is opened for the new user (Figure 9C.) Supporting multiple users was one of the early key design decisions for many reasons. First, enabling multiple simultaneous users maximizes the *honeypot* effect (Brignull & Rogers, 2003). Second, users interacting in groups tend to interact longer (Perry et al. 2010). Third, the lack of multi-user support may end some users' interaction prematurely, as single users who are part of a group may halt the interaction due to a social pressure from other party members (Hardy et al., 2011).

Figure 9. Information Wall 1. **A)** A user is tracked with a rectangular shape on the screen. An instruction dialog is displayed. **B)** A user interacts with an information cube. **C)** Two users interact with the wall simultaneously (Mäkelä et al., 2014b, Figure 1, © 2014 ACM).

Content is laid out on the sides of the cube, and the cube can be rotated in four directions to navigate the content. Rotating the cube up and down moves between sections, and rotating left and right cycles between pages within a particular section. Elements on the active page can be opened for more information. There are four content sections: daily menus for a nearby student cafeteria, public events from the university's public calendar, the latest news from a news portal, and a section providing more information on the university premises.

The user's hands are represented on the screen as hand-shaped cursors that move according to where the user is pointing. Interactive elements are triggered by moving the cursor atop the desired element and keeping the cursor steady for a short time. This interaction method is called *dwelling* (or point-and-dwell). During dwelling, a circular animation is overlaid on the active target to illustrate that it is being triggered.

Dwelling was chosen as the target selection method due to its intuitiveness and ease of use (Hespanhol et al., 2012; Walter et al., 2014), which we considered especially important in the serious, public context the Information Wall was designed for, as most users would be first-timers and are likely to have little knowledge of gestural interaction. Moreover, we hypothesized that a cursor-based design would be easier to understand due to its familiarity and would contribute to communicating the system's interactivity (Ojala et al., 2012). All interactions are designed to be performed with one hand, which follows the recommendations by Hardy et al. (2011), Walter et al. (2014), and Ackad et al. (2015). However, to enable flexible interactions, both hands can be used if the user so chooses, in which case both hands are depicted on the screen.

The information cube can be rotated by first triggering an element on the edge of the cube using dwelling, after which the cube can be rotated by swiping to the opposite direction. During rotation, the cube rotates in real time according to the user's swiping speed and position. Rotation can be canceled by swiping back to the starting point.

Technical Details

Information Wall 1 is developed on .NET Framework 4.0 using the Windows Presentation Foundation (WPF) library and the 3DTools plugin. The system utilizes the Microsoft Kinect for Windows sensor for tracking users' location and hand coordinates. For the data exchange between the Kinect and the Information Wall application, a dedicated socket-based middleware component called SkeletonServer was built.

The majority of the Information Wall's content is fetched from external sources using public APIs and RSS feeds, although some content is parsed from external websites.

Deployments

Information Wall 1 was publicly deployed for one year from April 2013 to April 2014 in the main building of the University of Tampere. The deployment space was chosen as it was a large, open space with two major landmarks, a cafeteria and a large auditorium (Figure 10). The cafeteria functioned as a popular meeting area and workspace for students, staff, and visitors. The auditorium was frequently used for lectures and exams. Both landmarks ensured a stable flow of foot traffic within the deployment space.

Figure 10. The deployment space of Information Wall 1 at a university campus from the deployment's perspective (Mäkelä et al., 2017c, Figure 4, © 2017 Taylor & Francis Group, LLC).

During the public deployment, we conducted occasional observations and gathered skeletal and interaction data of all passersby, which resulted in a data set of more than 100,000 passersby. This data was utilized in our semi-automated analysis, in which we studied how extensive data sets could be analyzed programmatically to evaluate the usage of the display without the prohibitive use of resources (Mäkelä et al., 2017c).

5.2 Information Wall 2

Based on lessons learned with Information Wall 1, we developed a new system, Information Wall 2, from scratch that utilized different design approaches. Content on the display is laid out in horizontally separated sections. Users activate sections of the wall by positioning themselves in front of an interesting section, and then use mid-air gestures to interact with the content (Figure 11).

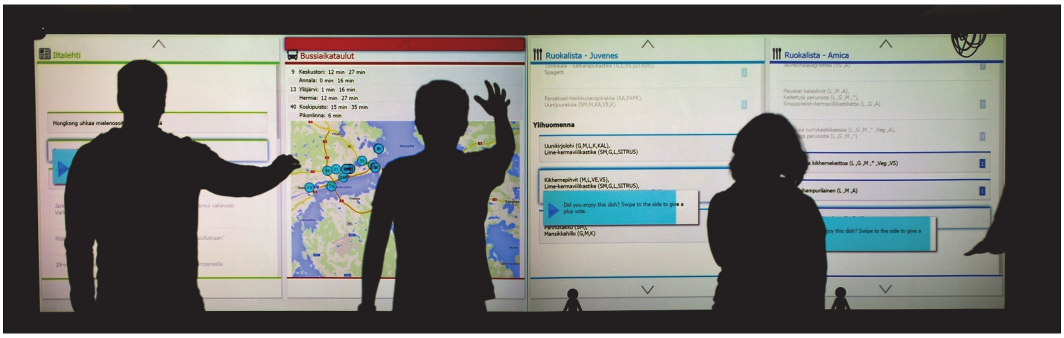


Figure 11. Information Wall 2.

When users have positioned themselves before the desired section, they interact with the content with a simple vertical elevation of the hand and sideways swipes. For example, to open up a news article, the user raises a hand to match the vertical positioning of the corresponding article and swipes to the right. The opened article can be closed by swiping back to the left.

The motivation for this design was to make interaction more efficient and to maximize the use of the available screen space. In the first Information

Wall, content was laid out on the sides of a 3D cube, meaning most of the content was hidden during interaction and hidden completely when no one was using the system. In the Information Wall 2's design, content is much more visible, therefore being potentially useful also to passersby who do not interact but glance at the screen as they walk past. This design also better communicates a summary of the available content. Moreover, with the position-based activation of content sections, we hypothesized that this would make interaction more effortless, as users do not have to navigate to the desired content onscreen, but simply stop at the corresponding location.

The content consists of context-relevant and timely information like the latest news, the lunch menus of nearby cafeterias, and real-time bus schedules. Like the previous Information Wall, the system supports multiple simultaneous users, though the Information Wall 2 supports running multiple instances with multiple Kinect sensors. One of the deployed versions of the system was run on two large parallel screens, allowing four simultaneous users.

Technical Details

The Information Wall 2 is a web application that runs on all modern browsers and is built using HTML, CSS, and Javascript (and the jQuery library).

Since the Information Wall 2 is a web application, we extended the existing SkeletonServer (used for the Information Wall 1) to support HTTP calls that would return responses in JSON format. Therefore, Kinect data could be fetched from the browser using HTTP calls and parsing the returned JSON data. This same approach was also used for the following two systems, the NFC Information Wall and SimSense.

Deployments

The Information Wall 2 was deployed in two spaces. First, it was deployed in the same space as the original Information Wall for three months and run with the same technical setup.

In addition, we ran the Information Wall 2 in SimSpace, which is a relaxed workspace and breakroom within the premises of the School of Information Sciences at the University of Tampere (Figure 12). It was run on two parallel screens using two Kinect sensors, therefore supporting four simultaneous users (Figure 11). The deployment in SimSpace was not officially regulated, as the space was actively used for myriad purposes, including lectures, during which the large screens were needed for other reasons. However, whenever the screens were free, Information Wall 2 was started.

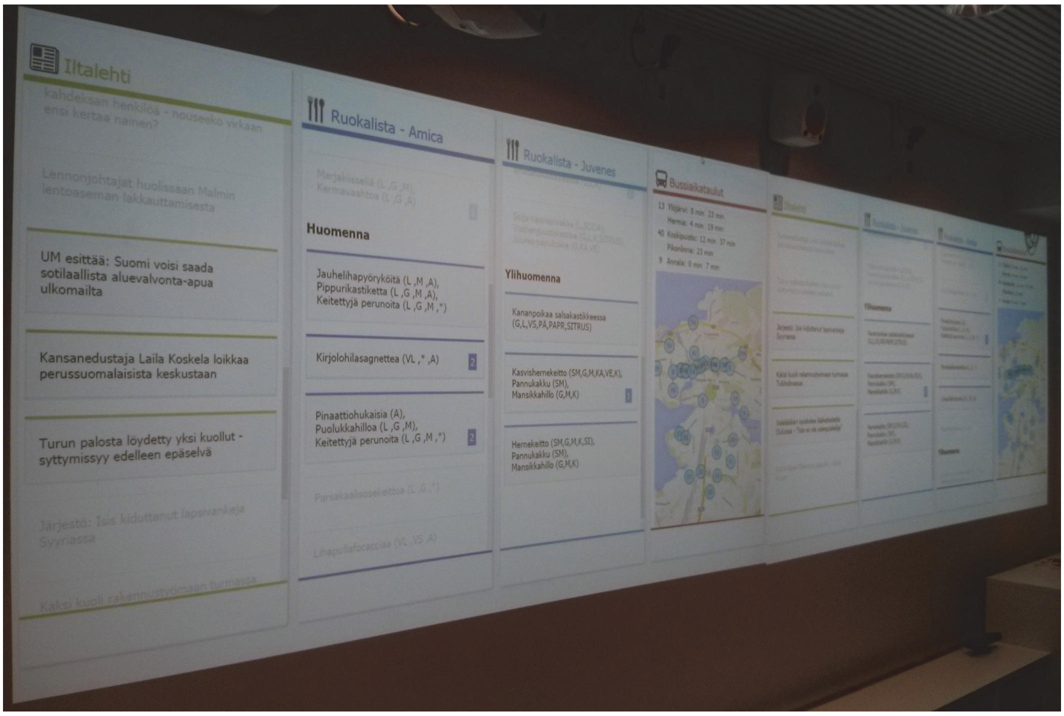


Figure 12. The Information Wall 2 running on two large screens deployed in SimSpace, an open office workspace.

5.3 NFC Information Wall

With NFC Information Wall, we extended the existing Information Wall 2 architecture and created a mobile application to allow users to transfer content between the large display and personal mobile devices.

Figure 13. NFC Information Wall. **Left:** The user is selecting an item on the large display. **Middle:** The user is transferring the selected item to his mobile device via the NFC reader. **Right:** The transferred content is ready to be explored on the mobile device (adapted from Mäkelä et al., 2016, Figure 1, © 2016 ACM).

Users can select items on the screen using point-and-dwell, and they can transfer the selected item to their mobile device by tapping the nearby NFC reader (Figure 13). Content items are laid out horizontally, four being visible at once. Users can cycle through content pages using the previous and next buttons located on the bottom corners of the screen with point-and-dwell. Content items are shown differently based on their nature; say, items that showcase an event with a date and time are depicted as calendar markings.

The mobile application is only mandatory if the user wants to create and upload items on the screen. The application presents a simple form through which users can add basic information, such as a title and a description, and optional information such as a date or a picture. When an item has been created, users can tap the NFC reader to upload it to the screen, which immediately updates to show the newly uploaded item.

Content transferred from the screen to the mobile device is presented differently based on whether the mobile application is installed. If the application is not installed, a link is sent to the mobile device that immediately opens on the mobile's default browser. If the app is installed, the content is opened inside the app in a scrollable list.

Technical Details

The software consists of the large display client, a web service, an Android client, the SkeletonServer, and another middleware component to handle data between the NFC reader and the web service. The NFC reader used with the system is an ACS ACR122 reader.

The large display client is a web-based application and runs on top of the Information Wall 2 architecture. A web service stores and provides the large display content. Another middleware component handles messages between the NFC reader and the web service. The large display client fetches the content directly from the web service, and the web service allows content moderation via a separate admin component.

Deployments

The NFC Information Wall was briefly deployed in the same two locations as the Information Wall 2. In the main building of the university campus, the NFC Information Wall was run for one week, during which time a researcher was present to demo the system and to gather initial feedback regarding the system and the concept.

The NFC Information Wall was deployed in SimSpace for a user study, in which we further investigated the user experience and interaction of content transfer using mid-air gestures and a physical transfer device (Mäkelä et al., 2016). This study is described in detail in Chapter 6.

5.4 SimSense

SimSense is the final, most advanced prototype that was built to carry out the research presented in this thesis. SimSense is a smart space system that automatically pairs mobile devices in the space with their owners based on location data and enables seamless transfer of content from a large display to a mobile device using mid-air gestures (Figure 14).[1] Thanks to automatic pairing, the mobile device can remain in one's pocket or bag throughout the interaction process, and no connection to the display needs to be established.

The system consists of a large display application, a mobile application, a Microsoft Kinect sensor, and five Kontakt.io Bluetooth beacons attached to the ceiling of the installation space. The Kinect sensor tracks people in space, whereas mobile devices are tracked using Bluetooth low energy (BLE) via the five Bluetooth beacons. The mobile device measures the strength of the Bluetooth signals received from these beacons with a dedicated Android application. The signals are called RSSI (Received Signal Strength Indicator) and are measured in dBm. The mobile application then sends this information to a location estimation server, which then estimates the location of the mobile device in the space. We base our work on the assumption that a mobile device belongs to a person whose location matches the mobile device's. Therefore, when a user transfers content from the large display, SimSense knows which device to send the content to.

The large display application functions as a gesture-controlled information display. Like Information Wall 1, hand cursors are displayed onscreen to visualize where the user is pointing. Content items can be opened for closer examination with the point-and-dwell method, that is, by pointing at an element for a short time. A progress bar is displayed during dwelling to visualize the selection process. Content feeds can be cycled through via two buttons at the bottom of the screen, similarly triggered using point-and-dwell. The contents consist mainly of news articles from numerous popular news portals (Figure 15).

To transfer content to their mobile device, users first point at an item (that is, hover over it with the hand cursor), and then form a fist with the pointing hand to "grab" the item, and pull it towards themselves (Figure 16A). When the hand is pulled back, the transfer of content is triggered, and the content is immediately sent to the recipient device. This gesture, which we call *grab-and-pull*, was chosen based on the results of a user study, in which we compared it to another gesture, *grab-and-drop* (Figure 16B), for content retrieval and measured the overall user experience of the system (Mäkelä et al., 2017). The user study will be presented in more detail in Chapter 6.

[1] https://www.youtube.com/watch?v=RkpjCsNBu3U

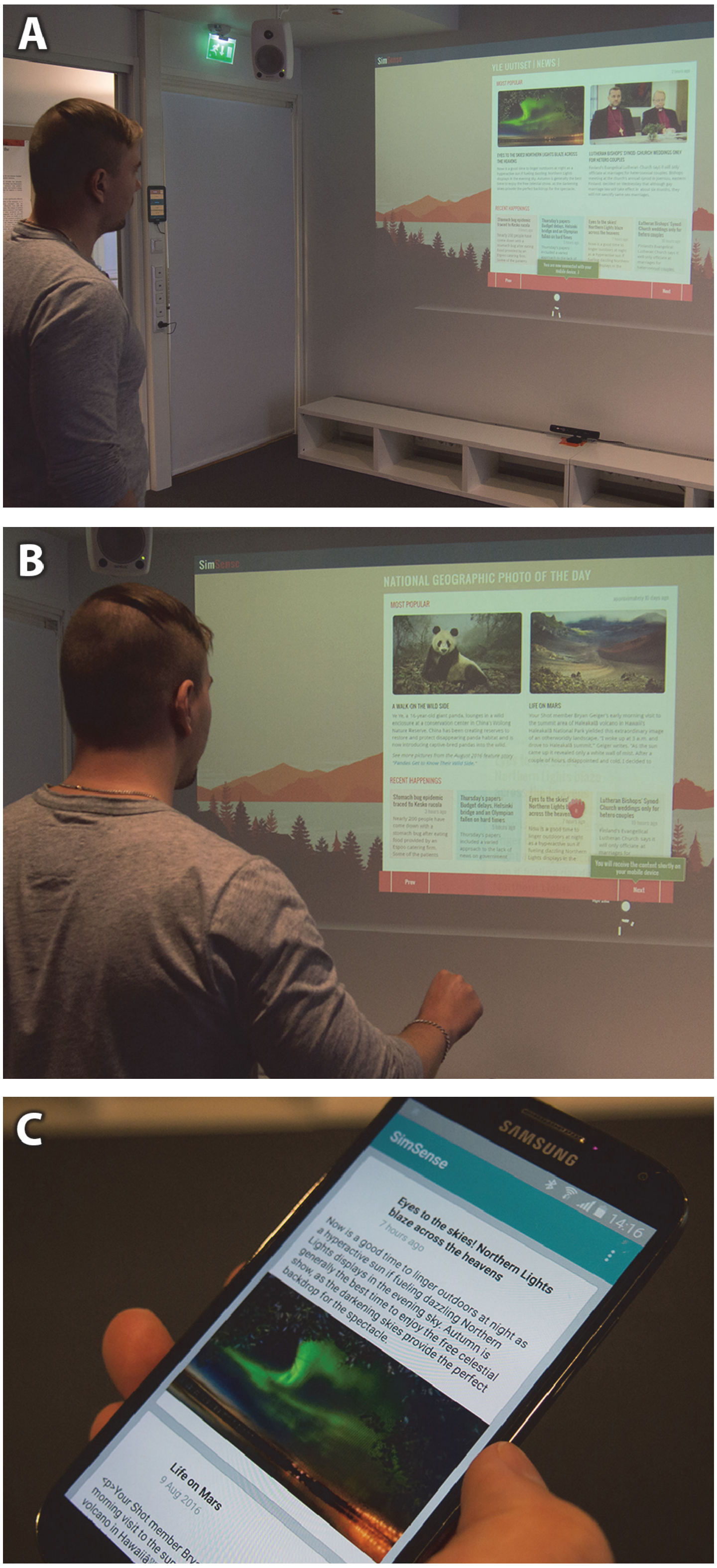


Figure 14. The SimSense system. **A)** A user enters the space, and the system immediately pairs the user with his mobile device based on location data. The system notifies the user that they have been paired. **B)** The user can then interact with the large display via mid-air gestures and transfer content to his mobile device. Here, the user has grabbed an interesting news article and is pulling it towards himself to trigger the transfer. **C)** The user can explore the transferred content on his mobile device.

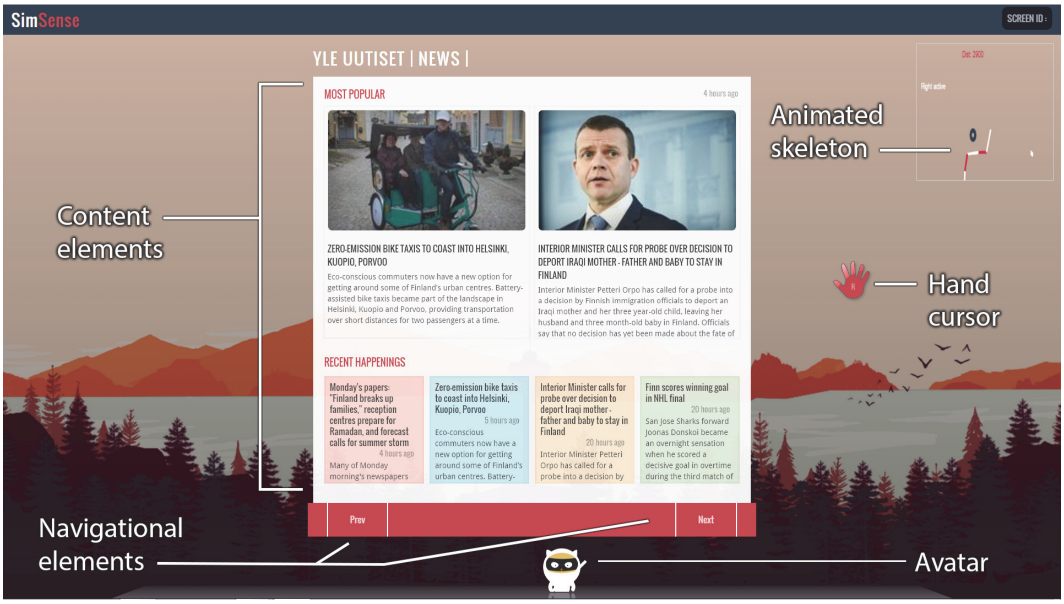


Figure 15. The SimSense screen elements overview (Mäkelä et al., 2017b, Figure 1B, © 2017 IEEE).

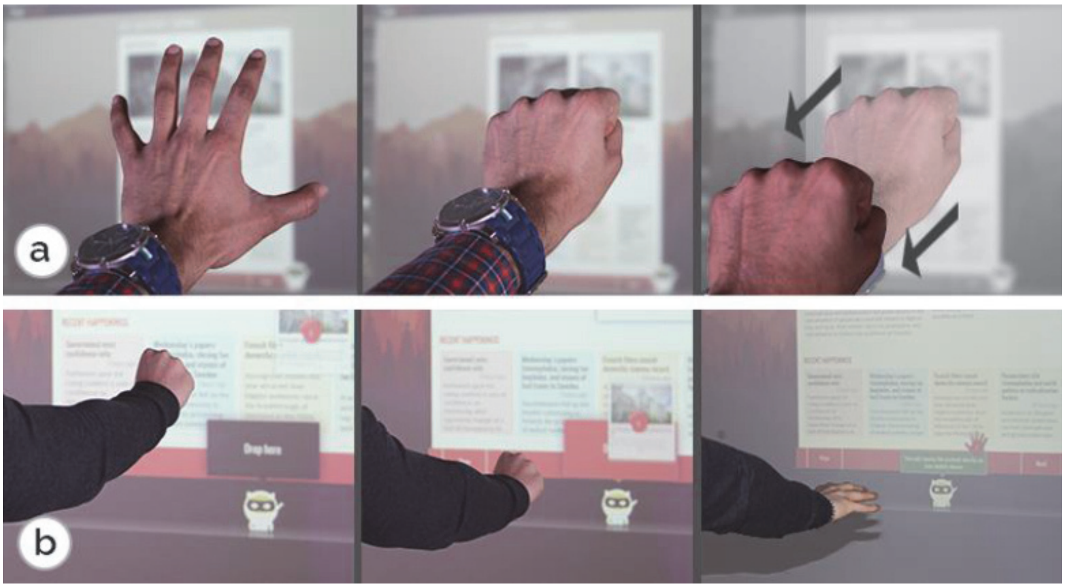


Figure 16. Two different gestures for content transfer were developed and evaluated. **A) Grab-and-pull**: the user points at a content item, grabs the item, and pulls the hand towards his or her body. **B) Grab-and-drop**: the user points at a content item, grabs the item, and drags it to the drop area (Mäkelä et al., 2017b, Figure 2AB, © 2017 IEEE).

One area of focus with SimSense was providing extensive feedback to the user. During each phase of the grab-and-pull gesture, small hand animations are shown to visualize the next step. Furthermore, while pulling, the grabbed element gets larger the further it is pulled to create a sense of the content being pulled out of the screen.

A skeletal representation mimicking the user's movement is displayed on the top right of the screen (Figure 15). On the bottom edge of the screen, an avatar representing the user is displayed that moves horizontally according to the position of the user and acts as an instructional agent. The avatar displays a “You are now connected!” notification when a user has entered the space and has been successfully paired with his or her mobile device. Likewise, a notification is displayed when content is transferred to the mobile device. All interactions are accompanied with sound effects.

On the mobile application, when content is received, the device vibrates and plays an audio tone, and an Android notification is created. Pressing the notification opens the SimSense mobile app, which shows the content. The mobile application displays all transferred content in a scrollable list (Figure 14C). Users can log in to the application with an existing Gmail account.

Technical Details

The SimSense large display web application is built using the AngularJS framework. The gesture recognition uses the same technology as the Information Wall 2. Otherwise, the SimSense client is developed from scratch.

For user–mobile pairing, a location prediction server was written using Node.js. Five Kontakt.io Bluetooth beacons are attached to the ceiling of the space to enable the location estimation of mobile devices. A kNN machine learning library (k = 3) is used to estimate the location of mobile devices.

Deployments

Since finishing the first version of SimSense, it has been semi-permanently deployed in SimSpace, much like Information Wall 2 was earlier. SimSense would run in the space most of the time outside official events, during which the large screen was needed for other purposes. In the same space, a user study was run regarding interaction with SimSense (Mäkelä et al., 2017b), which will be presented in the next chapter.

6 Introduction to the Publications

Three prominent themes arise through this research, the backgrounds for which were discussed in chapters 2, 3, and 4. Each of the six publications contribute to one of the themes. I will next present an overview of the themes and the research contributing to them, after which I will describe each publication in detail.

Theme 1: Deployment and Evaluation of Interactive Displays. In this theme, I focus on challenges and shortcomings regarding the deployment and evaluation of interactive displays. Publications I and II contribute to this theme. In Publication I (Mäkelä et al., 2017a), I investigate *external factors* around and within public spaces that affect interactive display deployments. I recognize six types of external factors and find that their effect on deployments is often surprising and negative. In Publication II (Mäkelä et al., 2017c), I study the usage of a semi-automated process to evaluate the usage of and behavior around public displays. This study stems from the fact that interactive display research in the field often requires considerable resources, for example, one or more researchers must be deployed in the field to observe the display's usage. In the study, I show that the semi-automated process can reach viable results while saving resources, especially in long-term deployments.

Theme 2: Interaction with Mid-air Gestures. After studying the deployment and evaluation of interactive displays, I move to study the interaction with large displays using mid-air gestures. Publications III and IV contribute to this theme. In Publication III (Mäkelä et al., 2014a), I propose and evaluate a target-acquisition technique called the *magnetic cursor*. The technique was designed to alleviate the physical stress and imprecision inherent in freehand pointing interfaces, wherein users interact with elements on the screen by controlling a cursor. The magnetic cursor sticks to interactive elements from further away, making it easier to acquire the desired target. In Publication IV (Mäkelä et al., 2014b), I presented and evaluated the Information Wall. We conducted a lab study wherein users tried out the system, and we additionally deployed the system in demo sessions and events for visitors to try out freely. Based on the study results as well as our observations, we provided guidelines for designing gesture-controlled systems.

Theme 3: Content Transfer between Interactive Displays and Mobile Devices Using Mid-air Gestures. After more general-purpose research of mid-air gestures in theme 2, I move on to my final research area, which centers on the transfer of content between public and personal devices using mid-air gestures. Publications V and VI contribute to this theme. The key difference from theme 2 is the advancement of gestures beyond single-screen interactions. In Publication V, I present and evaluate the NFC Information Wall, which combines mid-air gestures and a physical NFC reader to enable the transfer of content between the display and a personal mobile device. The findings from this study partially guided the design of the next and final system, SimSense. In Publication VI, we utilize the SimSense system to evaluate the transfer of content from a large display to a personal device using mid-air gestures. SimSense materializes a more advanced vision wherein users do not need to interact with their mobile device at all to transfer content to it. In this final study, we show that mid-air gestures for this purpose are efficient and offer an impressive user experience.

6.1 Publication I: Challenges in Public Display Deployments: A Taxonomy of External Factors

Reference

Mäkelä, V., Sharma, S., Hakulinen, J., Heimonen, T., & Turunen, M. (2017a). Challenges in Public Display Deployments: A Taxonomy of External Factors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '17)*, 3426–3475. New York, NY, USA: ACM. doi:10.1145/3025453.3025798

Objective and Methods

Deploying interactive displays in the wild is often more difficult than one might expect. After conducting several public display deployments, we came to the realization that various *external factors*—properties and phenomena unrelated to the deployment—often affect the deployment in a surprising way. For example, people in the space might interfere with the display, or weather issues like rain might destroy a piece of hardware.

To investigate external factors on a large scale, we conducted a literature review to search for issues caused by external factors and looked into our own deployment experiences to seek similar problems. Then, we combined the issues from the two sources and used affinity diagramming to categorize them. Our aim was to address the following questions: (1) Which types of external factors affect public deployments? And (2) How do researchers react to the effects of external factors?

For the literature review, we first used snowball sampling to gather an extensive set of papers. We started with a set of papers from well-known authors with whom we were already familiar due to our own work in the field. Subsequently, we checked the references of these publications to find additionally relevant papers. The relevance of the papers was decided based on their title or what was said about them in the citing paper. Then, we used Google Scholar to locate new publications that cited the papers in the collection. From the search results, we added papers based on their title or their abstract. After this process, our collection included 132 publications. We then narrowed down the collection by including papers that either presented a) an in-the-wild public display deployment or b) challenges or lessons learned regarding public deployments. After filtering, our final collection consisted of 71 publications.

Two researchers then reviewed all 71 publications and noted every issue, challenge, or observation from which an external factor could be identified. We did not include issues that were internal to the deployment due to, for example, bad design or poorly performing technology. From the 71 publications, 22 contained one or more issues due to external factors, combining into 51 issues. We then identified 10 issues from our own deployments and added them to the issues from the literature, forming the final data set of 61 issues.

We employed affinity diagramming to categorize the issues. Three researchers took part in this process: the two who conducted the original literature review and an additional researcher. Issues from both sources were treated equally, with no special emphasis on our own experiences. We considered the categorization complete when all three researchers agreed with the result.

Results and Discussion (Publication I)

Using the procedure described above, we identified six categories of external factors: *weather*, *events*, *surroundings*, *space*, *inhabitants*, and *vandalism*.

Weather refers to effects caused by multitudinous climate-related factors, such as sunlight, cold, wind, rain, ice, and humidity. For instance, sunlight can interfere with various sensors, rain can destroy hardware, and cold weather may make users unwilling to interact with an outdoor display.

Events refers to gatherings such as festivals that temporarily affect the deployment space or surrounding areas. For instance, events can bring unusually large numbers of people to the area, which may overcrowd the system or the interaction space. However, this may also lead to positive effects, such as more interest in the deployment and more users.

Surroundings refers to the areas around the deployment space, such as nearby buildings and their functions. For instance, a theater beside the deployment space may affect the general demographics, how much of a hurry the passersby are in, and when they are interested in the display.

Space refers to the physical properties and functions within the immediate deployment area. For instance, the layout of the area and the positioning and angle of the display may affect its visibility.

Inhabitants encapsulates the daily roles and routines of people frequenting the deployment space, such as guards, janitors, and students. For instance, in our deployments, janitors occasionally left their carts in such a way that they blocked part of the screen.

Vandalism refers to intentional interference with the deployment and destruction of hardware. For instance, a vandal may break the deployment's display, which leads not only to significant expenses, but also to downtime for the system while the display is being fixed.

We also assessed how current research reacts to, or deals with, the observed issues and identified four levels of reactions. Of the 61 issues, 39% were *ignored*. Another 39% were *adapted* to, and 13% were *solved*. The remaining 8% were *embraced*. *Ignoring* means taking no action against the issue. *Adapting* means taking some action against the issue while introducing a tradeoff. For example, if dealing with an issue requires continuous use of resources, it would count as adapting. *Solving* refers to a permanent solution, after which the issue no longer exists. *Embracing* means taking advantage of the observed effect or simply seeing the effect as positive.

Our results are useful to researchers and practitioners in various ways. First, we show that the effects from external factors are almost always negative, highlighting the need to identify potential threats to the deployment in advance and being prepared for surprising effects. Second, our taxonomy of external factors helps in the identification of potential threats in a step-by-step manner by providing six distinct categories from which these threats can arise. We furthermore provide an extensive set of exemplar issues within each category. This knowledge is also helpful in the pre-deployment phase, as practitioners can now more easily identify the most optimal locations for their deployments. We present detailed recommendations for preparing for and dealing with the effects from external factors in Mäkelä et al. (2017a).

6.2 Publication II: Semi-Automated, Large-Scale Evaluation of Public Displays

Reference

Mäkelä, V., Heimonen, T., & Turunen, M. (2017c). Semi-Automated, Large-Scale Evaluation of Public Displays. *International Journal of Human–Computer Interaction*. doi:10.1080/10447318.2017.1367905

Objective and Methods

Public interactive displays have been widely researched in a plethora of contexts. As such displays have become increasingly common, public display research has moved from the lab into the wild where, for example, the display's usage and the behavior of passersby can be studied in a realistic environment.

Research in the field has commonly utilized a mix of different methods, most notably observations, interviews, and interaction logs. The primary method in past research has been observations. While observing users and passersby presents a multitude of advantages and is a straightforward study method, it often consumes a considerable amount of resources. Especially when studying the long-term usage of a display, deploying one or more researchers in the field for such an extended period is not desirable—and often impossible. At the same time, interaction logs and depth sensor data have been mostly used as supporting data for observations and interviews. Moreover, logged data has been analyzed manually in many cases. This practice is similarly time-consuming, and we assume it overlooks much of what could be potentially learned from the data.

Therefore, we set out to investigate whether we could utilize logged interaction and skeletal data more extensively to reach meaningful evaluation results, without having to be present to observe the users.

We deployed one of the gesture-controlled information displays presented earlier in this thesis, the Information Wall, in a large open space in the main building of a university campus for one year. The system collected all interactions as well as anonymized skeletal data of all passersby in separate log files. By *skeletal data*, we mean a collection of 3D points of every passersby's upper body joints, which were collected every 60 milliseconds.

Our one-year deployment resulted in a data set containing data of 106 637 passersby. We then wrote a collection of scripts to clean up, combine, and finally, transform the log files into CSV form with dozens of variables on which statistical analyses could be conducted.

Results and Discussion (Publication II)

The main result of our work is the semi-automated process that can be used to gather, transform, and analyze data from a public display's deployment. We recognized four distinct phases to the process: (1) *data collection*, (2) *preparation*, (3) *feature extraction*, and (4) *analysis*. By *data collection*, we refer to the automatic gathering of interaction and skeletal information during the deployment. *Preparation* refers to the process of combining and cleaning up the extensive amount of data, to get the data ready for the next phase. *Feature extraction* is the most crucial and unique phase to this process. The ultimate result from feature extraction is data that is transformed into such a form (e.g., a CSV file) on which statistical analysis can be performed in the A*nalysis* phase to answer deployment-specific research questions.

For the feature extraction phase, we wrote scripts to transform the data into a large CSV file containing dozens of variables for each passerby, which were ultimately used to reach meaningful findings regarding our public display deployment. To produce the required variables, one needs to carefully consider the characteristics of the installation, how it is interacted with, and how the space around it might affect its use. A more detailed discussion of this process is presented in Mäkelä et al. (2017c).

We applied the semi-automated process to our Information Wall deployment. Although the main result of this work is the process itself, I briefly present some of the results from applying this process to demonstrate the findings that can be reached.

We were able to categorize passersby into direct, subtle, and passive users. We somewhat followed the Audience Funnel framework, wherein user engagement is classified into six different categories (Michelis & Müller, 2011). Although not all of these categories could be recognized through logged data alone, we were nonetheless interested in whether we could classify passersby based on whether they interacted with the display, and if so, how much and for how long they interacted. Out of all passersby, we recognized 1,489 as direct users (1.4%) and 6,241 as subtle users (5.9%), leaving the remaining 98,907 passersby as passive users (92.8%).

In analyzing individual and group behavior, we found that 45.8% of direct users interacted without anyone else in the scene. Of the direct users, 18.5% were accompanied by one or more passive users, that is, people who did not interact at all. In addition, 35.7% were accompanied by subtle or other direct users, or a combination of all three user types.

Direct users who were alone interacted for an average of 55 seconds and triggered 5.8 elements on the screen. Direct users who had company interacted much longer, with an average of 103 seconds and 7.8 element triggers. The highest increase in interaction time was observed in larger groups that mixed all three user types, with an average duration of 153 seconds and 12.4 element triggers.

As presented earlier in this thesis, the Information Wall exhibits a two-level reaction to passersby. Distant passersby are depicted as transparent rectangles on the screen, whereas people stepping closer to the screen get assigned their personal information cube with which they can interact. The rectangles are meant to catch their attention; the rectangles are not shown if someone is already interacting with an information cube so the rectangles do not distract direct users. Our analysis showed that, of those passive users who were accompanying direct users, the clear majority (81.7%) stood far from the screen, so an information cube was not opened to them. This could suggest a phenomenon we call *multi-user interaction blindness*, wherein the people observing a direct user do not interact themselves simply because they are not aware the system supports multiple simultaneous users.

Finally, we analyzed how the number of subtle and direct users develops over time. As expected, the amount of users dropped significantly during the first few months. We were nonetheless surprised by the very high percentage of users in the first month, wherein 25.7% of all passersby were subtle users, and 4% were direct users. In the second month, the figures were 14.5% and 3.1% respectively, and in the third month, 13.4% and 3.4%, which we also consider high numbers. However, in the following months, subtle users dropped to around 3–4%, and direct users dropped below 1%.

While our analysis results indicate many things, two main implications can be deduced. First, considering long interaction times as well as the significant decline in usage rates over time, it seems that most users of the system were first-time users. In other words, not many returned to the system, which could suggest that the Information Wall suffered from usability issues or that people simply did not want to interact with gestures after the initial experience. Second, we found that direct users were often accompanied by passive users who stood further away and did not interact at all. This could suggest that passersby were simply not aware the system supports multiple simultaneous users. A concrete implication from this is that multi-user systems should still, in some way, react to passive users even when someone is already using the system.

Through this work, we proposed and demonstrated a semi-automated process to evaluate public displays. Our main goal with this process is to offer an alternative to existing research methods, namely, to conduct long-term evaluations without having to send researchers into the field. This process is not entirely self-sufficient and requires a mix of other research methods, such as observations, to be efficient. Nevertheless, the process can still save a considerable amount of resources. In addition, as the process is applied to more and more deployments, the process also becomes increasingly straightforward and automated, as many high-level aspects of the process can be directly applied to any deployment, requiring only minor changes in between.

6.3 Publication III: Magnetic Cursor: Improving Target Selection in Freehand Pointing Interfaces

Reference

Mäkelä, V., Heimonen, T., & Turunen, M. (2014a). Magnetic Cursor: Improving Target Selection in Freehand Pointing Interfaces. In *Proceedings of the International Symposium on Pervasive Displays (PerDis '14)*, 112–117. New York, NY, USA: ACM. doi:10.1145/2611009.2611025

Objective and Methods

Freehand pointing interfaces, wherein users point from a distance at interactive targets on the screen, can be tiring and slow. Keeping the arm extended and steady in mid-air may result in fatigue, especially in prolonged interactions (Hincapié-Ramos et al., 2014; Jang et al., 2017). Because the hand must remain steady also while interacting with the target, focus and steadiness are required beyond merely the initial acquisition of the target.

To alleviate this problem, we investigated assisting techniques that make target acquisition easier. While past research has evaluated and proposed many such techniques, they have been evaluated in different contexts, using modalities other than mid-air gestures.

Based on existing research as well as our experiences with freehand pointing interfaces, we recognized three properties that should be considered when designing target acquisition techniques for freehand pointing. First, it may not be desirable to use techniques that significantly alter the visual space of the interface by, for example, manipulating the size of targets close to the cursor. Doing so imposes limitations on the interface and may appear odd, especially to first-time users in public spaces. Second, because the location of the onscreen cursor is directly mapped to the location of the pointing hand in 3D space, the cursor's location cannot be dramatically manipulated (slight changes, however, may increase performance). Third, gesturing requires constant effort, meaning pointing at the screen, unlike, say, mouse or touch interaction, where staying on target requires little to no effort. Therefore, the assisting technique should not only make *acquiring* targets easier, but also make it easier to *stay* on target.

Per these definitions, we produced the concept of a *magnetic cursor*, which combines characteristics from several existing techniques. Most notably, it a) jumps to the closest target when a certain threshold is crossed, and b) it moves slower while on target, but does not remain entirely stationary, thereby giving the user hint regarding the actual pointing location of their hand. In short, the magnetic cursor increases the size of interactive targets in motor space, while aiming to affect the visual space only slightly.

We conducted a 19-participant lab user study following a within-subjects design. Participants first used a cursor with no assisting techniques enabled to serve as a practice session. Subsequently, participants used and evaluated two different versions of the magnetic cursor (referred to as MC1 and MC2) as well as the snap-to-target method proposed by existing research (Parker et al., 2005). The two magnetic cursors utilized different thresholds, that is, how strong the magnetic effect is, with the strength of 100 and 200 pixels.

Employing all the described techniques, the participants' task was to select a target element on the screen using point-and-dwell, with a dwell time of 1.2 seconds. The target was always surrounded by distractor elements that were visually dissimilar from the target. Independent variables were (1) the techniques, (2) the distance to the target from the previous target, (3) the distance between the target and the distractors, and (4) the size of the target and the distractors.

We measured the objective performance of each technique, including, for example, time to acquire the target, time to select the target, number of reacquisitions, and error rate. Additionally, participants filled out a subjective questionnaire regarding their experience with each technique.

Results and Discussion (Publication III)

Our results showed that all assisting techniques, unsurprisingly, significantly outperformed the normal cursor. MC1 was the fastest in terms of selection time (1.6 s), followed by MC2 (1.7 s) and snap-to-target (1.8 s). MC1 also had the fewest errors (2.4%), followed by snap-to-target (3.2%) and MC2 (8.6%). As an error, we counted the selection of a distractor element instead of the target.

Of subjective measures, the three assisting techniques were mostly evaluated as equal with no significant differences. However, MC2 was found to be more mentally demanding and present less control than MC1. MC1 and snap-to-target were also perceived as faster than MC2. Nine out of 19 participants chose MC1 as the most pleasant technique, and another nine preferred snap-to-target, while MC2 was preferred by only one participant. Conversely, eight participants named MC2 as their least favorite technique, and another eight preferred the normal cursor the least. Snap-to-target received three votes, while nobody preferred MC1 the least.

Our results are useful in the design of freehand pointing interfaces. We proposed the Magnetic Cursor, which was found to be an effective assisting technique for freehand pointing, making target selection faster and less cumbersome. However, we also discovered that, if the effect of the magnetic cursor is too strong, it can become frustrating and result in more errors. Thus, the strength should be relative to the screen size and resolution. In our study, with a full HD display, the magnetic cursor with an effect of 100 pixels was the most efficient, and was also preferred by users. The effect of 200 pixels was more error-prone, and users liked it significantly less. In objective measures, the snap-to-target performed almost as well as MC1, but was still slightly slower and slightly more error-prone. Similarly, MC1 received slightly higher user experience and preference scores than snap-to-target.

6.4 Publication IV: Information wall: Evaluation of a Gesture-Controlled Public Display

Reference

Mäkelä, V., Heimonen, T., Luhtala, M., & Turunen, M. (2014b). Information Wall: Evaluation of a Gesture-Controlled Public Display. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia (MUM '14, Melbourne, Australia)*, 228–231. New York, NY, USA: ACM. doi:10.1145/2677972.2677998

Objective and Methods

Interactive displays are regularly interacted with through touch input. Interaction through mid-air gestures, however, has been very rare. While mid-air gestures present many challenges, they also present a number of advantages over touch input and other modalities. Hence, they provide interesting opportunities for research in the public space.

It is still relatively unclear how gestural interactions should be designed, especially for public spaces, and how people in such spaces react to and interact with gesture-controlled displays.

To investigate these aspects in more detail, we conducted a 19-participant lab user study, wherein participants used and evaluated the Information Wall system presented earlier in this thesis. Moreover, we deployed the Information Wall in many public events in which we had the opportunity to observe first-time users of the system in a free-form manner.

In the user study, participants were given simple tasks without further instructions, that is, they discovered the system's range of functionalities on their own. For instance, the Information Wall contained a public university calendar with information on public lectures. One of the participants' tasks was to seek a specific lecture and check the location of its lecture hall. We used the SUXES method (Turunen et al., 2009) to measure the users' expectations and experiences with the Information Wall, as well as the AttrakDiff (2018) questionnaire. AttrakDiff measures a system's appeal in two primary categories: pragmatic quality (PQ) and hedonic quality (HQ). PQ mainly measures the system's usability, while HQ measures properties such as novelty and how users generally experience the system.

Results and Discussion (Publication IV)

Based on the SUXES results, the participants' expectations of the system were mostly met. While this is a promising result, it is interesting to note that their expectations were rather low to start. This might be due to a certain level of skepticism towards mid-air gestures and the related technology. The AttrakDiff results show that the Information Wall received an average score in PQ and a slightly above-average score in HQ. This primarily suggests that there is room for improvement, especially concerning the system's usability. However, the positive result is that users seemed to value the novelty and new experiences provided by the Information Wall.

Connecting our findings from the user study as well as the observations, we produced five guidelines to help with the design of gesture-controlled information displays. A more detailed discussion and arguments for the guidelines are presented in the paper (Mäkelä et al., 2014b), though two guidelines warrant explicit discussion here because they relate strongly to existing research:

First, we recommended that gesture-controlled information displays should *support simultaneous interaction*. During the public events, we observed that users preferred interacting with a known associate. Moreover, visitors were much more likely to interact with the system after the first few had the courage to try interacting with the display. While this recommendation is not new per se and follows the honeypot effect (Brignull & Rogers, 2003; Hardy et al., 2011; Müller et al., 2012), our observations suggest that this effect is enhanced by gestural interaction. Because users interact with the system from a distance, bystanders can easily observe how the system works. Moreover, even during interaction, users can observe how the other person (perhaps a more experienced user) is interacting.

Second, we recommended that only *one hand cursor per user should be depicted* at a time. Other research has pointed out that one-handed interaction should be supported (Hardy et al., 2011; Walter et al., 2014; Ackad et al., 2015). We also made the same observation and completely supported one-handed interactions. However, we introduced a flexible solution to allow users to utilize both hands if they wanted, or change the interacting hand on the fly. Therefore, both hands would be portrayed onscreen if they were lifted. This feature seemed to work against itself for two reasons. First, users had surprising trouble recognizing which hand cursor was mapped to which hand, even though the cursors had a slightly different visual appearance. Second, first-time users especially often unknowingly interacted with the "other" hand, leading to confusing situations. Most users did not realize this until it was pointed out to them. These observations suggest that stronger visual feedback should be used, for example, in the form of skeletal representations, as we did with the gesture-controlled systems that followed the original Information Wall.

6.5 Publication V: Investigating Mid-Air Gestures and Handhelds in Motion Tracked Environments

Reference

Mäkelä, V., Korhonen, H., Ojala, J., Järvi, A., Väänänen, K., Raisamo, R., & Turunen, M. (2016). Investigating Mid-Air Gestures and Handhelds in Motion Tracked Environments. In *Proceedings of the 5th ACM International Symposium on Pervasive Displays (PerDis '16)*, 45–51. New York, NY, USA: ACM. doi:10.1145/2914920.2915015

Objective and Methods

Researchers have proposed and studied numerous ways of moving content between personal mobile devices and public devices like interactive displays. However, these methods generally employ touch as the interaction modality, either through touching the public device (i.e., a screen), the recipient mobile device, or both. To our knowledge, no such work has been done with mid-air gestures, wherein users could, for example, select items they want to transfer to their mobile device by using gestures from a distance. In such scenarios, the relationships between different devices and mid-air gestures are unclear, for example, whether a mobile device affects how users interact with gestures.

Consequently, we conducted a 13-participant user study, wherein participants used and evaluated the NFC Information Wall presented earlier in this thesis. Our primary research questions were: (1) How does a mobile device affect mid-air gestures? (2) How is the mobile device held during gesturing and when the user does not intend to interact with any device? (3) How is a gesture-controlled large display affected by the presence of another interactive device in the same space?

Participants carried out six tasks using the NFC Information Wall smart space system. In three tasks, participants sought a specific item on the screen and transferred it to their mobile device to investigate the content. In the remaining three tasks, participants created different kinds of content items on their smartphones and uploaded them onto the public screen.

Results and Discussion (Publication V)

When interacting with the display, we observed that most participants preferred to gesture with one hand and hold the mobile device in the other (8 out of 13 participants). Two participants left the phone on a table for the duration of the interaction and picked it up again when they needed it. Two participants used the hand holding the mobile device to gesture, and one participant frequently switched between the aforementioned methods.

A surprising observation during interactions with the display was that participants were holding the mobile device relatively high, at chest or shoulder height. This occasionally made it very difficult for the system to figure which hand should be tracked for interaction, resulting in, for example, errors and accidental interactions. However, in their feedback, participants did not think that holding the mobile device disturbed their gestures, suggesting they did not realize that the mobile device actually had an impact.

Another interesting observation was what we call the *freezing effect*. Six out of 13 participants frequently "froze" when they wanted to pause interaction to observe the newly displayed content on the screen, for example. Users would first point at a screen to trigger whichever functionality they wanted, but when they wanted to stop, they would still continue pointing, triggering further (unwanted) interactions.

Similar issues continued when users were interacting with the mobile device, that is, either observing received content or creating new items for the screen. Users were holding the mobile device relatively high, close to their face, which often triggered accidental interactions on the display. The freezing effect was similarly apparent in these cases. When interacting with the handheld device, users would lift their heads when they realized that the screen was reacting to them, but chose to remain still as if assuming the display would eventually understand that the user is not interacting. Intuitively, no users tried to do anything else to stop unwanted interaction, such as lower their hands or turn away from the screen.

Many of our observations are closely tied to the Midas touch issue (Kjeldsen & Hartman, 2001). The Midas touch issue refers to the challenge of recognizing whether the user wants to interact with the system. This issue is prominent in modalities that exhibit an always-on nature, such as mid-air gestures, speech, and eye gaze, wherein any kind of movement (or speech) can be falsely interpreted as input by the system.

The Midas touch issue is well-known from existing research. We also encountered this issue in our previous studies, such as during the public deployments of the Information Wall (Mäkelä et al., 2014b). However, in this study, the issue occurred significantly more frequently and rather dramatically affected how easily participants could complete their tasks. This result shows that cross-device interaction in motion-tracked smart spaces, wherein users may handle multiple devices, is problematic and not only presents unique challenges, but also reinforces existing challenges.

The main implication from this study is that holding the mobile device affects users' stance and pose, and it may interfere with gestural interaction in such a way that users themselves do not realize it. This presents a particular challenge for content transfer between public and personal devices when mid-air gestures are involved. This problem can be solved with a more advanced and innovative solution in which users can keep their mobile device in their pocket or bag throughout the interaction process, as presented next in the final publication of this thesis.

6.6 Publication VI: "It's Natural to Grab and Pull": Retrieving Content from Large Displays Using Mid-Air Gestures

Reference

Mäkelä, V., James, J., Keskinen, T., Hakulinen, J., & Turunen, M. (2017b). "It's Natural to Grab and Pull": Retrieving Content from Large Displays Using Mid-Air Gestures. *Pervasive Computing, 16*(3), 70–77. doi:10.1109/MPRV.2017.2940966

Objective and Methods

In the final publication of my thesis, I move towards my vision of utilizing mid-air gestures for seamless cross-device interaction, in particular, for transferring content from public interactive displays to personal mobile devices.

In the previous study, we used an NFC reader to transfer content between the screen and the mobile device, and we found the combination of mid-air gestures, mobile device interaction, and NFC interaction problematic. For this study, we built the SimSense system, presented earlier in this thesis, which allows users to keep the mobile device in their pocket throughout the interaction process by automatically pairing users with their mobile devices based on location data.

Despite some existing research regarding gestural interaction and interface design, we are not aware of any other existing work that utilized mid-air gestures for seamless content transfer. It is unclear how users experience mid-air gestures for this purpose, and how such interactions should be designed.

Therefore, we conducted a 12-participant user study. Subjects used the SimSense system to carry out content retrieval tasks using two gestures, *grab-and-pull* and *grab-and-drop*. With both gestures (in a counterbalanced order), the participants' task was to repeatedly transfer a highlighted content item from the screen to the mobile device that was placed in their pocket. The highlighted item was chosen randomly and was visualized with thick, red borders. All participants performed 20 content retrieval tasks with both gestures.

We measured the objective performance of both gestures via interaction logs, namely, their speed and error rate. In addition, participants filled out their expectations and experiences regarding both techniques using a modified version of the SUXES method (Turunen et al., 2009), which was tailored to suit mid-air gestures in particular. Prior to the tasks, we showed very short video clips to participants demonstrating both gestures, based on which they filled out the forms describing their expectations. After the tasks, participants filled out the SUXES experience questionnaire as well as a questionnaire comparing the techniques. Subjects also answered questions in an interview.

Results and Discussion (Publication VI)

Based on the SUXES results, expectations were very high for both gestures. We hypothesized that this would be the case, given the rather novel and, in a sense, futuristic context in which we utilized mid-air gestures. We also assumed that various sci-fi movies and other entertainment, which occasionally exhibit futuristic gesture-controlled interfaces, may play a role in people's elevated expectations. This is especially evident when comparing the expectations of one of our earlier prototypes, the Information Wall (Mäkelä et al., 2014b), the expectations of which were very low.

Regardless of high expectations, a very promising result is that we could primarily meet and, in some cases, even exceed them. In comparing the two gestures directly, *grab-and-pull* received higher median scores in 10 out of 14 statements included in the SUXES questionnaire, whereas *grab-and-drop* scored higher in the remaining four statements.

The grab-and-pull also performed significantly faster, only taking an average 0.65 seconds to complete from when the gesture was started. Grab-and-drop took an average of 1.99 seconds to complete. This was unsurprising, as *grab-and-pull* only exhibits a one-dimensional pull after the target element has been grabbed, whereas with *grab-and-drop*, the grabbed element needs to be dragged and dropped onto a specified area on the screen.

Users made almost no errors at all. With *grab-and-pull*, only one user transferred the wrong content item once, and with *grab-and-drop*, this happened only eight times in total. When measuring the number of unfinished gestures (i.e., when a gesture was started but not correctly finished), a relatively comparable amount was found with both techniques. There was an average of 3.0 unfinished gestures with *grab-and-pull* and 2.75 unfinished gestures with *grab-and-drop* throughout the whole study session. This measurement seemed to be more related to the performance and accuracy of the Kinect sensor, though, as it occasionally did not correctly track the status of the active hand, sometimes requiring participants to start over.

Grab-and-pull was preferred over *grab-and-drop* for transferring content from a distance, as 10 out of 12 participants preferred *grab-and-pull*. However, despite this clear preference, it is notable that *grab-and-drop* still received highly positive feedback. Moreover, in addition to the two content transfer gestures, we utilized the point-and-dwell method for other, more conventional interactions with the screen, such as for moving from one content feed to another. We did not observe a single instance in which these mechanics interfered with each other. This is a promising result, as it suggests that gestural interfaces could be employed for more advanced and multifaceted interactions than previously thought. *Grab-and-drop* could be simply used for another purpose, such as transferring content from one display to another, while using *grab-and-pull* for personal content transfer and point-and-dwell for, say, navigational interactions.

Both gestures and the SimSense system in general received highly positive and seemingly enthusiastic feedback. Both techniques scored a median value between 6 and 7 in all SUXES statements (1 being the lowest value and 7 the highest). In addition, participants gave positive comments in the interview, for instance, making remarks about how the interaction felt as if they were in a sci-fi movie.

Our results show promise for mid-air gestures in ubiquitous computing and cross-device interactions through three main findings. (1) We show that mid-air gestures are well suited for content transfer scenarios and provide a highly positive user experience. (2) Distinct gestures, such as *grab-and-pull* and *grab-and-drop*, can be successfully mixed with point-and-dwell interactions, making way for advanced, versatile gesture interfaces. (3) *Grab-and-pull* was found to be an ideal gesture for transferring content to a personal device, whereas *grab-and-drop* could be utilized for another purpose, such as transferring from one display to another.

7 Discussion

In this thesis, I investigated three themes revolving around interactive displays, mid-air gestures, and seamless content transfer. In this section, I discuss the results from each theme and tie together the three themes.

Deployment and Evaluation of Interactive Displays

RQ1: *What external factors affect interactive displays, and how can long-term evaluations of interactive displays be streamlined without considerable use of resources?*

I provided answers to this research question through Publications I and II (Mäkelä et al., 2017a; Mäkelä et al., 2017c).

In Publication I, I investigated *external factors* that have an effect on a public display deployment, either directly or indirectly. This was carried out as an extensive literature survey, complemented by our own experiences from several public deployments. Through this work, I identified six categories of external factors: *weather*, *events*, *surroundings*, *space*, *inhabitants*, and *vandalism*. I found that the effects from all these factors on the deployment are almost always negative and frequently surprising. I furthermore provided examples of the issues experienced in each category and provided guidelines to help recognize and deal with potential issues.

The work in Publication I is valuable for designers and practitioners conducting public display deployments. I raise awareness of external factors and show their (negative) effects on public display deployments, which should motivate others to aim for the pre-emptive recognition of potential threats to their deployments and to reserve resources for dealing with them. The examples and guidelines provided in this work offer concrete ways of recognizing and dealing with potential issues in a step-by-step manner.

In Publication II, I proposed and evaluated a semi-automated process to evaluate the usage of public interactive displays by utilizing logged movement and interaction data of passersby. Past research has mainly utilized on-site observations and lightweight log data analysis to evaluate public displays, which is time-consuming and not always possible, especially for long-term deployments.

The proposed process identifies four phases: *data collection*, *preparation*, *feature extraction*, and *analysis*. I applied this process to the logged data collected during a one-year deployment of the Information Wall, containing the data of more than 100,000 passersby, to evaluate the feasibility of the proposed process in practice. Accoredingly, I was able to extract valuable information about the Information Wall's usage as well as passerby behavior. Of particular interest was that many users were accompanied by people who stood back to observe the interaction, but did not interact themselves. This finding would not have been achieved through traditional interaction logs. Moreover, how the usage of the display develops over time was easy to analyze from the logs. During the first three months, the usage rates were very high. Only from the fourth month onwards did they remain somewhat stable—and much lower than in the first months.

Through the proposed semi-automated evaluation process I, in a sense, extend the research space of public displays, providing more options for researchers when they decide which research methods to utilize in their deployments. Based on the results, it seems that the semi-automated evaluation process can indeed streamline the evaluation of public displays and is helpful especially in a large-scale evaluation of a display's long-term usage. Our results further show that it took more than three months for the display's usage to stabilize, highlighting the fact that short-term evaluations are unlikely to paint a realistic picture of the usage of a display. This factor should motivate researchers to conduct more long-term evaluations.

Publications I and II contribute to the overall domain of interactive displays especially in public spaces, assisting researchers and practitioners in conducting successful deployments as well as providing tools for the evaluation of interactive displays. The external factors I systematically investigated in Publication I have not received much attention in the past. Prior research has reported experiences and lessons regarding problems with public deployments (Harris et al., 2004; Storz et al., 2006; Dalsgaard & Halskov, 2010), but none of them has focused on the *source* of these issues. Similarly, to the best of my knowledge, this work is the first to delve into how researchers *reacted* to the issues they encountered.

Publication II offers a new method for evaluating interactive displays. Earlier, Williamson and Williamson (2014) and Elhart et al. (2017) presented tools for pedestrian and passerby analytics. I, on the other hand, synthesize a combination of passerby movement as well as interaction with the display, resulting in more opportunities for an in-depth evaluation of the display.

It should be noted that I do not limit myself to public spaces when researching gestural interaction and content transfer in the next two themes. Nonetheless, in researching the deployment and evaluation of interactive displays in this theme, it made sense to do it in the public context. I argue that the findings in this theme apply to other, more closed environments as well: the challenges in more controlled environments are simply perhaps less prominent. However, in contrast, if the work in this theme was done in a non-public context, many aspects that are potentially involved in both the deployment and evaluation of interactive displays would have been overlooked.

Additionally, I argue that the work done in this theme is especially useful for systems using motion tracking. First, regarding external factors, a considerable number of issues had to do with the performance of sensors, on which gestures are particularly dependent. Moreover, I also observed that the novelty of gestures resulted in various surprising issues, more so than more traditional modalities. Therefore, the taxonomy of external factors and their implications should be particularly valuable to those who plan to deploy gesture-controlled systems in the wild.

The presented semi-automated process is also especially well-suited for motion tracking systems, as the movement and positioning of passersby may also convey meaning. As such, this creates the opportunity to evaluate, for instance, how (well) users react to certain position-triggered visual cues. Therefore, this process can provide much richer insight into how users perceived and used a gesture-controlled system. Moreover, there is the matter of convenience. The necessary hardware and much of the programming logic concerning the tracking of users and passersby are already in the system due to the chosen method of interaction. Simply storing this (anonymous) data for later analysis is trivial.

Finally, understanding and researching the public space was crucial in understanding interactive systems as a whole. It helped to identify what kinds of spaces such interactions might fit in and what kinds of challenges one might expect to encounter. Currently, people can be hesitant to interact with technology in public (Ojala et al., 2012; Deterding et al., 2015). As we found in Publications II and IV, this may be especially true for gestural interaction. In a way, regarding gestural interaction and cross-device communication, I see public spaces as the final step in a long process that is still ahead of us, wherein public spaces are made truly ubiquitous, with interactive devices and their information readily accessible from anywhere in the space.

Interaction with Mid-Air Gestures

RQ2: *What techniques and design principles should be used for gesture-controlled information displays to provide an efficient and usable experience?*

I provided answers to this research question through Publications III and IV (Mäkelä et al., 2014a; Mäkelä et al., 2014b).

In Publication III, I evaluated the performance of a custom-made technique, the *magnetic cursor*, designed specifically to improve target acquisition in freehand pointing interfaces. Pointing in such a manner can be demanding, requiring mental and physical focus. Moreover, the physical demand can easily lead to the *Gorilla arm* syndrome (Hincapié-Ramos et al., 2014; Jang et al., 2017).

Based on the results, the magnetic cursor makes interaction faster and easier, reducing the mental load and physical precision required for successful interaction. Moreover, the magnetic cursor was favored over an unassisted cursor as well as a competitive assisting technique, *snap-to-target* (Parker et al., 2005). While the error rates were slightly higher with both assisting techniques than with an unassisted cursor, the error rates were nonetheless low across the board. In summary, target acquisition techniques should be used in freehand pointing interfaces, as they result in a variety of positive effects, with little to no drawbacks.

In Publication IV, I presented an evaluation of the Information Wall. This work was carried out as a combination of a 19-participant lab study, in which subjective feedback was gathered, and demo sessions, in which groups of people with various backgrounds could engage the system with little to no prior knowledge, offering us the opportunity to observe first-time users in a realistic setting.

Based on the results, I derived five guidelines to help with the design of gesture-controlled, cursor-based information displays. In addition, I found that the system and gestural interaction in general gathered substantial interest from the people to whom it was introduced. I also observed uncertainty, as people were somewhat hesitant to test the system, especially when others were watching. The same finding was later presented by Gentile et al. (2017). Fortunately, I believe that supporting two simultaneous users played a crucial part in overcoming the initial barrier for interaction, as interacting with a known associate seemed to greatly reduce the threshold. After this initial barrier was broken, the honeypot effect (Brignull & Rogers, 2003) took effect, motivating increasingly more people to interact.

Both of these publications contribute to the overall design space of gesture-controlled systems, more precisely in systems that incorporate freehand pointing and illustrate it via an onscreen cursor. Publication III evaluated and proposed a suitable targeting assistance technique. To my knowledge, it is the first to conduct such a study in the context of device-free freehand pointing interfaces. This is important, as freehand pointing differs, for instance, in terms of accuracy, hand jitter, and fatigue, from the contexts wherein such techniques have been previously evaluated (e.g., Parker et al., 2005; Grossman & Balakrishnan, 2005; Ahlström et al., 2006; Bateman et al., 2013). Publication IV presented guidelines for designing gesture-controlled interfaces. Other research has also provided guidelines in a similar context (Walter et al., 2014; Ackad et al., 2015; Ackad et al., 2016). This is only a positive, as the guideline sets do not contradict, but complement each other, focusing on different aspects of and approaches to gestural interaction and system design. N.b. that the guidelines provided by other researchers appeared in the same year or later as Publication IV.

Towards Cross-Device Interaction Using Mid-Air Gestures

RQ3: *Can mid-air gestures be employed to enable seamless, low-effort exchanging of information between displays and mobile devices, and what is the user experience of such interactions?*

I provided answers to this research question through Publications V and VI (Mäkelä et al., 2016; Mäkelä et al., 2017b).

In Publication V, I investigated a combination of mid-air gestures and a proxemic device (NFC reader) for two-way content transfer between a public display and a mobile device. With this approach, I identified many challenges. The *Midas touch* issue (Kjeldsen & Hartman, 2001), which is common in gestural interaction, was particularly problematic when interaction with a mobile device as well as a proxemic device was brought into the mix. Nonetheless, despite the shortcomings of the system, the general concept and gestural interaction received positive feedback and was seen as having much potential.

In Publication VI, the challenges identified in the previous study were largely overcome, even though we only focused on one-way content transfer from the display to the mobile device. This was partly thanks to our innovative user-mobile pairing, wherein we used location data to automatically recognize which mobile device belonged to which person. Consequently, users could pull content from a distance to their mobile devices, without any kind of setup phase, while keeping the mobile device in their pocket. I am not aware of any other work that enables similarly seamless content transfer scenarios.

With this approach, the main finding was that mid-air gestures are well suited for seamless content transfer, and offer an impressive and efficient user experience. Another positive finding was that SimSense also allowed the exploration of content on the display using mid-air gestures, and we successfully mixed interaction mechanics for both exploration and content transfer. That is, we used point-and-dwell and separate content transfer gestures in the same interface with no issues. This is encouraging and suggests that mid-air gestures have the potential for more advanced and intuitive interactions than what has been currently proposed.

I believe the highly positive user experience of using mid-air gestures for content transfer can and will apply on a higher level, hopefully making way for even more advanced interaction scenarios in which users can freely interact with and transfer content between a multitude of devices in the space. As motion recognition and other technology advance, increasingly more opportunities for utilizing such interactions will become available.

One realization during my work on this research question was that seamless cross-device content transfer using gestures is not *only* about gestures, but also about creating innovative solutions to inherent challenges and obstacles. In this case, with the SimSense prototype, such a solution was the automatic pairing approach, wherein we used Bluetooth and skeletal data to automatically recognize which mobile device belonged to which person. It was thanks to this innovation that the vision of transferring content seamlessly with mid-air gestures was possible in the first place. In essence, ubiquitous computing is not just about designing novel interactions; it is about designing whole *platforms* in which these interactions can be carried out, bringing forth an array of entirely new possibilities. While other technical solutions have been presented to identify the owners of mobile devices (Rofouei et al., 2012; Wilson & Benko, 2014), they have not been used for such purposes.

Future Work

While the three themes I researched in this dissertation are closely related, their relationship still leaves room for future work. To my knowledge, I am the first to present seamless content transfer methods that allow content to be transferred with mid-air gestures while the recipient device remains in a pocket or bag. Simultaneously, mid-air gestures as an interaction modality are still not widely adopted. Therefore, it would be worthwhile to investigate how such interactions perform and how they are perceived in different settings—public, semi-public, and others. Of particular interest should be the *acceptance* of such interactions and the related technology, accompanied with any related concerns, such as privacy issues.

The design of seamless cross-device interactions as well as the spaces in which they are performed will likely become more complicated as more opportunities are taken advantage of, that is, when more and more devices are brought into the mix. This is when solutions will become even more important such as that by Schwarz et al. (2014), wherein they utilized a combination of body pose, gaze, and gestures to ascertain whether the user intends to interact with a system. With a multitude of devices available for interaction, however, the problem will extend beyond the intent to interact, as we must also effectively ascertain which device the user aims to address. Solutions could be provided by extending current solutions, such as that by Schwarz et al. (2014), or by investigating effortless ways for the user to communicate which device they want to interact with.

I conclude the discussion with a positive outlook on future advancements. Many of this work's aspects were affected by the state of current technology, which still introduces a number of limitations affecting, for example, what kinds of gestural interactions are possible and how stable their performance is. The primary limitations stem from sensing technologies, particularly motion-tracking sensors, which a) require a direct line of sight to the user, b) do not always track users accurately, and c) have a limited field of view. However, my point here is not to bash current technology—quite the contrary. Even *with* these limitations, I achieved promising results regarding the use of mid-air gestures in ubiquitous computing, especially with respect to seamless cross-device interaction. Imagine, then, what could be achieved with more advanced technology. Even now, practical improvements are available. For instance, current off-the-shelf smart watches can be used to track hand movements, removing the need for external sensors (Katsuragawa et al., 2016; Pietroszek et al., 2017; Siddhpuria et al., 2017). Such advancements bring forth further possibilities, allowing the integration of gestural interaction into a variety of environments. Another direction that can be taken is to investigate different modalities for similarly seamless interactions, as we have already done in a follow-up study (Mäkelä et al., 2018), in which we use the SimSense technology to develop and evaluate seamless content transfer techniques using not only mid-air gestures, but also touch, eye gaze, and a multimodal combination of eye gaze and mid-air gestures.

8 Conclusion

In this dissertation, I investigated (1) the deployment and evaluation of interactive displays, (2) the interaction with, and the design of, mid-air gestures, and (3) the seamless transferring of content from interactive displays to personal mobile devices using mid-air gestures. To carry out this research, I developed four gesture-controlled applications and deployed them in various locations to study their usage as well as people's perception of and experience with them.

This research consisted of six original publications. The conclusions from each are as follows:

- Many external factors exist that predominantly have a negative effect on interactive displays. These factors can be placed into six categories: weather, events, surroundings, space, inhabitants, and vandalism. Gesture-controlled applications are particularly susceptible to external factors, as they rely on sensors and direct line of sight. With the help of this categorization and the accompanying guidelines and real-world examples, researchers and practitioners can more easily identify external factors and either deal with them prematurely or devise a plan about how to adapt to the possible effect(s), should issues arise.

- Evaluation of interactive displays largely relies on field observations. While observations can provide deep insights into the usage of the display, they are time-consuming and often unrealistic for long-term evaluations. However, the evaluation process can be, to an extent, automated by utilizing log data gathered during the deployment, consisting of interaction data as well as movement data of users and passersby. This data can be analyzed to reach valid findings and is especially useful for studying long-term usage, as well as phenomena that rarely occur (e.g., studying how specific types of groups use the system).

- In freehand pointing interfaces, the interaction with onscreen elements can be difficult and tiring, as users need to point at the desired target accurately and keep their pointing hand stable. This problem can be alleviated with a targeting assistance technique called the *magnetic cursor*, which makes interaction faster, improves the user experience, and reduces physical fatigue.

- Users in public spaces tend to be first-time users and may be unsure about interacting with novel interaction modalities. Simultaneous interactions should be supported, as users are often more comfortable interacting together. One-handed interactions should be supported. Due to a lack of tactile feedback and known conventions, special emphasis should be placed on clear, dynamic visual feedback.

- Using mid-air gestures for conventional means of content transfer between public and personal devices is problematic. Combining mid-air gestures with mobile interactions as well as proxemic devices, such as NFC readers, results in an emphasized Midas touch issue. This suggests that mid-air gestures are better suited for quick, one-way transfers, and that innovative solutions should be sought wherein users would not have to manipulate a proxemic device or the recipient device to transfer content.

- Mid-air gestures can be leveraged to allow efficient one-way content transfer from public to personal devices, resulting in an impressive user experience. This is primarily due to being able to seamlessly pull content from a public device from a distance. Such interactions are dependent on innovative solutions, such as those that automatically recognize the mobile device of the user, allowing the user to keep the mobile device in a pocket or bag during the interaction.

The findings have three primary implications. First, the work on external factors as well as the semi-automated evaluation process of interactive displays contribute to the success of future interactive display deployments, especially gesture-controlled systems, offering researchers and practitioners tools to avoid problems with their deployments as well as more options to evaluate their usage. Second, the guidelines and techniques presented in this work help in the design of usable gesture-controlled interfaces. Third, this work motivates researchers and practitioners to utilize the strengths of mid-air gestures in a variety of new contexts, especially for cross-device scenarios such as content transfer.

References

Ackad, C., Wasinger, R., Gluga, R., Kay, J., & Tomitsch, M. (2013). Measuring Interactivity at an Interactive Public Information Display. In *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration (OzCHI '13)*, 329-332. New York, NY, USA; ACM. doi:10.1145/2541016.2541091

Ackad, C., Clayphan, A., Tomitsch, M., & Kay, J. (2015). An in-the-wild study of learning mid-air gestures to browse hierarchical information at a large interactive public display. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*, 1227-1238. New York, NY, USA: ACM. doi:10.1145/2750858.2807532

Ackad, C., Tomitsch, M., & Kay, J. (2016). Skeletons and Silhouettes: Comparing User Representations at a Gesture-based Large Display. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*, 2343-2347. New York, NY, USA: ACM. doi:10.1145/2858036.2858427

Ahlström, D., Hitz, M., & Leitner, G. (2006). An evaluation of sticky and force enhanced targets in multi target situations. In *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles (NordiCHI '06)*, 58-67. New York, NY, USA: ACM. doi:10.1145/1182475.1182482

All about UX. (2018). All about UX: User experience definitions: http://www.allaboutux.org/ux-definitions. Referenced January 8th 2018.

AttrakDiff. (2018). AttrakDiff website: http://attrakdiff.de/. Referenced January 8th 2018.

Alt, F., Schneegaß, S., Schmidt, A., Müller, J., & Memarovic, N. (2012). How to evaluate public displays. In *Proceedings of the 2012 International Symposium on Pervasive Displays (PerDis '12)*, Article 17, 6 pages. New York, NY, USA: ACM. doi:10.1145/2307798.2307815

Alt, F., Shirazi, A.S., Kubitza, T., & Schmidt, A. (2013). Interaction techniques for creating and exchanging content with public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*, 1709-1718. New York, NY, USA: ACM. doi:10.1145/2470654.2466226

Baldauf, M., Lasinger, K., & Fröhlich, P. (2012). Real-World Drag'n'Drop – Bidirectional Camera-based Media Transfer between Smartphones and Large Displays. In *Adj. Proceedings of the 10th International Conference on Pervasive Computing*.

Bateman, S., Mandryk, R., Gutwin, C., & Xiao, R. (2013). Analysis and comparison of target assistance techniques for relative ray-cast pointing. *International Journal of Human–Computer Studies 71(5)*, 511-532. doi:10.1016/j.ijhcs.2012.12.006

Bazo, A., & Echtler, F. (2014). Phone proxies: effortless content sharing between smartphones and interactive surfaces. In *Proceedings of the 2014 ACM SIGCHI symposium on Engineering interactive computing systems (EICS '14)*, 229-234. New York, NY, USA: ACM. doi:10.1145/2607023.2610276

Bolt, R. (1980). "Put-that-there": Voice and gesture at the graphics interface. In *Proceedings of the 7th annual conference on Computer graphics and interactive techniques (SIGGRAPH '80)*, 262-270. New York, NY, USA: ACM. doi:10.1145/800250.807503

Boring, S., Altendorfer, M., Broll, G., Hilliges, O., & Butz, A. (2007). Shoot & copy: phonecam-based information transfer from public displays onto mobile phones. In *Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology (Mobility '07)*, 24-31. New York, NY, USA: ACM. doi:10.1145/1378063.1378068

Brignull, H., & Rogers, Y. (2003). Enticing people to interact with large public displays in public spaces. In *Proceedings of the 2003 IFIP TC13 international conference on Human-Computer Interaction (INTERACT'05)*, 17-24. IOS Press.

Broll, G., Reithmeier, W., Holleis, P., & Wagner, M. (2011). Design and evaluation of techniques for mobile interaction with dynamic NFC-displays. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction (TEI '11)*, 205-212. New York, NY, USA: ACM. doi:10.1145/1935701.1935743

Clinch, S., Davies, N., Kubitza, T., & Schmidt, A. (2012). Designing application stores for public display networks. In *Proceedings of the 2012 International Symposium on Pervasive Displays (PerDis '12)*, Article 10, 6 pages. New York, NY, USA: ACM. doi:10.1145/2307798.2307808

Dalsgaard P., & Halskov, K. (2010). Designing urban media façades: cases and challenges. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*, 2277-2286. New York, NY, USA: ACM. doi:10.1145/1753326.1753670

Dalton, N.S., Collins, E., & Marshall, P. (2015). Display Blindness?: Looking Again at the Visibility of Situated Displays using Eye-tracking. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*, 3889-3898. New York, NY, USA: ACM. doi:10.1145/2702123.2702150

Davies, N., Friday, A., Newman, P., Rutlidge, S., & Storz, O. (2009). Using bluetooth device names to support interaction in smart environments. In *Proceedings of the 7th international conference on Mobile systems, applications, and services (MobiSys '09)*, 151-164. New York, NY, USA: ACM. doi:10.1145/1555816.1555832

Davies, N., Langheinrich, M., Clinch, S., Elhart, I., Friday, A., Kubitza, T., & Surajbali, B. (2014). Personalisation and privacy in future pervasive display networks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*, 2357-2366. New York, NY, USA: ACM. doi:10.1145/2556288.2557287

Deterding, S., Lucero, A., Holopainen, J., Min, C., Cheok, A., Waern, A., & Walz, S. (2015). Embarrassing Interactions. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*, 2365-2368. New York, NY, USA: ACM. doi:10.1145/2702613.2702647

Elhart, I., Langheinrich, M., Memarovic, N., & Rubegni, E. (2016). A good balance of costs and benefits: convincing a university administration to support the installation of an interactive multi-application display system on campus. In *Proceedings of the 5th ACM International Symposium on Pervasive Displays (PerDis '16)*, 197-203. New York, NY, USA: ACM. doi:10.1145/2914920.2915029

Elhart, I., Mikusz, M., Mora, C.G., Langheinrich, M., & Davies, N. (2017). Audience monitor: an open source tool for tracking audience mobility in front of pervasive displays. In *Proceedings of the 6th ACM International Symposium on Pervasive Displays (PerDis '17)*, Article 10, 8 pages. New York, NY, USA: ACM. doi:10.1145/3078810.3078823

Freeman, E., Brewster, S., & Lantz, V. (2016). Do That, There: An Interaction Technique for Addressing In-Air Gesture Systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*, 2319-2331. New York, NY, USA: ACM. doi:10.1145/2858036.2858308

Gentile, V., Khamis, M., Sorce, S., & Alt, F. (2017). They are looking at me!: understanding how audience presence impacts on public display users. In *Proceedings of the 6th ACM International Symposium on Pervasive Displays (PerDis '17)*, Article 11, 7 pages. New York, NY, USA: ACM. doi:10.1145/3078810.3078822

Grossman, T., & Balakrishnan, R. (2005). The bubble cursor: enhancing target acquisition by dynamic resizing of the cursor's activation area. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*, 281-290. New York, NY, USA: ACM. doi:10.1145/1054972.1055012

Hardy, R., & Rukzio, E. (2008). Touch & interact: touch-based interaction of mobile phones with displays. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services (MobileHCI '08)*, 245-254. New York, NY, USA: ACM. doi:10.1145/1409240.1409267

Hardy, J., Rukzio, E., & Davies, N. (2011). Real world responses to interactive gesture based public displays. In *Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia (MUM '11)*, 33-39. New York, NY, USA: ACM. doi:10.1145/2107596.2107600

Harris, E., Fitzpatrick, G., Rogers, Y., Price, S., Phelps, T., & Randell, C. (2004). From snark to park: lessons learnt moving pervasive experiences from indoors to outdoors. In *Proceedings of the fifth conference on Australasian user interface - Volume 28 (AUIC '04), A. Cockburn (Ed.), Vol. 28*, 39-48. Darlinghurst, Australia, Australia: Australian Computer Society, Inc.

Hespanhol, L., Tomitsch, M., Grace, K., Collins, A. & Kay, J. (2012). Investigating intuitiveness and effectiveness of gestures for free spatial interaction with large displays. In *Proceedings of the 2012 International Symposium on Pervasive Displays (PerDis '12)*, Article 6, 6 pages. New York, NY, USA: ACM. doi:10.1145/2307798.2307804

Hincapié-Ramos, J., Guo, X., Moghadasian, P., & Irani, P. (2014). Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems (CHI '14)*, 1063-1072. New York, NY, USA: ACM. doi:10.1145/2556288.2557130

Hosio, S., Goncalves, J., Kukka, H., Chamberlain, A., & Malizia, A. (2014). What's in it for me: Exploring the Real-World Value Proposition of Pervasive Displays. In *Proceedings of The International Symposium on Pervasive Displays (PerDis '14)*, 174-179. New York, NY, USA: ACM. doi:10.1145/2611009.2611012

Huang, E.M., Koster, A., & Borchers, J. (2008). Overcoming assumptions and uncovering practices: when does the public really look at public displays. In *Proceedings of the 6th International Conference on Pervasive Computing (Pervasive '08)*, 228-243. Berlin, Heidelberg: Springer-Verlag. doi:10.1007/978-3-540-79576-6_14

Jang, S., Stuerzlinger, W., Ambike, S., & Ramani. K. (2017). Modeling Cumulative Arm Fatigue in Mid-Air Interaction based on Perceived Exertion and Kinetics of Arm Motion. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*, 3328-3339. New York, NY: ACM. doi:10.1145/3025453.3025523

Karaman, S., Bagdanov, A., Landucci, L., D'Amico, G., Ferracani, A., Pezzatini, D., & Del Bimbo, A. (2016). Personalized multimedia content delivery on an interactive table by passive observation of museum visitors. *Multimedia Tools and Applications, 75*(7), 3787-3811. doi:10.1007/s11042-014-2192-y

Karam, M., & schraefel, m. c. (2005). A Taxonomy of Gestures in Human Computer Interactions. Technical report, Electronics and Computer Science, University of Southampton, Southampton, United Kingdom.

Katsuragawa, K., Pietroszek, K., Wallace, J.R., & Lank, E. (2016). Watchpoint: Freehand Pointing with a Smartwatch in a Ubiquitous Display Environment. In *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI '16)*, 128-135. New York, NY, USA: ACM. doi:10.1145/2909132.2909263

Kjeldsen, R., & Hartman, J. (2001). Design issues for vision-based computer interaction systems. In *Proceedings of the 2001 workshop on Perceptive user interfaces (PUI '01)*, 1-8. New York, NY, USA: ACM. doi:10.1145/971478.971511

ten Koppel, M., Bailly, G., Müller, J., & Walter, R. (2012). Chained displays: configurations of public displays can be used to influence actor-, audience-, and passer-by behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*, 317-326. New York, NY, USA: ACM. doi:10.1145/2207676.2207720

Kjeldsen, R., & Hartman, J. (2001). Design issues for vision-based computer interaction systems. In *Proceedings of the 2001 workshop on Perceptive user interfaces (PUI '01)*, 1-98. New York, NY, USA; ACM. doi:10.1145/971478.971511

Langner, R., von Zadow, U., Horak, T., Mitschick, A., & Dachselt, R. (2016). Content Sharing Between Spatially-Aware Mobile Phones and Large Vertical Displays Supporting Collaborative Work. In *Collaboration Meets Interactive Spaces*, 75-96. Springer, Cham. doi:10.1007/978-3-319-45853-3_5

Mäkelä, V., Heimonen, T., & Turunen, M. (2014a). Magnetic Cursor: Improving Target Selection in Freehand Pointing Interfaces. In *Proceedings of The International Symposium on Pervasive Displays (PerDis '14)*, 112-117. New York, NY, USA: ACM. doi:10.1145/2611009.2611025

Mäkelä, V., Heimonen, T., Luhtala, M., & Turunen, M. (2014b). Information wall: evaluation of a gesture-controlled public display. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia (MUM '14, Melbourne, Australia)*, 228-231. New York, NY, USA: ACM. doi:10.1145/2677972.2677998

Mäkelä, V., Korhonen, H., Ojala, J., Järvi, A., Väänänen, K., Raisamo, R., & Turunen, M. (2016). Investigating mid-air gestures and handhelds in motion tracked environments. In *Proceedings of the 5th ACM International Symposium on Pervasive Displays (PerDis '16)*, 45-51. New York, NY, USA: ACM. doi:10.1145/2914920.2915015

Mäkelä, V., Sharma, S., Hakulinen, J., Heimonen, T., & Turunen, M. (2017a). Challenges in Public Display Deployments: A Taxonomy of External Factors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '17)*, 3426-3475. New York, NY, USA: ACM. doi:10.1145/3025453.3025798

Mäkelä, V., James, J., Keskinen, T., Hakulinen, J., & Turunen, M. (2017b). "It's Natural to Grab and Pull:" Retrieving Content from Large Displays Using Mid-Air Gestures. *Pervasive Computing, 16*(3), 70-77. doi:10.1109/MPRV.2017.2940966

Mäkelä, V., Heimonen, T., & Turunen, M. (2017c). Semi-Automated, Large-Scale Evaluation of Public Displays. *International Journal of Human–Computer Interaction*. doi:10.1080/10447318.2017.1367905

Mäkelä, V., Khamis, M., Mecke, L., James, J., Turunen, M., & Alt, F. (2018). Pocket Transfers: Interaction Techniques for Transferring Content from Situated Displays to Mobile Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '18)*. New York, NY, USA: ACM. doi:10.1145/3173574.3173709

Marshall, P., Morris, R., Rogers, Y., Kreitmayer, S., & Davies, M. (2011). Rethinking 'multi-user': an in-the-wild study of how groups approach a walk-up-and-use tabletop interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*, 3033-3042. New York, NY, USA: ACM. doi:10.1145/1978942.1979392

Memarovic, N., Clinch, S., & Alt, F. (2015). Understanding Display Blindness in Future Display Deployments. In *Proceedings of the 4th International Symposium on Pervasive Displays (PerDis '15)*, 7-14. New York, NY, USA: ACM. doi:10.1145/2757710.2757719

Michelis, D., & Müller, J. (2011). The audience funnel: Observations of gesture based interaction with multiple large displays in a city center. *International Journal of Human-Computer Interaction, 27*(6), 562-579. doi:10.1080/10447318.2011.555299

Mikusz, M., Clinch, S., & Davies, N. (2017). Design considerations for multi-stakeholder display analytics. In *Proceedings of the 6th ACM International Symposium on Pervasive Displays (PerDis '17)*, Article 18, 10 pages. New York, NY, USA: ACM. doi:10.1145/3078810.3078830

Müller, J., Wilmsmann, D., Exeler, J., Buzeck, M., Schmidt, A., Jay, T., & Krüger, A. (2009). Display Blindness: The Effect of Expectations on Attention towards Digital Signage. In *Proceedings of the 7th International Conference on Pervasive Computing (Pervasive '09)*, 1-8. Berlin, Heidelberg: Springer-Verlag. doi:10.1007/978-3-642-01516-8_1

Müller, J., Alt, F., Michelis, D., & Schmidt, A. (2010). Requirements and design space for interactive public displays. In *Proceedings of the 18th ACM international conference on Multimedia (MM '10)*, 1285-1294. New York, NY, USA: ACM. doi:10.1145/1873951.1874203

Müller, J.,Walter, R., Bailly, G., Nischt, M., & Alt, F. (2012). Looking glass: a field study on noticing interactivity of a shop window. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*, 297-306. New York, NY, USA: ACM. doi:10.1145/2207676.2207718

Müller, J., Eberle, D., & Tollmar, K. (2014a). Communiplay: a field study of a public display media space. In *Proceedings of the SIGCHI Conference* *on Human Factors in Computing Systems (CHI '14)*, 1415-1424. New York, NY, USA: ACM. doi:10.1145/2556288.2557001

Müller, J., Bailly, G., Bossuyt, T., & Hillgren, N. (2014b). MirrorTouch: combining touch and mid-air gestures for public displays. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services (MobileHCI '14)*, 319-328. New York, NY, USA: ACM. doi:10.1145/2628363.2628379

Nancel, M, Wagner, J., Pietriga, E., Chapuis, O., & Mackay, W. (2011). Mid-air pan-and-zoom on wall-sized displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*, 177-186. New York, NY, USA: ACM. doi:10.1145/1978942.1978969

Ng, P.C., She, J., Jeon, K.E., & Baldauf, M. (2017). When Smart Devices Interact With Pervasive Screens: A Survey. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 13*(4), Article 55, 23 pages. doi:10.1145/3115933

Ojala, T., Kukka, H., Lindén, T., Heikkinen, T., Jurmu, M., Hosio, S., & Kruger, F. (2010). UBI-Hotspot 1.0: Large-Scale Long-Term Deployment of Interactive Public Displays in a City Center. In *Proceedings of the Fifth International Conference on Internet and Web Applications and Services (ICIW '10)*, 285-294. IEEE. doi: 10.1109/ICIW.2010.49

Ojala, T., Kostakos, V., Kukka, H., Heikkinen, T., Linden, T., Jurmu, M., & Zanni, D. (2012). Multipurpose interactive public displays in the wild: Three years later. *Computer, 45*(5), 42-49. doi:10.1109/MC.2012.115

Parker, J.K., Mandryk, R.L., Nunes, M.N., & Inkpen, K.M. (2005). TractorBeam selection aids: improving target acquisition for pointing input on tabletop displays. In *Proceedings of the 2005 IFIP TC13 international conference on Human-Computer Interaction (INTERACT'05)*, 80-93. Heidelberg, Germany: Springer-Verlag Berlin. doi:10.1007/11555261_10

Parra, G., Klerkx, J., & Duval, E. (2014). Understanding Engagement with Interactive Public Displays: an Awareness Campaign in the Wild. In *Proceedings of The International Symposium on Pervasive Displays (PerDis '14)*, Sven Gehring (Ed.), 180-185. New York, NY, USA: ACM. doi:10.1145/2611009.2611020

Perry, M., Beckett, S., O'Hara, K., & Subramanian, S. (2010). WaveWindow: public, performative gestural interaction. In *ACM International Conference on Interactive Tabletops and Surfaces (ITS '10)*, 109-112. New York, NY, USA: ACM. doi:10.1145/1936652.1936672

Pietroszek, K., Tahai, L., Wallace, J. R., & Lank, E. (2017). Watchcasting: Freehand 3D interaction with off-the-shelf smartwatch. In *2017 IEEE Symposium on 3D User Interfaces* (3DUI ‘17). IEEE. doi:10.1109/3DUI.2017.7893335

Rofouei, M., Wilson, A., Brush, A.J., & Tansley, S. (2012). Your phone or mine?: fusing body, touch and device sensing for multi-user device-display interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*, 1915-1918. New York, NY, USA: ACM. doi:10.1145/2207676.2208332

Schmidt, D., Chehimi, F., Rukzio, E., & Gellersen, H. (2010). PhoneTouch: a technique for direct phone interaction on surfaces. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology (UIST '10)*, 13-16. New York, NY, USA: ACM. doi:10.1145/1866029.1866034

Schmidt, C., Müller, J., & Bailly, G. (2013). Screenfinity: extending the perception area of content on very large public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*, 1719-1728. New York, NY, USA: ACM. doi:10.1145/2470654.2466227

Schwarz, J., Marais, C., Leyvand, T., Hudson, S., & Mankoff, J. (2014). Combining body pose, gaze, and gesture to determine intention to interact in vision-based interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*, 3443-3452. New York, NY, USA: ACM. doi:10.1145/2556288.2556989

Seewoonauth, K., Rukzio, E., Hardy, R., & Holleis, P. (2009). Touch & connect and touch & select: interacting with a computer by touching it with a mobile phone. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '09)*, Article 36, 9 pages. New York, NY, USA: ACM. doi: 10.1145/1613858.1613905

She, J., Crowcroft, J., Fu, H., & Ho, P.H. (2013). Smart signage: A draggable cyber-physical broadcast/multicast media system. *IEEE Transactions on Emerging Topics in Computing, 1*(2), 232-243. doi:10.1109/TETC.2013.2282618

Siddhpuria, S., Katsuragawa, K., Wallace, J.R., & Lank, E. (2017). Exploring At-Your-Side Gestural Interaction for Ubiquitous Environments. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17)*, 1111-1122. New York, NY, USA: ACM. doi:10.1145/3064663.3064695

Storz, O., Friday, A., Davies, N., Finney, J., Sas, C., & Sheridan, J. (2006). Public Ubiquitous Computing Systems: Lessons from the e-Campus Display Deployments. *Pervasive Computing, 5*(3), 40-47. doi:10.1109/MPRV.2006.56

Tomitsch, M., Ackad, C., Dawson, O., Hespanhol, L., & Kay, J. (2014). Who cares about the Content? An Analysis of Playful Behaviour at a Public Display. In *Proceedings of The International Symposium on Pervasive Displays (PerDis '14)*, Sven Gehring (Ed.), 160-165. New York, NY, USA: ACM. doi:10.1145/2611009.2611016

Turner, J., Alexander, J., Bulling, A., Schmidt, D., & Gellersen, H. (2013a). Eye Pull, Eye Push: Moving Objects between Large Screens and Personal Devices with Gaze and Touch. In *Proceedings of the 2015 IFIP TC13 international conference on Human-Computer Interaction (INTERACT '13)*, 170-186. Heidelberg, Germany: Springer-Verlag Berlin. doi:10.1007/978-3-642-40480-1_11

Turner, J., Bulling, A., Alexander, J., & Gellersen, H. (2013b). Eye drop: an interaction concept for gaze-supported point-to-point content transfer. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia (MUM '13)*, Article 37, 4 pages. New York, NY, USA: ACM. doi:10.1145/2541831.2541868

Turner, J., Bulling, A., Alexander, J., & Gellersen, H. (2014). Cross-device gaze-supported point-to-point content transfer. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '14)*, 19-26. New York, NY, USA: ACM. doi:10.1145/2578153.2578155

Turunen, M., Hakulinen, J., Melto, A., Heimonen, T., Laivo, T., & Hella, J. (2009). SUXES - User Experience Evaluation Method for Spoken and Multimodal Interaction. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association (INTERSPEECH '09)*, 2567-2570.

Valkama, V., & Ojala, T. (2011). Stakeholder value propositions on open community testbed of interactive public displays. In *Proceedings of the International Conference on Management of Emergent Digital EcoSystems (MEDES '11)*, 107-113. New York, NY, USA: ACM. doi:10.1145/2077489.2077509

Vogel, D., & Balakrishnan, R. (2004). Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of the 17th annual ACM symposium on User interface software and technology (UIST '04)*, 137-146. New York, NY, USA: ACM. doi:10.1145/1029632.1029656

Vrzakova, H., & Bednarik, R. (2013). That's not norma(n/l): a detailed analysis of midas touch in gaze-based problem-solving. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*, 85-90. New York, NY, USA: ACM. doi:10.1145/2468356.2468373

Walter, R., Bailly, G., & Müller, J. (2013). StrikeAPose: revealing mid-air gestures on public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*, 841-850. New York, NY, USA: ACM. doi:10.1145/2470654.2470774

Walter, R., Bailly, G., Valkanova, N., & Müller, J. (2014). Cuenesics: using mid-air gestures to select items on interactive public displays. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services (MobileHCI '14)*, 299-308. New York, NY, USA: ACM. doi:10.1145/2628363.2628368

Williamson, J.R., & Williamson, J. (2014). Analysing Pedestrian Traffic Around Public Displays. In *Proceedings of The International Symposium on Pervasive Displays (PerDis '14)*, 13-18. New York, NY, USA: ACM. doi:10.1145/2611009.2611022

Wilson, A., & Sarin, R. (2007). BlueTable: connecting wireless mobile devices on interactive surfaces using vision-based handshaking. In *Proceedings of Graphics Interface 2007 (GI '07)*, 119-125. New York, NY, USA: ACM. doi:10.1145/1268517.1268539

Wilson, A., & Benko, H. (2014). CrossMotion: Fusing Device and Image Motion for User Identification, Tracking and Device Association. In *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14)*, 216-223. New York, NY, USA: ACM. doi:10.1145/2663204.2663270

Yoo, S., Parker, C., Kay, J., & Tomitsch, M. (2015). To Dwell or Not to Dwell: An Evaluation of Mid-Air Gestures for Large Information Displays. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction (OzCHI '15)*, 187-191. New York, NY, USA: ACM. doi:10.1145/2838739.2838819

Paper I

Mäkelä, V., Sharma, S., Hakulinen, J., Heimonen, T., & Turunen, M. (2017). Challenges in Public Display Deployments: A Taxonomy of External Factors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '17)*, 3426–3475. New York, NY, USA: ACM. doi:10.1145/3025453.3025798

Challenges in Public Display Deployments: A Taxonomy of External Factors

Ville Mäkelä[1], Sumita Sharma[1], Jaakko Hakulinen[1], Tomi Heimonen[2], Markku Turunen[1]
[1]University of Tampere, Tampere, Finland
[2]University of Wisconsin - Stevens Point, Wisconsin, USA
{ville.mi.makela, sumita.s.sharma, jaakko.hakulinen, markku.turunen}@sis.uta.fi,
tomi.heimonen@uwsp.edu

ABSTRACT
Public display deployments are often subjected to various surprising and unwanted effects. These effects are frequently due to *external factors* – properties and phenomena that are unrelated to the deployment. Therefore, we conducted a literature review within the public display domain to investigate the causes behind the reported issues. This work presents a taxonomy of external factors affecting deployments, consisting of six categories: *weather*, *events*, *surroundings*, *space*, *inhabitants*, and *vandalism*. Apart from a few positive examples, we predominantly found negative effects arising from these factors. We then identified four ways of addressing the effects: *ignoring*, *adapting, solving*, and *embracing*. Of these, ignoring and adapting are substantially more frequent responses than solving and embracing – emphasizing the need for researchers to adapt. We present real-world examples and insights on how researchers and practitioners can address the effects to better manage their deployments.

Author Keywords
Public displays; pervasive displays; challenges; taxonomy; external factors; deployments; literature review.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Public displays are actively deployed and studied in various contexts, from shopping and transportation to entertainment. Much of the current focus in public display research is on understanding user behavior and interaction, and on models describing the dynamics of users [10,23,47, 49,69] and the components of the deployment [46]. Although public display applications presented in prior research generally appear to be successful, researchers often experience various surprising challenges.

CHI 2017, May 06-11, 2017, Denver, CO, USA
© 2017 ACM. ISBN 978-1-4503-4655-9/17/05…$15.00
DOI: http://dx.doi.org/10.1145/3025453.3025798

Some past works provide retrospective guidelines based on subjective experiences with various challenges in public deployments [19,31,62]. However, a large segment of articles describing public deployments are less explicit in their reporting of these challenges. This is in part because the challenges faced are not necessarily vital to the research questions - or the "take-away message" - of their studies. Any experienced challenges are typically mentioned very briefly, for example, to explain why certain research approaches were not possible, or to provide rationale for sub-optimal design decisions that had to be made.

The work presented in this paper is based on the realization that many of these challenges (or *effects*) are due to *external factors*, i.e., they are caused by properties and phenomena that are unrelated to the deployment. The motivation for this work stems from our own experiences with various public display deployments in which we have encountered surprising issues due to external factors that have been difficult to identify and control. We believe there is now enough collective experience within the pervasive display community to shed light on the extent and type of challenges caused by external factors. This led us to conduct a literature review of reported issues to better understand their role in public display deployments. To our knowledge, we are the first to approach challenges with public display deployments from this perspective. This work addresses the following research questions:

- **RQ1:** Which types of external factors affect public deployments?
- **RQ2:** How do researchers react to the effects of external factors?

We identified and collected a set of 61 issues with public deployments caused by external factors (10 issues from our own experiences, and 51 from the literature). Then, using affinity diagramming, we categorized both the causes for the issues as well as the researchers' reactions to the issues. In addressing RQ1, we identified six categories of external factors: *weather, events, surroundings, place, inhabitants*, and *vandalism*. In addressing RQ2, we identified that researchers' reaction to these issues can be categorized as: *embracing*, *solving*, *ignoring*, and *adapting*. We found that ignoring (39%) and adapting to (39%.) issues are much more prominent than solving (13%) and embracing (8%).

The experienced issues are rarely fully solved during deployments, which we believe is because external factors are difficult or impossible to control, and their effects are often surprising. Hence, researchers must instead be prepared to adapt to such situations by adjusting the properties of the deployment. The infrequency of embracing, on the other hand, shows that such effects are most often negative, and are difficult to take advantage of. Both conclusions show the importance of recognizing external factors, and being prepared to address them.

By building on the collective experiences of the HCI community working within the public deployments domain, we provide a taxonomy of external factors. We highlight examples of each type of external factors, and provide guidelines for identifying potential factors that could affect their deployments, and describe common approaches for tackling these issues.

The paper is organized as follows. We first discuss the methodology, in which we describe the deployments from which we draw our real-world experiences, and the literature review, and the process of creating the taxonomy of external factors. Following, we present each category of external factors in detail and discuss approaches for addressing each of them. Finally, we discuss the significance of our findings.

METHODOLOGY

To systematically investigate external factors with public deployments, we gathered issues caused by external factors from our own experiences as well as from existing literature, and categorized the found issues using the affinity diagramming approach.

We only included issues wherein a clear external cause could be identified. Therefore, we did not account for internal issues, that is, issues arising from the deployment itself, due to e.g., poor design solutions.

Case Deployments

We first gathered issues from two of our own deployments. Since our focus in this paper is on external factors, and not on interaction with the installations, we present the deployments only briefly, primarily describing the deployment locations.

The *Information Wall* is a gesture-controlled public information display. We implemented two different versions, however in both versions, users navigate location-relevant content, such as local events and lunch menus of nearby cafeterias, by using mid-air gestures in front of the display (Figure 1 – B and C). The first Information Wall was deployed at a university campus for one year, from April 2013 to April 2014. The next version was deployed in the same location with identical equipment from October 2014 to April 2015. The deployment setting (Figure 1A) is a large, open space with a high volume of pedestrian traffic. It contains two major landmarks: a cafeteria that students, staff and visitors regularly use, and a large auditorium with several hundred seats. Information Wall was conveniently projected on a wall roughly between these two landmarks with a ceiling-mounted full HD projector.

The second deployment, *EnergyLand*, is a gesture-based system for initiating conversations on novel energy-efficient solutions for smart homes. Users interact with virtual objects in game-like tasks to understand ways to incorporate green energy practices inside their homes (Figure 1D). EnergyLand was a part of a Smart House installation at an annual Finnish housing fair in summer

Figure 1. Public displays and their locations. A) Panorama of the deployment location for Information Wall from the installation's perspective. Cafeteria on the far left, auditorium on the far right. B) Information Wall. C) Information Wall 2. D) EnergyLand. E) The deployment location for EnergyLand at a housing fair. The installation is at the far back of the room.

2012 (Figure 1E). EnergyLand was installed in a 7 by 5 meter rectangular area inside a large 400 square meter Smart House tent. EnergyLand consisted of three large interactive screens arranged along the length of a wall and three poles on the opposite wall held a projector each, one for each screen. Microsoft Kinect sensors were used for gesture recognition, and a computer/laptop was placed under each screen. Two sets of speaker arrays were placed above the left and right screens, and a 5.1 channel speaker system under the center screen and under the projectors.

From these two deployments, we gathered a total of 10 issues that were caused by external factors. We will describe these issues later in the paper, when we present the taxonomy of external factors.

Literature Review

For the literature review, we gathered a collection of papers using snowball sampling. First, we included papers that we were already familiar with due to our work on the field, such as those by Ojala et al. [54], Alt et al. [6], and Müller et al. [49,52]. Then, we reviewed the references in these papers, and added cited papers to the collection based on their title, or what was said about them in the citing paper. Following, we used Google Scholar to look for additional papers citing the papers already in the collection, and added papers based on the title and abstract. With this method, we gathered a collection of 132 papers. Of these, we further identified those that presented an in-the-wild public display deployment, or challenges or lessons learned regarding such, or discussed user behavior around public displays. This resulted in the final collection of 71 publications that were used in the literature review.

Two researchers carefully reviewed the 71 publications and noted down every issue, challenge, or observation that was attributable to an external factor. We found a total of 51 issues from 22 publications (several papers reported more than one issue). The 71 reviewed publications are listed in the references of this paper, with the 22 papers that were part of the taxonomy marked with an asterisk.

The fact that most reviewed papers did not report any issues suggests that a large segment of experienced issues within the HCI community may go unreported. Furthermore, as noted earlier, such issues are rarely the focus of research papers. Rather, we specifically had to look for them within the fine details of the papers – for example, issues were often only briefly mentioned to explain why certain study decisions were made.

Outcome

The final data set consisted of 61 issues, including 10 issues from our own deployments and 51 issues from the literature. Issues from both sources were treated equally during the process, i.e., no special emphasis was put on our own experiences. Therefore, our own experiences formed a ~16% contribution to RQ1 and RQ2.

To address RQ1, we utilized affinity diagramming to categorize the findings. The process was conducted by three researchers – the two researchers who conducted the literature review, and a third researcher for a fresh perspective. We considered that the process was complete when a *consensus* was reached, that is, when all three researchers agreed with every issue and the category they were assigned to. This resulted in six categories: *weather*, *events*, *surroundings*, *space*, *inhabitants*, and *vandalism*, which will be discussed in detail in the next chapter.

As a second layer of analysis (RQ2), we considered the strategies researchers used to deal with the observed issues, and identified four categories of responses. *Ignoring* means doing nothing to address or control the observed effect. *Adapting* means either changing some property of the deployment that introduces a trade-off (a suboptimal solution), or addressing the effect once while not ensuring that the same effect will not happen again. *Solving* refers to a permanent solution that either eliminates the original cause, or changes some property of the deployment to eliminate the effect, without introducing a trade-off. Finally, *embracing* means taking advantage of the observed effect, to turn the negative into positive.

EXTERNAL FACTORS IN PUBLIC DEPLOYMENTS

Our taxonomy of external factors consists of *weather*, *events*, *surroundings*, *space*, *inhabitants*, and *vandalism*. In this chapter, we present each category in detail, and discuss how to recognize and deal with these factors All sources that contributed towards this taxonomy are presented in Table 1.

The distribution of issues per cause and reaction are presented in Figure 2. Of the 61 issues, 39% were *ignored*, while 39% were *adapted* to. Only 13% of the issues were *solved*, and 8% were *embraced*. In this section, we will also provide examples of how researchers dealt with the experienced issues with respect to these reaction types.

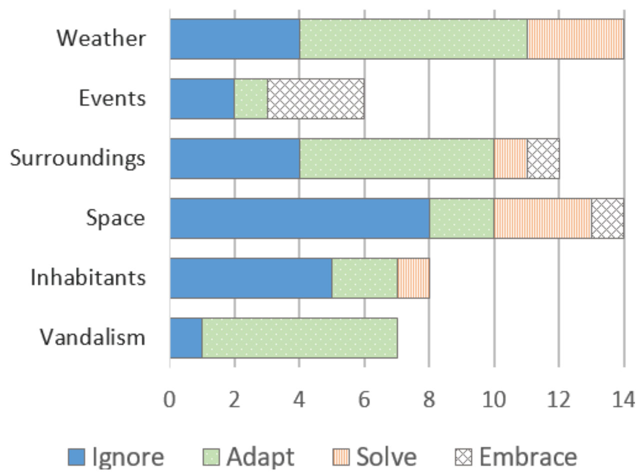


Figure 2. Distribution of issues per cause and reaction.

	Ignored	Adapted	Solved	Embraced
Weather	**EnergyLand** Ojala et al. [54] x 2 Valkanova et al. [68]	**EnergyLand** Ackad et al. [2] x 2 Ackad et al. [3] Harris et al. [31] x 2 Fischer & Hornecker [25]	**EnergyLand** Dalsgaard & Halskov [19] Harris et al. [31]	
Events	**Information Wall** O'Hara et al. [53]	O'Hara et al. [53]		Schroeter et al. [60] Akpan et al. [4] Finke et al. [23]
Surroundings	Ackad et al. [2], Storz et al. [62] Dalsgaard & Halskov [19] Fischer et al. [24]	Harris et al. [31] x 4 Storz et al. [62] Memarovic et al. [46]	Harris et al. [31]	Tomitsch et al. [67]
Space	Schroeter et al. [60] x 2 Fischer et al. [24] Churchill et al. [17], Kukka et al. [44] Peltonen et al. [57], Müller et al. [51] Fischer & Hornecker [25]	**EnergyLand** Fischer & Hornecker [25]	Müller et al. [52] Fischer et al. [24] Fischer & Hornecker [25]	Cao et al. [13]
Inhabitants	**Information Wall** x 2 Storz et al. [62] Dalsgaard & Halskov [19] x 2	**Information Wall** x 2	**Information Wall**	
Vandalism	Taylor et al. [65]	Dalsgaard & Halskov [19] x 4 Heikkinen et al. [33] Müller et al. [51]		

Table 1. Summary of reported challenges that contributed to the taxonomy, divided by cause and reaction.

Weather

Weather is a diverse, unstable and unpredictable phenomenon, and can affect both indoor and outdoor public deployments in various ways. We found that the primary causes for issues in the *weather* category are sunlight, rain, temperature, and humidity.

Reported Causes

Sunlight has caused issues especially with display visibility and sensor performance. In our Energyland installation, even though the installation was located inside a large, completely covered tent, the sunlight was bright enough to penetrate the tent's fabric and interfere with the Kinect sensors, which rendered the system unusable. We solved the problem by hanging thick, dark black curtains above our installation (Figure 3) to improve the gesture recognition.

Ackad et al. [2] also reported two issues with sunlight with their outdoor gesture-controlled display. Sunlight interfered with the sensors, as well as hindered the visibility of screen content. They adapted to the issues by only running the system from 6 pm until midnight, when sunlight no longer hit the deployment space. Similarly, Fischer and Hornecker [25] only ran their media façade system after dusk, since the projections were not visible enough in daylight. Harris et al. [31] reported varying visibility with their screen in an outdoor setting, and added hoods to the screens to minimize the effect from external lighting. Likewise, Ojala et al. [54] report low visibility with their UBI-hotspots in direct sunlight, however this issue was ignored.

Figure 3. Curtains hung above the EnergyLand installation to prevent issues with sunlight.

Rain resulted in serious issues with our EnergyLand installation near the end of the fair when a rain storm soaked parts of the floor and destroyed the computer running the system and one of the speakers. While the computer was quickly replaced with another unit, we were unable to replace the speakers within the available time frame. Hence, visitors had to experience the system without a soundscape for the remainder of the fair.

Harris et al. [31] had issues with rain and other harmful weather conditions with their deployment that was setup in the woods. They needed to keep an eye on the weather forecast at all times. To offer some protection for the equipment, polythene bags were used. A back-up procedure was to simply not run the setup on very rainy days. Dalsgaard and Halskov [19] reported an interesting effect from rain on their motion-tracking media façade installation. Reflections from puddles after rainfall were incorrectly identified as people by the sensor. This was eventually solved when colorful carpets were installed to more clearly mark the interaction areas.

More general-purpose weather effects were reported in several papers. A pilot study of a public voting installation by Valkanova et al. [68] suffered from unpleasant weather conditions (rain, wind, cold) that resulted in fewer users (and interviewees), as most passersby were not motivated to stay outdoors. Ackad et al. [3] studied user behavior on an outdoor installation, and had to schedule their studies based on weather conditions. Ojala et al. [54] noticed that winter months affected the UBI-hotspot usage as their capacitive touch screen foils did not respond to gloves. During the deployment of their setup in the woods, Harris et al. [25] noticed that high levels of humidity will attenuate radio waves, which they needed to tune on the spot.

Recommendations

The effects of *weather* can range from screen visibility issues to destruction of equipment. Since weather cannot be controlled, researchers must be aware of its potential effects and be ready to adapt. Representative examples of such adaptations are those by Ackad et al. [2] and Fischer and Hornecker [25], wherein the deployment was run only during evenings to avoid sunlight issues. However, while they could work around the issue by decreasing the deployment time, this workaround is far from optimal.

Weather can also have positive effects. Ojala et al. [55] found that sunny and warm days result in more users as well as increased interaction times, and interestingly, this phenomenon was also observed in indoor locations. This is likely because even for indoor activities, people still need to go out and move to the location – and they may be encouraged to do so by the nice weather.

We present the following considerations:

- Consider the effects of sunlight on the deployment: sunlight affects screen visibility as well as the performance of many motion tracking sensors.
- Cover sensitive hardware. Rain and humidity can cause surprising issues even in seemingly well covered environments.
- Consider how weather conditions affect people. Warm and sunny weather can result in more interactions and increased interaction times, even for indoor locations [55]. Inclement weather, such as cold and wind, decreases the number of people outside and their willingness to interact [68].

Events

As one category, we consider *events*, such as festivals, that temporarily (but significantly) alter the characteristics of the deployment space or its surroundings, hence affecting the installation.

Reported Causes

With our Information Wall installation, we observed issues caused by an event for the elderly that was held in the nearby auditorium and the adjacent lecture rooms. Despite the large, open deployment space, it became unusually full when hundreds of people poured out of the rooms. Some noticed the display and came to investigate, and some tried interacting with it. However, other people constantly disrupted the interaction (and observation) by moving past the display between the users and the sensor, effectively blocking both the sensor and the screen. We also observed some discomfort from the users, as they ultimately felt they were in the way of others when interacting. These observations did not lead to any action on our part, as there were no easy solutions available.

O'Hara et al. [53] reported that a temporary outdoor urban beach was installed in front of the large BBC screen in Birmingham. Because the BBC screen required users to execute large physical movements, and because the beach installation expected crowds, this was a safety hazard. To overcome this, barriers were erected around the active gameplay area to keep the crowds safely separated, and a professional speaker was also introduced to help manage the crowd and gameplay, which increased financial costs.

Some positive effects due to events have also been reported. A deployment by Finke et al. [23] was surrounded by tables, chairs and a coffee bar. At the end of the student semester, more chairs and tables were brought into the deployment space to offer students sufficient room to prepare for their final exams. This resulted in a larger audience and more interactions with the system. In addition, Akpan et al. [4] reported that a party held at the premises in which their interactive Shadow Wall was installed resulted in a much higher amount of interactions as well as longer durations of interactions, as the party transformed the configuration of the space to a more relaxed one. O'Hara et al. [53] also mention that the above-mentioned temporary beach somewhat changed the nature of the space, as people who were interviewed mentioned that they came to the place to rest, take a break, wait, sit or socialize.

A general observation of the effects of events was made by Schroeter et al. [60] who noted that events change the demographic in the area. For instance, they state that on normal days, people around public displays can be "broad and diverse in terms of their interests". However, during events, usually a more specific crowd is present. Thus, the effect of events can be embraced, for example by focusing the content of the display towards this specific group.

Recommendations

Events typically have two effects. First, they alter the social setting of the space. For instance, O'Hara et al. [53] and Akpan et al. [4] experienced a positive effect, as an event resulted in a more favorable, relaxed setting. Second, events affect crowds and people flow. Events often equal to higher amounts of people, which can lead to security risks and interference with interaction, but can also lead to positive effects, such as an increased number of users. We derive three practical considerations from the reported issues:

- Stay aware of the magnitude and nature of upcoming events in the area. Events affect the social setting, demographics, and the amount of people.
- The deployment should be robust enough to handle interference caused by crowds, even if such crowds are not normally present. For instance, people in the background should not prevent others from interacting.
- It is possible to turn events to one's advantage, by tailoring the content of the deployment to serve the participants of the event. Events tend to bring in more users and foster interaction in general.

Surroundings

As the third category, we consider *surroundings*, which refers to areas and their persistent characteristics surrounding the deployment space. This includes nearby buildings, their functions, and traffic. Research on public deployments has focused extensively on the effects of immediate surroundings near a display, such as people flow and its subsequent interaction effects. However, several issues reported in literature point towards the effects of buildings and artifacts surrounding the area of deployment, from which the display or installation is likely not visible.

Reported Causes

Tomitsch et al. [67] report a higher flow of people before and after a show in the theater, adjacent to their installation, and how the nature of each show would affect the demographic of people in the area. For Fischer et al. [24], the day-time pedestrian traffic near their installations was dominated by students from a nearby university and at night by visitors to a nearby theater. Ackad et al. [2] further note that due to an adjacent theater near their installation, passersby often had limited time to interact, as they were on their way to see a show. Yet contrary to *events*, the functions of the surroundings are perennial. Therefore, public displays should be designed to account for the effects of surroundings, for example by designing content that serves the theater visitors [1,2].

Surroundings can also impact the deployment's hardware. Storz et al. [62] had an installation that was in a tunnel under a heavily trafficked street. Diesel exhaust continuously interfered with their equipment, for example by clogging the projector filters used in the installation, despite custom-built cases, housings, and other prior preparations. Furthermore, maintaining the hardware required closing the road, which was especially difficult during high traffic periods such as university semesters. For Harris et al. [31], their setup in the woods needed to be disassembled on a daily basis to prevent moisture deposits from accumulating on electrical boards. Dalsgaard and Halskov [19] reported windows across the street from their installation causing reflections, which were incorrectly interpreted as people by their motion-controlled system.

Moreover, surroundings can create adverse deployment conditions, including lack of electricity or unreliable network connectivity. Harris et al. [31] had no access to electricity for their deployment that was set up in the woods. Generators were not an option due to noise and running cables to each device would have been impractical. They adapted to the issue by using batteries and low-power devices where possible. Further, trees blocked GPS signals and radio frequencies. This was solved by adding three booster antennas for their own network. Due to the dynamic and unstable nature of woods and similar areas, the ground can change from hard to soft and shift unexpectedly. Solutions included placing network access points to trees and adding base plates to the ground where required. Issues with network connectivity within the deployment location were solved by Memarovic et al. [46] by connecting wireless receivers from the rooftops of local houses to the installation. It was also important to make at least a part of the content available offline in case of network issues.

Recommendations

Understanding the functions of surroundings, for example nearby buildings, traffic conditions and vegetation, provides a better understanding of the potential user demographics and technical limitations. We present three considerations:

- Identify key locations in the surrounding areas, such as theaters and malls, as they provide further insight for times of day with large crowds and specific demographics.
- Consider the possibility of designing for these locations. Ackad et al. [1,2] tailored the content of their public display to serve the visitors to a nearby theater. Ideally one should take this into account a priori in the fundamental design decisions of the deployment.
- Technical limitations need to be studied before the deployment to accommodate network failures, power issues, and other potential factors that may interfere with the equipment.

Factors in *surroundings* may have similar effects as those in the *events* category, such as increased people flow; however, the differentiation is based on the prevalence of cause. Properties of the *surroundings* are permanent, and cannot be made to go away without architectural interventions.

Space

We consider *space* as the physical attributes of the deployment area, whereas with the previously discussed *surroundings* category we refer to areas outside of the deployment area. With *space*, we follow the definition of Harrison and Dourish [32]: space is the actual physical structure of a "three-dimensional environment in which objects and events occur". A space is made a place by its inhabitants [32]; therefore, place is the social and cultural construct within a physical space. Issues arising due to inhabitants of a space are discussed later in the *inhabitants* category. Here, we report issues that were caused by properties of the deployment space.

Reported Causes

In EnergyLand, we had access to the floorplan of the tent and used it to define the installation layout. However, a few days before the deployment, the location managers informed us that details provided of the space were incorrect – there was a double wall structure inside the tent and the thickness of the wall was not provided. We had to adapt by setting the first and third display at an angle (Figure 3), when originally all displays were supposed to be parallel to the wall. Although such a small change may seem trivial, studies have shown that display orientation can have a significant effect on the dynamics of the installation [41]. Moreover, the angles made the interaction spaces for each display somewhat overlap. Fischer and Hornecker [25] also report discrepancies between their initial plan for a media façade installation and the actual physical properties of the façade, and they had to adjust accordingly.

Schroeter et al. [60] note that due to their installation's size and positioning amongst other booths at their deployment location, the visibility of their display was limited. For Fischer et al. [24], a railway track ran across their installation space, with the media façade being on one side of the tracks and the interactive controller device on the other. This prevented them from connecting their system with a network cable, and regular Wi-Fi was not an option due to electrical interference from the trams. A solution was found in the form of a GSM USB adapter. In another case, Fischer and Hornecker [25] had to cover the illumination of an entrance to an underground parking space, as it interfered with their media façade. Müller et al. [52] reported that after moving the installation to a different space, patches on the floor would be recognized by system as users while actual far away users were not. Both issues were solved by moving the camera to floor level.

For Müller et al. [51], nearby elevators influenced the use of their public display installation as people would often interact with the system while they waited but abruptly stop when the elevator arrived. In CityWall [57], people waited out the rain under the sun shade on the installation window without even realizing that it is interactive. Furthermore, Fischer et al. [24] state that their installation was jammed when around 70 people suddenly came out of a bus from a stop nearby. Churchill et al. [17] noted that by installing a display in a foyer with a T-junction, people moving in the space were colliding with those interacting with the display. Kukka et al. [44] noted that one of their displays was sandwiched between two bulletin boards, located in a busy walkway opening to a cafeteria. Instead of making the display more noticeable, this instigated display avoidance.

However, not all space dynamics are perceived as being negative by researchers. An installation by Cao et al. [13] was located at a public university atrium, which had open seating, a café nearby, and access to classrooms. This resulted in increased traffic, especially when classes were ending or starting. This was desirable for the researchers as it promoted the usage of their installation. Moreover, Fischer and Hornecker [25] found that the built environment, such as pillars and trees, can create "comfort spaces" wherein people can comfortably observe the installation. While Fischer and Hornecker [25] also note that these comfort spaces might "interfere with designers' intentions of achieving a certain situation in a given setting", we note that they also encourage people to engage with the installation, as people can first observe how it works without being "out in the open".

Recommendations

Researchers must study and understand the functions of the space, comparable to the previously presented *surroundings* category. A somewhat similar approach was utilized by Müller et al. [50] when designing interactive displays for shop windows, as they observed the deployment space with regards to people flow and passersby attention, to identify the most suitable windows for the displays. However, we emphasize the focus on identifying *external factors*. We present the following considerations:

- Observe the physical characteristics of the space, including the layout, placement of doorways, pillars, elevators, machinery, and other displays and installations. Further, identify how said characteristics affect people flow and the use of the space.
- Identify permanent structures, such as pillars and other objects, within the layout that can potentially create "comfort spaces" for the users. Fischer and Hornecker [25] found that these spaces allow people to comfortably observe the installation.

By knowing the physical characteristics of the space, one can predict some of the effects on the deployment. Once there is a clear understanding of the space, researchers can begin to map out the use of the space by its *inhabitants*.

Inhabitants

Deployment spaces, as well as the surrounding areas, are usually filled with a multitude of people, each with varying roles. While current pervasive display research largely investigates people as passersby, i.e., potential users, with the term *inhabitants* we consider their usual, day-to-day roles within the space, such as students, janitorial staff, security guards, and store owners.

For clarification, stakeholders are not part of the *inhabitants* category, as stakeholders are an internal part of the deployment project, and are therefore not an external factor.

It is notable, however, that the same person can be both - the key is the *role* the person is acting on. For instance, location managers can be stakeholders, and the actions they undertake *as stakeholders* would not be considered an external factor, however they would still be inhabitants in their usual roles as location managers. For further reading on stakeholders, we refer to Elhart et al. [22] and Hosio et al. [34].

Reported Causes

With our Information Wall installation at a university campus, security guards kept turning off the projector at night. We were unaware of this for a long time, but eventually found out when we asked the local managers to investigate if the projector was faulty. One of the managers contacted the security guards to ask if they, by any chance, are turning the screen off during their nightly rounds. Indeed, the guards were turning off the projector because they assumed that it should not be on during the night. The real cause was surprising, as we were at no point in contact with the guards. In fact, we never even *saw* the guards.

We also experienced various other surprises regarding local inhabitants. The large, open deployment space is occasionally used to set up temporary stands. For instance, members of student organizations occasionally set up stands to sell tickets to parties, and members of various clubs set up stands to promote their club and recruit new members. We experienced several occurrences wherein the stands had been set up so that they obstructed the use of our installation, or at the very least made interacting uncomfortable, even though there was plenty of room in the space for the stands to be positioned elsewhere.

Sometimes we found furniture or other equipment temporarily positioned so that they obstructed or hindered the use of Information Wall. Some of these were presumably left behind from the above-mentioned stands when they were taken down. Each time, we simply moved the furniture slightly to ease the situation, but left the furniture to be cleaned up by the people responsible for them. Additionally, we observed two occasions where a janitor’s cart was positioned partly in front of the installation – but the janitor was nowhere in sight. We took no action to address the issue, however we would have sought out the janitors if the issue had been more frequent.

Another interesting occurrence happened with local students. As part of the Information Wall, a cupboard (which contained the PC running the system) was placed underneath the screen, and the Kinect sensor was placed on top of the cupboard. Once, we found that the nearby coat hanger was (almost) full, so a group of people had left their jackets on top of the cupboard, completely covering the Kinect, which prevented the system from being used.

Dalsgaard and Halskov [19] installed a webcam-enabled public display behind the windows in the grocery section of a local department store. A local manager lowered sun-blinds in some of the windows, preventing the webcam from seeing passersby. Storz et al. [62] discuss that in one of their deployment locations, local managers tested the fire alarms weekly, during which time the power was cut off from the deployment space.

Recommendations

An important insight arises from the *inhabitants* category. Inhabitants are unrelated to the deployment - none of the reported issues involved intentional interference with the deployment (perhaps apart from the guards, however it was not their *intention* to disrupt the deployment). Rather, issues arising from this category are primarily due to inhabitants conducting their usual, everyday routines, and the deployment is simply in the way. From the inhabitants’ perspective, it could be argued that it is the deployment that is invading *their* space.

To further enable communication with the inhabitants, a practical suggestion is to add background and contact information near the installation in case the inhabitants have issues or questions. With our Information Wall installation, we received a request by staff from another school to add temporary content about an upcoming event to our installation. In the absence of contact information near the display, these people had to go through multiple channels to finally ascertain who to contact about the display.

For *inhabitants*, we present the following considerations:

- Observe and identify different groups of inhabitants, and further aim to identify their routines and tasks in the space. Then, analyze how these routines and tasks could interfere with the deployment, and adjust the design accordingly.
- Consider engaging in a discussion with the inhabitants, and informing them of the deployment and its purpose, particularly if the deployment is likely to interfere with their normal routines.
- Consider adding contact details near the deployment.

Vandalism

With *vandalism*, we primarily refer to intentional interference with the deployment as well as the destruction of equipment. Public deployments are sometimes at the mercy of people who may not be regular inhabitants. Most notably, it is not realistic to consider vandalism to be a person’s primary activity as an inhabitant, which is why vandalism should be considered a separate category.

Reported Causes

Heikkinen et al. [33] (also mentioned in Ojala et al., [54]) report that the safety glass of one of their outdoor installations was purposefully broken – the system was down for two weeks, and the expenses for repairs were significant. Similarly, Dalsgaard and Halskov [19] report that their information stand displaying videos on a bus shelter was destroyed. Dalsgaard and Halskov [19] further report that in one of their outdoor media façade installations, someone tried to remove a colorful carpet from the ground, possibly with the assumption that the carpet was enabling the interactions. In reality, the carpet was there to merely mark the interaction area more clearly.

In addition to physical harm, vandalism can stem from rude or unacceptable behavior. Dalsgaard and Halskov [19] reported a few instances wherein users would upload erotic content to one of their public systems which allowed uploading low-resolution animations. They also reported a user recording obscene gestures in an installation that allowed users to record short clips to be shown in displays around the city. The recorded clips had to be checked and edited before showing them to the public.

We also found an interesting link between misuse of an installation and playful behavior. Taylor et al. [65] installed a display aimed at gathering opinions on issues relevant to the local community. Users could give their opinion on polls by pressing corresponding buttons below the display. However, it was discovered that some people were interacting merely to play around with the system (to experience it) and pressing the buttons uncontrollably. Therefore, the results of the polls, which otherwise could have been useful to the community, were skewed.

Müller et al. [51] observed users engaging in (unintended) harmful activities through playful behavior, when studying transitioning between mid-air gestures and touch in their MirrorTouch installation. They reported two occasions where users first interacted with increasingly expressive mid-air gestures, and when transitioning to touch, they interacted with the screen so roughly that researchers needed to intervene to prevent damage to the installation.

Many more authors have observed playful behavior where they have not expected it, even though no actual issues have been encountered due to it. Examples of such are presented in [19,52,57,67].

Recommendations

The effects of *vandalism* range from destruction of equipment to unintentional interference with the deployment. There may also be a link between playful behavior and misuse. While playful behavior has many benefits, such as enticing interaction [67,70], researchers should be aware of the possible negative consequences as well. We found a case where playful behavior turned too extreme and was about to lead to damaged hardware [51], and a case where playful interaction interfered with the serious purpose of an installation [65].

It is difficult to predict vandalism or misuse; however, to address some of them, we offer the following considerations:

- Go into the "mindset" of a vandal. What are the easy targets for vandalism? Consider attaching all parts of the deployment to something sturdy, and hiding components, such as wires, from plain sight.
- Consider moderating user-generated content. Dalsgaard and Halskov [19] experienced several occurrences wherein improper content was uploaded. Fortunately, such instances seem to be rare. For instance, Elhart et al. [22] had to go to great lengths to prepare for improper content due to requirements from stakeholders; however, they experienced no issues.

DISCUSSION AND CONCLUSION

In this work, we investigated *external factors* that cause many surprising and unwanted effects on public display deployments. We conducted a literature review and, using affinity diagramming, identified a taxonomy of external factors: *weather*, *events*, *surroundings*, *space*, *inhabitants*, and *vandalism*. We found that our experience about the primarily negative effects of these factors hold true within the larger HCI community as well.

Our work differentiates itself from past research in two ways. First, we draw upon the *collective experiences* within the HCI community. Some past work has discussed challenges with public deployments, and while they offer valuable insights, the challenges have only been individual research groups' experiences [19,31,62]. However, many of the issues we collected have come from publications in which these issues were not the focus. This, in part, underlines the need to bring these experiences together. Second, we are the first to approach these challenges from the perspective of *cause*. This is important in offering a practical approach towards identifying and adapting to potential issues. From our viewpoint, researchers should be encouraged to look beyond the observed challenges, as identifying the cause may help in addressing the issue.

In addition to the taxonomy of external factors, we investigated how researchers deal with issues arising from these factors, and identified four types of reactions: *ignoring*, *adapting*, *solving*, and *embracing*. Ignoring and adapting to issues were much more common responses, which suggests that the issues are often surprising, and are difficult to solve completely.

We note that in this work, we quantified how researchers *eventually* reacted to the issues. Based on our experiences and some examples found in the literature, for many issues that were ignored or adapted to, researchers initially reviewed their options and, in the absence of good and easy solutions, had to resort to sub-optimal solutions (adapting), or had to ignore the issue. In deciding the action to take, researchers must evaluate how severe the issue is, and consider available options and how much effort they require as opposed to the resources available.

When evaluating the severity of an issue, the purpose of the deployment is worth considering, as many public deployments are conducted to study a specific phenomenon. For instance, for a public deployment investigating display blindness in some form, an interaction-related issue might not be as severe as it would in many other scenarios. In

Weather	• Consider the effects of sunlight on the deployment as it affects screen visibility and the performance of sensors. • Cover sensitive hardware, as rain and humidity can cause surprising issues. • Consider how weather affects people. Good weather may increase the number of interactions, and bad weather may have the opposite effect – even for indoor deployments.
Events	• Stay aware of the magnitude and nature of upcoming events in the area - events may alter the social setting, demographics, and the amount of people. • Design the deployment to handle large crowds, even if such crowds are not normally present. • Consider turning events to the deployment's advantage, for instance, by tailoring the content to serve the participants of the event. Events tend to bring in more users and foster interaction.
Surroundings	• Identify key locations in the surrounding areas, such as theaters and malls, and consider how they affect the demographics in the area at different times of day. • Consider designing content for the deployment that is specific to these locations. • Identify technical limitations to accommodate network failures, power issues, and other potential factors that may interfere with the equipment.
Space	• Observe the physical characteristics of the space, e.g., the layout, doorways, pillars, elevators, and machinery. Identify how these characteristics affect people flow and the use of the space. • Identify pillars and other objects within the layout that can potentially create "comfort spaces" for the users, which may encourage people to observe and subsequently interact.
Inhabitants	• Observe and identify different groups of inhabitants, and their routines and tasks in the space. Analyze how their routines and tasks could interfere with the deployment. • Consider informing and engaging the inhabitants, particularly if the deployment is likely to interfere with inhabitants' normal routines. • Add contact details near the deployment.
Vandalism	• Go into the "mindset" of a vandal, and identify easy and obvious targets for vandalism. Attach all parts of the deployment to something sturdy, and hide easy targets, such as wires. • Consider moderating user-generated content, as people may upload improper material.

Table 2. Considerations for external factors in public deployments.

addition, many research-oriented deployments are short-term in nature, due to which some issues might receive less attention than they would in long-term deployments.

In defining the available options in each situation, an important factor is the level of control one has over the deployment and the surrounding area. Researchers rarely own the deployment space, which significantly limits the amount and extent of available options. This brings us back to the importance of *adapting* to issues. Solving issues completely is difficult and, as we observed in this work, not always possible.

In conclusion, with this work, we hope to bring attention to external factors and help researchers reflect on and report issues experienced in public deployments. The prevailing approach in reporting deployments is to focus on *successes*; experienced issues are often considered threats to the integrity of the studies. In fact, experienced challenges can often be more informative than the successes, and can aid researchers and practitioners in carrying out more successful deployments. Moreover, there are now good venues to publish such experiences (e.g., CHI Stories).

We provide a go-to summary of the taxonomy of external factors with practical considerations in Table 2. The main contribution of our work is to help researchers and practitioners design *for* external factors, and be more prepared to address the effects. This is achieved in three ways. First, by using the taxonomy of the external factors, one can more easily identify such factors within the deployment. Second, we provide practical insights for each category, and report several examples on how past work has addressed the encountered issues. Third, we show that one must be prepared to adapt – many observed negative effects usually require changes in the deployment. Overall, we wish to provide a useful framework for understanding external factors and hope that it is built upon by the community, by sharing how the issues were reacted to in future deployments.

REFERENCES

1. Christopher Ackad, Rainer Wasinger, Richard Gluga, Judy Kay, and Martin Tomitsch. 2013. Measuring interactivity at an interactive public information display. In *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration* (OzCHI '13), Haifeng Shen, Ross Smith, Jeni Paay, Paul Calder, and Theodor Wyeld (Eds.). ACM, New York, NY, USA, 329-332. DOI: http://dx.doi.org/10.1145/2541016.2541091

2. Christopher Ackad, Andrew Clayphan, Martin Tomitsch, and Judy Kay. 2015. An in-the-wild study of learning mid-air gestures to browse hierarchical information at a large interactive public display. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp '15). ACM, New York, NY, USA, 1227-1238. DOI: http://dx.doi.org/10.1145/2750858.2807532 *

3. Christopher Ackad, Martin Tomitsch, and Judy Kay. 2016. Skeletons and Silhouettes: Comparing User Representations at a Gesture-based Large Display. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). ACM, New York, NY, USA, 2343-2347. DOI: http://dx.doi.org/10.1145/2858036.2858427 *

4. Imeh Akpan, Paul Marshall, Jon Bird, and Daniel Harrison. 2013. Exploring the effects of space and place on engagement with an interactive installation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13). ACM, New York, NY, USA, 2213-2222. DOI: http://dx.doi.org/10.1145/2470654.2481306 *

5. Florian Alt, Albrecht Schmidt, and Jörg Müller . 2012. Advertising on Public Display Networks. *IEEE Computer* 45, 5 (May 2012), 50-56. DOI= https://doi.org/10.1109/MC.2012.150

6. Florian Alt, Stefan Schneegaß, Albrecht Schmidt, Jörg Müller, and Nemanja Memarovic. 2012. How to evaluate public displays. In *Proceedings of the 2012 International Symposium on Pervasive Displays* (PerDis '12). ACM, New York, NY, USA, , Article 17 , 6 pages. DOI=http://dx.doi.org/10.1145/2307798.2307815

7. Moritz Behrens, Ava Fatah gen. Schieck, Efsathia Kostopoulou, Steve North, Wallis Motta, Lei Ye, and Holger Schnadelbach. 2013. Exploring the effect of spatial layout on mediated urban interactions. In *Proceedings of the 2nd ACM International Symposium on Pervasive Displays* (PerDis '13). ACM, New York, NY, USA, 79-84. DOI=http://dx.doi.org/10.1145/2491568.2491586

8. Gisele Bennett, Gitte Lindgaard, Bruce Tsuji, Kay H. Connelly, and Katie A. Siek. 2006. Reality testing: HCI challenges in non-traditional environments. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '06). ACM, New York, NY, USA, 1679-1682. DOI=http://dx.doi.org/10.1145/1125451.1125761

9. Gilbert Beyer, Florian Alt, Jörg Müller, Albrecht Schmidt, Karsten Isakovic, Stefan Klose, Manuel Schiewe, and Ivo Haulsen. 2011. Audience behavior around large interactive cylindrical screens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11). ACM, New York, NY, USA, 1021-1030. DOI=http://dx.doi.org/10.1145/1978942.1979095

10. Harry Brignull and Yvonne Rogers. 2003. Enticing people to interact with large public displays in public spaces. In *Proceedings of INTERACT*, 3, 17-24.

11. Barry Brown, Stuart Reeves, and Scott Sherwood. 2011. Into the wild: challenges and opportunities for field trial methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11). ACM, New York, NY, USA, 1657-1666. DOI=http://dx.doi.org/10.1145/1978942.1979185

12. Susanne Bødker. 2006. When second wave HCI meets third wave challenges. In *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles* (NordiCHI '06), Anders Mørch, Konrad Morgan, Tone Bratteteig, Gautam Ghosh, and Dag Svanaes (Eds.). ACM, New York, NY, USA, 1-8. DOI=http://dx.doi.org/10.1145/1182475.1182476

13. Xiang Cao, Michael Massimi, and Ravin Balakrishnan. 2008. Flashlight jigsaw: an exploratory study of an ad-hoc multi-player game on public displays. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work* (CSCW '08). ACM, New York, NY, USA, 77-86. DOI=http://dx.doi.org/10.1145/1460563.1460577 *

14. John M. Carroll. 2003. The Blacksburg Electronic Village: A Study in Community Computing. *Lecture Notes in Computer Science* 3081 (September 2003), 43-65. DOI= http://dx.doi.org/10.1007/11407546_3

15. Scott Carter and Jennifer Mankoff. 2005. Prototypes in the Wild: Lessons from Three Ubicomp Systems. *IEEE Pervasive Computing* 4, 4 (October 2005), 51-57. DOI=http://dx.doi.org/10.1109/MPRV.2005.84

16. Chien-Hsu Chen, Hsiao-Mei Hung, I-Jui Lee, Yu-Wen Chen, and Fong-Gong Wu. 2011. Observe the user interactive behavior with a large multi-touch display in public space. In *Proceedings of the 6th international conference on Universal access in human-computer interaction: context diversity - Volume Part III* (UAHCI'11), Constantine Stephanidis (Ed.), Vol. Part III. Springer-Verlag, Berlin, Heidelberg, 141-144.

17. Elizabeth F. Churchill, Les Nelson, Laurent Denoue, Jonathan Helfman, and Paul Murphy. 2004. Sharing multimedia content with interactive public displays: a case study. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques* (DIS '04). ACM, New York, NY, USA, 7-16. DOI=http://dx.doi.org/10.1145/1013115.1013119 *

18. Céline Coutrix, Kai Kuikkaniemi, Esko Kurvinen, Giulio Jacucci, Ivan Avdouevski, and Riikka Mäkelä. 2011. FizzyVis: designing for playful information browsing on a multitouch public display. In *Proceedings of the 2011 Conference on Designing Pleasurable Products and Interfaces* (DPPI '11). ACM, New York, NY, USA, , Article 27 , 8 pages. DOI=http://dx.doi.org/10.1145/2347504.2347534

19. Peter Dalsgaard and Kim Halskov. 2010. Designing urban media façades: cases and challenges. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10). ACM, New York, NY, USA, 2277-2286. DOI=http://dx.doi.org/10.1145/1753326.1753670 *

20. Peter Dalsgaard, Christian Dindler, and Kim Halskov. 2011. Understanding the dynamics of engaging interaction in public spaces. In *Proceedings of the 13th IFIP TC 13 international conference on Human-computer interaction - Volume Part II* (INTERACT'11), Pedro Campos, Nuno Nunes, Nicholas Graham, Joaquim Jorge, and Philippe Palanque (Eds.), Vol. Part II. Springer-Verlag, Berlin, Heidelberg, 212-229.

21. Paul Dourish. 2006. Re-space-ing place: "place" and "space" ten years on. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work* (CSCW '06). ACM, New York, NY, USA, 299-308. DOI=http://dx.doi.org/10.1145/1180875.1180921

22. Ivan Elhart, Marc Langheinrich, Nemanja Memarovic, and Elisa Rubegni. 2016. A good balance of costs and benefits: convincing a university administration to support the installation of an interactive multi-application display system on campus. In *Proceedings of the 5th ACM International Symposium on Pervasive Displays* (PerDis '16). ACM, New York, NY, USA, 197-203. DOI: http://dx.doi.org/10.1145/2914920.2915029

23. Matthias Finke, Anthony Tang, Rock Leung, and Michael Blackstock. 2008. Lessons learned: game design for large public displays. In *Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts* (DIMEA '08). ACM, New York, NY, USA, 26-33. DOI=http://dx.doi.org/10.1145/1413634.1413644 *

24. Patrick Tobias Fischer, Christian Zöllner, and Eva Hornecker. 2010. VR/Urban: Spread.gun - design process and challenges in developing a shared encounter for media façades. In *Proceedings of the 24th BCS Interaction Specialist Group Conference* (BCS '10). British Computer Society, Swinton, UK, UK, 289-298. *

25. Patrick Tobias Fischer and Eva Hornecker. 2012. Urban HCI: spatial aspects in the design of shared encounters for media facades. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12). ACM, New York, NY, USA, 307-316. DOI=http://dx.doi.org/10.1145/2207676.2207719 *

26. Adrian Friday, Nigel Davies, and Christos Efstratiou. 2012. Reflections on Long-Term Experiments with Public Displays. IEEE Computer 45, 5 (May 2012), 42-49. DOI= https://doi.org/10.1109/MC.2012.155

27. Kazjon Grace, Rainer Wasinger, Christopher Ackad, Anthony Collins, Oliver Dawson, Richard Gluga, Judy Kay, and Martin Tomitsch. 2013. Conveying interactivity at an interactive public information display. In *Proceedings of the 2nd ACM International Symposium on Pervasive Displays* (PerDis '13). ACM, New York, NY, USA, 19-24. DOI=http://dx.doi.org/10.1145/2491568.2491573

28. Sukeshini A. Grandhi, Gina Joue, and Irene Mittelberg. 2011. Understanding naturalness and intuitiveness in gesture production: insights for touchless gestural interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11). ACM, New York, NY, USA, 821-824. DOI=http://dx.doi.org/10.1145/1978942.1979061

29. Saul Greenberg and Michael Rounding. 2001. The notification collage: posting information to public and personal displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '01). ACM, New York, NY, USA, 514-521. DOI=http://dx.doi.org/10.1145/365024.365339

30. Jonna Häkkilä, Olli Koskenranta, Maaret Posti, Leena Ventä-Olkkonen, and Ashley Colley. 2013. Clearing the virtual window: connecting two locations with interactive public displays. In *Proceedings of the 2nd ACM International Symposium on Pervasive Displays* (PerDis '13). ACM, New York, NY, USA, 85-90. DOI=http://dx.doi.org/10.1145/2491568.2491587

31. Eric Harris, Geraldine Fitzpatrick, Yvonne Rogers, Sara Price, Ted Phelps, and Cliff Randell. 2004. From Snark to Park: Lessons learnt moving pervasive experiences from indoors to outdoors. In *Proceedings of the fifth conference on Australasian user interface* (AUIC ‘04), Australian Computer Society, Inc., 39-48. *

32. Steve Harrison and Paul Dourish. 1996. Re-place-ing space: the roles of place and space in collaborative systems. In *Proceedings of the 1996 ACM conference on Computer supported cooperative work* (CSCW '96), Mark S. Ackerman (Ed.). ACM, New York, NY, USA, 67-76. DOI=http://dx.doi.org/10.1145/240080.240193

33. Tommi Heikkinen, Tomas Lindén, Timo Ojala, Hannu Kukka, Marko Jurmu, and Simo Hosio. 2010. Lessons learned from the deployment and maintenance of ubi-hotspots. In *Proceedings of the Fourth International Conference on Multimedia and Ubiquitous Engineering* (MUE '10), 1-6. http://dx.doi.org/10.1109/MUE.2010.5575054 *

34. Simo Hosio, Jorge Goncalves, Hannu Kukka, Alan Chamberlain, and Alessio Malizia. 2014. What's in it for me: Exploring the Real-World Value Proposition of Pervasive Displays. In *Proceedings of The International Symposium on Pervasive Displays* (PerDis '14), Sven Gehring (Ed.). ACM, New York, NY, USA, , Pages 174 , 6 pages. DOI=http://dx.doi.org/10.1145/2611009.2611012

35. Simo Hosio, Hannu Kukka, Jorge Goncalves, Vassilis Kostakos, and Timo Ojala. 2016. Toward Meaningful Engagement with Pervasive Displays. *IEEE Pervasive Computing* 15, 3 (July 2016), 24-31. DOI= https://doi.org/10.1109/MPRV.2016.58

36. Elaine M. Huang and Elizabeth D. Mynatt. 2003. Semi-public displays for small, co-located groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '03). ACM, New York, NY, USA, 49-56. DOI=http://dx.doi.org/10.1145/642611.642622

37. Elaine M. Huang, Anna Koster, and Jan Borchers. 2009. Overcoming Assumptions and Uncovering Practices: When Does the Public Really Look at Public Displays?. In *Proceedings of the 6th International Conference on Pervasive Computing* (Pervasive '08), Jadwiga Indulska, Donald J. Patterson, Tom Rodden, and Max Ott (Eds.). Springer-Verlag, Berlin, Heidelberg, 228-243. DOI=10.1007/978-3-540-79576-6_14 http://dx.doi.org/10.1007/978-3-540-79576-6_14

38. Giulio Jacucci, Ann Morrison, Gabriela T. Richard, Jari Kleimola, Peter Peltonen, Lorenza Parisi, and Toni Laitinen. 2010. Worlds of information: designing for engagement at a public multi-touch display. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10). ACM, New York, NY, USA, 2267-2276. DOI=http://dx.doi.org/10.1145/1753326.1753669

39. Wendy Ju and David Sirkin. 2010. Animate objects: how physical motion encourages public interaction. In *Proceedings of the 5th international conference on Persuasive Technology* (PERSUASIVE'10), Thomas Ploug, Per Hasle, and Harri Oinas-Kukkonen (Eds.). Springer-Verlag, Berlin, Heidelberg, 40-51. DOI=http://dx.doi.org/10.1007/978-3-642-13226-1_6

40. Marko Jurmu, Masaki Ogawa, Sebastian Boring, Jukka Riekki, and Hideyuki Tokuda. 2013. Waving to a touch interface: descriptive field study of a multipurpose multimodal public display. In *Proceedings of the 2nd ACM International Symposium on Pervasive Displays* (PerDis '13). ACM, New York, NY, USA, 7-12. DOI=http://dx.doi.org/10.1145/2491568.2491571

41. Maurice Ten Koppel, Gilles Bailly, Jörg Müller, and Robert Walter. 2012. Chained displays: configurations of public displays can be used to influence actor-, audience-, and passer-by behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12). ACM, New York, NY, USA, 317-326. DOI=http://dx.doi.org/10.1145/2207676.2207720

42. Matthias Korn and Susanne Bødker. 2012. Looking ahead: how field trials can work in iterative and exploratory design of ubicomp systems. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (UbiComp '12). ACM, New York, NY, USA, 21-30. DOI=http://dx.doi.org/10.1145/2370216.2370221

43. Kai Kuikkaniemi, Giulio Jacucci, Marko Turpeinen, Eve Hoggan, and Jörg Müller. 2011. From Space to Stage: How Interactive Screens Will Change Urban Life. *IEEE Computer* 6, 44 (June 2011), 40-47. DOI=http://dx.doi.org/ 10.1109/MC.2011.135

44. Hannu Kukka, Heidi Oja, Vassilis Kostakos, Jorge Gonçalves, and Timo Ojala. 2013. What makes you click: exploring visual signals to entice interaction on public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13). ACM, New York, NY, USA, 1699-1708. DOI: http://dx.doi.org/10.1145/2470654.2466225 *

45. Scott Mcquire. 2010. Rethinking media events: Large screens, public space broadcasting and beyond. *New Media & Society* 12, 4 (June 2010), 567-582. DOI= http://dx.doi.org/10.1177/1461444809342764

46. Nemanja Memarovic, Marc Langheinrich, Keith Cheverst, Nick Taylor, and Florian Alt. 2013. P-LAYERS -- A Layered Framework Addressing the Multifaceted Issues Facing Community-Supporting Public Display Deployments. ACM *Transactions in Computer-Human Interaction* 20, 3, Article 17 (July 2013), 34 pages. DOI=http://dx.doi.org/10.1145/2491500.2491505 *

47. Daniel Michelis and Jörg Müller. 2011. The Audience Funnel: Observations of Gesture Based Interaction with Multiple Large Displays in a City Center. In *International Journal of Human–Computer Interaction* 27,6, 562-579. http://dx.doi.org/10.1080/10447318.2011.555299

48. Jörg Müller, Dennis Wilmsmann, Juliane Exeler, Markus Buzeck, Albrecht Schmidt, Tim Jay, and Antonio Krüger. 2009. Display Blindness: The Effect of Expectations on Attention towards Digital Signage. In *Proceedings of the 7th International Conference on Pervasive Computing* (Pervasive '09), Hideyuki Tokuda, Michael Beigl, Adrian Friday, A. J. Brush, and Yoshito Tobe (Eds.). Springer-Verlag, Berlin, Heidelberg, 1-8. DOI=http://dx.doi.org/10.1007/978-3-642-01516-8_1
49. Jörg Müller, Florian Alt, Daniel Michelis, and Albrecht Schmidt. 2010. Requirements and design space for interactive public displays. In *Proceedings of the 18th ACM international conference on Multimedia* (MM '10). ACM, New York, NY, USA, 1285-1294. DOI=http://dx.doi.org/10.1145/1873951.1874203
50. Jörg Müller, Robert Walter, Gilles Bailly, Michael Nischt, and Florian Alt. 2012. Looking glass: a field study on noticing interactivity of a shop window. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12). ACM, New York, NY, USA, 297-306. DOI=http://dx.doi.org/10.1145/2207676.2207718
51. Jörg Müller, Gilles Bailly, Thor Bossuyt, and Niklas Hillgren. 2014. MirrorTouch: combining touch and mid-air gestures for public displays. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services* (MobileHCI '14). ACM, New York, NY, USA, 319-328. DOI=http://dx.doi.org/10.1145/2628363.2628379
52. Jörg Müller, Dieter Eberle, and Konrad Tollmar. 2014. Communiplay: a field study of a public display mediaspace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14). ACM, New York, NY, USA, 1415-1424. DOI=http://dx.doi.org/10.1145/2556288.2557001 *
53. Kenton O'Hara, Maxine Glancy, and Simon Robertshaw. 2008. Understanding collective play in an urban screen game. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work* (CSCW '08). ACM, New York, NY, USA, 67-76. DOI=http://dx.doi.org/10.1145/1460563.1460576 *
54. Timo Ojala, Hannu Kukka, Tomas Lindén, Tommi Heikkinen, Marko Jurmu, Simo Hosio, and Fabio Kruger. 2010. UBI-hotspot 1.0: Large-scale long-term deployment of interactive public displays in a city center. In *Fifth International Conference on Internet and Web Applications and Services* (ICIW), IEEE, 285-294. IEEE. *
55. Timo Ojala, Vassilis Kostakos, Hannu Kukka, Tommi Heikkinen, Tomas Lindén, Marko Jurmu, Simo Hosio, Fabio Kruger, and Daniele Zanni. 2012. Multipurpose Interactive Public Displays in the Wild: Three Years Later. *IEEE Computer* 45, 5 (May 2012), 42-49. DOI=http://dx.doi.org/10.1109/MC.2012.115
56. Gonzalo Parra, Joris Klerkx, and Erik Duval. 2014. Understanding Engagement with Interactive Public Displays: an Awareness Campaign in the Wild. In *Proceedings of The International Symposium on Pervasive Displays* (PerDis '14), Sven Gehring (Ed.). ACM, New York, NY, USA, , Pages 180 , 6 pages. DOI=http://dx.doi.org/10.1145/2611009.2611020
57. Peter Peltonen, Esko Kurvinen, Antti Salovaara, Giulio Jacucci, Tommi Ilmonen, John Evans, Antti Oulasvirta, and Petri Saarikko. 2008. It's Mine, Don't Touch!: interactions at a large multi-touch display in a city centre. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '08). ACM, New York, NY, USA, 1285-1294. DOI=http://dx.doi.org/10.1145/1357054.1357255 *
58. Mark Perry, Steve Beckett, Kenton O'Hara, and Sriram Subramanian. 2010. WaveWindow: public, performative gestural interaction. In *ACM International Conference on Interactive Tabletops and Surfaces* (ITS '10). ACM, New York, NY, USA, 109-112. DOI=http://dx.doi.org/10.1145/1936652.1936672
59. Constantin Schmidt, Jörg Müller, and Gilles Bailly. 2013. Screenfinity: extending the perception area of content on very large public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13). ACM, New York, NY, USA, 1719-1728. DOI: http://dx.doi.org/10.1145/2470654.2466227
60. Ronald Schroeter, Marcus Foth, and Christine Satchell. 2012. People, content, location: sweet spotting urban screens for situated engagement. In *Proceedings of the Designing Interactive Systems Conference* (DIS '12). ACM, New York, NY, USA, 146-155. DOI=http://dx.doi.org/10.1145/2317956.2317980 *
61. Fabius Steinberger, Marcus Foth, and Florian Alt. 2014. Vote With Your Feet: Local Community Polling on Urban Screens. In *Proceedings of The International Symposium on Pervasive Displays* (PerDis '14), Sven Gehring (Ed.). ACM, New York, NY, USA, , Pages 44 , 6 pages. DOI=http://dx.doi.org/10.1145/2611009.2611015
62. Oliver Storz, Adrian Friday, Nigel Davies, Joe Finney, Corina Sas, and Jennifer Sheridan. 2006. Public Ubiquitous Computing Systems: Lessons from the e-Campus Display Deployments. *IEEE Pervasive Computing* 5, 3 (July 2006), 40-47. DOI=http://dx.doi.org/10.1109/MPRV.2006.56 *
63. Mirjam Struppek. 2006. Urban Screens – The Urbane Potential of Public Screens for Interaction. *Intelligent Agent* 6, 2.
64. Nick Taylor and Keith Cheverst. 2012. Supporting Community Awareness with Interactive Displays.

IEEE Computer 45, 5 (May 2012), 26–32. DOI=http://dx.doi.org/10.1109/MC.2012.113

65. Nick Taylor, Justin Marshall, Alicia Blum-Ross, John Mills, Jon Rogers, Paul Egglestone, David M. Frohlich, Peter Wright, and Patrick Olivier. 2012. Viewpoint: empowering communities with situated voting devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12). ACM, New York, NY, USA, 1361-1370. DOI=http://dx.doi.org/10.1145/2207676.2208594 *
66. Nick Taylor, Keith Cheverst, Peter Wright, and Patrick Olivier. 2013. Leaving the wild: lessons from community technology handovers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13). ACM, New York, NY, USA, 1549-1558. DOI: http://dx.doi.org/10.1145/2470654.2466206
67. Martin Tomitsch, Christopher Ackad, Oliver Dawson, Luke Hespanhol, and Judy Kay. 2014. Who cares about the Content? An Analysis of Playful Behaviour at a Public Display. In *Proceedings of The International Symposium on Pervasive Displays* (PerDis '14), Sven Gehring (Ed.). ACM, New York, NY, USA, , Pages 160 , 6 pages. DOI=http://dx.doi.org/10.1145/2611009.2611016 *
68. Nina Valkanova, Robert Walter, Andrew Vande Moere, and Jörg Müller. 2014. MyPosition: sparking civic discourse by a public interactive poll visualization. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (CSCW '14). ACM, New York, NY, USA, 1323-1332. DOI=http://dx.doi.org/10.1145/2531602.2531639 *
69. Daniel Vogel and Ravin Balakrishnan. 2004. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of the 17th annual ACM symposium on User interface software and technology* (UIST '04). ACM, New York, NY, USA, 137-146. DOI=http://dx.doi.org/10.1145/1029632.1029656
70. Robert Walter, Gilles Bailly, and Jörg Müller. 2013. StrikeAPose: revealing mid-air gestures on public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13). ACM, New York, NY, USA, 841-850. DOI: http://dx.doi.org/10.1145/2470654.2470774
71. Robert Walter, Gilles Bailly, Nina Valkanova, and Jörg Müller. 2014. Cuenesics: using mid-air gestures to select items on interactive public displays. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services* (MobileHCI '14). ACM, New York, NY, USA, 299-308. DOI: http://dx.doi.org/10.1145/2628363.2628368

* Part of the taxonomy of external factors.

All referenced papers were part of the literature review.

Paper II

Mäkelä, V., Heimonen, T., & Turunen, M. (2017). Semi-Automated, Large-Scale Evaluation of Public Displays. *International Journal of Human–Computer Interaction*. doi:10.1080/10447318.2017.1367905

This is an Accepted Manuscript of an article published by Taylor & Francis in International Journal of Human-Computer Interaction on 8/24/2017, available online: http://www.tandfonline.com/doi/full/10.1080/10447318.2017.1367905.

SEMI-AUTOMATED, LARGE SCALE EVALUATION OF PUBLIC DISPLAYS

Ville Mäkelä[1], Tomi Heimonen[2], Markku Turunen[1]

ville.mi.makela@sis.uta.fi, tomi.heimonen@uwsp.edu, markku.turunen@sis.uta.fi

[1]Tampere Unit for Computer-Human Interaction (TAUCHI),
University of Tampere, Tampere, Finland

[2]Department of Computing and New Media Technologies,
University of Wisconsin-Stevens Point, Wisconsin, USA

Abstract: We present a scalable, semi-automated process for studying the usage of public displays. The process consists of gathering anonymous interaction and skeletal data of passersby during public display deployment and programmatically analyzing the data. We demonstrate the use of the process with the analysis of the Information Wall, a gesture-controlled public information display. Information Wall was deployed in a university campus for one year and collected an extensive data set of more than 100 000 passersby. The main benefits of the process include (1) gathering of large data sets without considerable use of resources, (2) fast, semi-automated data analysis, and (3) applicability to studying the effects of long-term public display deployments. In analyzing the usage and passersby data of the Information Wall in our validation study, the main findings uncovered using the method were (i) most users were first-time users exploring the system, and not many returned to use the system again, and (ii) many users were accompanied by passive users who observed interaction from further away, which could suggest a case of multi-user interaction blindness. In the past, logged data has mainly been used as a supporting method for in situ observations and interviews, and its use has required a considerable amount of manual work. In this article, we argue that logged data analysis can be automated to complement other methods, particularly in the evaluation of long-term deployments.

Keywords: *Public displays, evaluation, logged data, observation, pervasive displays, long-term studies, deployment-based research.*

INTRODUCTION

Evaluating public displays and their users in a real-world setting is challenging. Evaluations often rely in mixed-method approaches, combining observations, interviews, and interaction logs [3]. This requires significant resources in time and personnel, which calls for the development of analysis methods that can automate some or all of the manual steps.

As is noted in past research, interaction with public displays is a multi-faceted process and can be considered to have begun even before the user directly interacts with the display [12]. This is especially true in the case of gesture-controlled interfaces, where detection of a passerby can be utilized to trigger visual interaction cues to entice use. Hence, being able to analyze not only users but also other passersby can give important insights into the design of public display applications, and cannot be achieved by traditional interaction logs. Field observations, on the other hand, while usually effective and often the preferred method in in-the-wild studies, are time-consuming and limited by human capabilities in terms of how much, and what kind of, data can be meaningfully gathered.

To supplement the currently prevailing research methods, we present a semi-automated process for evaluating public displays both extensively and without significant use of resources. Our method consists of automated gathering of a large set of interaction and skeletal data of passersby during the deployment of a public display and the subsequent computational analysis of the data. In this article, we investigate the benefits and limitations of the proposed approach. Our starting point was to investigate whether we could programmatically analyze the collected data to reach findings that we would have likely identified if we had been on-site observing users the whole time. At the same time, we envisioned that the semi-automatic analysis can help capture results not achievable with human observers alone, when cognitive limitations and time constraints are eliminated.

Logged skeletal motion data has been collected for analysis purposes in past research [1;16;28], however it has not been used to its full extent. Primarily, such data has acted as supporting evidence for other methods such as observations and interviews, and has involved a considerable amount of work in the form of manually going through the recorded depth sensor data. For clarification, in this article, *skeletal data* refers to simple, anonymous location data of skeletal joints tracked in 3D space, e.g., shoulders, elbows, and hands of passersby.

To investigate the semi-automated approach, we used it to analyze the long-term deployment of a gesture-controlled public display application, the Information Wall [11] (Figure 1). We deployed the Information Wall in a large indoor space at a university campus for one year, during which we recorded all interaction data from the system as well as skeletal data from the Microsoft Kinect sensor used to control the display. The resulting large data set contained traces of more than 100,000 passersby. Using the semi-automated process, we were able to produce many meaningful results. For instance, we found that many users interacting with the system were accompanied by *non-interacting* people. With this large data set, we also show that the process scales well with large-scale public deployments.

Figure 1. Two users interacting with the Information Wall at the public deployment setting.

In this article, we discuss the possibilities, benefits and challenges of utilizing interaction and skeletal data in semi-automated usage analysis. The proposed approach has several benefits to other data collection approaches, and can be considered an ecologically valid method for evaluating public displays. We also discuss the relationship between our approach and conducting field observations, and how the two evaluation methods can support each other. We demonstrate the semi-automated process in practice with the Information Wall deployment.

The remainder of this paper is organized as follows: first, we present related work in which we focus on how past research has gathered data and evaluated public displays. Then, we present the Information Wall prototype in detail as well as the long-term deployment setting. Next, we present the four phases of the semi-automated process and discuss each of them: *data collection*, *preparation*, *feature extraction*, and *analysis*. Following, we present the results of applying the semi-automated process to the Information Wall deployment data to provide an example of the type of insights that can be acquired. Finally, we discuss the characteristics, benefits, and challenges of the semi-automated approach, and conclude with a discussion of future work.

RELATED WORK

In this section, we present existing work relevant to this study. First, we show the need for improved and less time-consuming research methods for large display evaluations. Then, we discuss frameworks and different phases of interaction with large displays that we utilize when analyzing the data from our long-term large display deployment.

Approaches to Evaluating Users and Public Displays

Müller *et al.* [15] note that the majority of public display evaluations are descriptive field studies, and recognize five categories of metrics that are commonly used in quantitative evaluations of public displays: (1) absolute number of users is used e.g. to determine the number of interactions or views towards a display; (2) percentage of users out of the total number of users showing a certain behavior; (3) absolute number of interactions e.g. to determine how often an application was started; (4) duration of interactions to determine how long the interaction in general or a specific type of interaction lasted; (5) number of simultaneous users is used to round up the description of usage.

Williamson and Williamson presented the open source Pedestrian Tracker tool [29], which utilizes a camera to track passersby using motion detection and background subtraction. The tool can be used, e.g., to recognize directions from which pedestrians approach the display the most and how passersby adjust their route when a public display is installed in a space. The tool's strength is that it is separate from the public display installation and can thus be deployed in spaces from which a large area can be seen and analyzed, e.g. high above the public display. However, passerby movement is only one of the features we aim to include in our approach, and recognizing features such as gestures of a particular user with the same tool is difficult or even impossible by utilizing a camera that is set in a location with a good view of a large area.

Some past studies have gathered extensive log data via a motion detection sensor in gestural or proxemic interfaces. Most studies have combined the approach with methods such as observations and interviews, and only performed lightweight, manual analysis of the motion data. Ackad *et al.* [1] presented the Media Ribbon, a gesture-controlled public information display relatively similar to the Information Wall. They included a lightweight analysis of the data provided by the installation's motion detection sensor, such as session duration, performed gestures, and the number of people in front of the display. In addition, they manually analyzed the depth data to evaluate how users interacted with the system. Similarly, Müller *et al.* [16] recorded depth data from six different locations and converted it into simple behavioral variables such as time of entry and exit. However, classifying passersby into categories was done manually. Walter *et al.* [28] captured raw depth video to support their on-site observations. However, they too reviewed the video data manually. Similar analysis was conducted by Schmidt *et al.* for their Screenfinity display [24].

Based on prior work, it seems that not many public display deployments utilize automatically gathered interaction and skeletal data. The few studies that do mostly utilize the data to support existing findings and/or invest a considerable amount of manual work, such as manually reviewing video logs of the installation.

The absence of large-scale data collection in public deployments is likely explained by their research-focused nature. The deployments tend to be relatively short, and focus on investigating specific phenomena. With such installations more frequently becoming a part of the urban environment, the need for automated evaluation methods is emphasized to provide a cost-effective assessment method. Therefore, we investigated if (a) interaction and skeletal data can be semi-automatically analyzed and (b) whether said data could lead to actionable findings, and consequently act as the primary research method in the quantitative assessment of public displays.

User Behavior with Public Displays

People tend to notice public displays more easily and be more interested in them when there are already people using the display. This effect is known as the *honeypot* effect [4]. Furthermore, Müller *et al.* [16] conducted a study on public displays in six different locations and not only found a significant rise in users through the honeypot effect, but also found an increase in the duration of interaction.

Walter *et al.* note that a public setting results in most users being first time users, and people are prepared to invest only a short period of time to investigate a display [27]. In their following work, Walter et al. [28] further argue that immediate usability and clear interactions are key concepts for a public display. In addition, Marshall *et al.* [9] noticed that people trying out public displays tend to be impatient: interaction usually ends if users do not succeed with what they are trying to achieve.

One of the most influential frameworks describing public display interaction, the Audience Funnel, was presented by Michelis and Müller [12]. The framework divides public display interaction into six phases:

1. **Passing by.** The user is in the same space with the display with no intention of interacting with it.
2. **Viewing and reacting.** The user glances at or reacts to the display.
3. **Subtle interaction.** The user commits an action, e.g. waves his hand, to see what effect it causes on the display.
4. **Direct interaction.** The user begins interacting with the display in more depth.
5. **Multiple interaction.** The user leaves and comes back after a while, or switches to another display if multiple displays are available.
6. **Follow-up actions.** The user e.g. takes pictures of the display or himself interacting with it.

Mostly based on the Audience Funnel framework, Müller *et al.* [14] present three issues that public displays specifically need to address. First, the audience is not necessarily even aware of the public display in the first place (*display blindness*), or they might not be aware that the display can be interacted with (*interaction blindness*). It is thus important that the display aims to catch the attention of passersby in some way. Second, users need to be motivated to start interacting with the display. It should be noted that users typically are not specifically looking for a public display but rather stumble upon it. Therefore, displays should offer ways to pass time or contain information that is relevant to the user. Third, the fact that interaction with the display happens in public should be accounted for. For example, people may avoid interaction completely or partially because of their role (such as police officer or custodial services), physical limitations (for example, an elderly person not being able to commit certain gestures), or other traits (being afraid of public embarrassment).

Parra *et al.* [21] add a fourth issue that public displays should address: users should reach a goal or “final stage” of interaction with the system. In many cases, the goal may be straightforward, for example to provide the user with information (s)he is looking for, or as is the case with playful installations, to make the user have a good time. As a more complex example, Parra *et al.* [21] presented an interactive display developed as part of an awareness campaign. Users were provided with a web address after they had successfully interacted with the display, with the aim of having the users visit the website. In this case, the success of the installation was partly measured by how many users actually visited the site.

Akpan *et al.* [2] conducted a large-scale study on the effects of space and place on social behavior around public displays. Their findings suggest that an optimal social context can encourage interaction and help overcome several issues related to public displays, even if the display is deployed in a poor physical space. For example, entertainment-oriented environments can encourage people to engage in playful behavior and to try out new systems.

We will utilize the related work presented here when evaluating and discussing our public display deployment later in this article. Most notably, we will aim to classify passersby based on the Audience Funnel framework [12], and evaluate the use of our public display with respect to phenomena such as the honeypot effect [4] as well as display blindness and interaction blindness [14].

INFORMATION WALL

Information Wall[1] is a gesture-controlled public information display [11] that contains information relevant to the deployment location, such as the lunch menus of nearby restaurants as well as events taking place around the campus and the city. In the following, we describe the interaction design of the application, discuss the rationale behind the design choices, and present the system's one-year deployment and setup.

Reacting to Users

We wanted to investigate different ways to react to passersby in an effort to attract people to explore the display. Thus, when no users are present, the wall displays a static background image and a dialog encouraging users to try out the system (Figure 2A). When a user passes by or approaches the display, a subtle reaction is activated: a rectangular shape appears on the screen and reacts to the user's movement (Figure 2A). The shape moves horizontally along with the user, and grows as the user gets closer and shrinks as the user gets further away. Whenever the user gets close enough to the display (less than 2.8 meters away), a three-dimensional information cube is opened on the screen (direct reaction) (Figure 2B).

The system supports two simultaneous users. In the case of one user, the single information cube is placed in the middle of the screen. When another user steps in, the first cube scales down to make room for another information cube, which appears from the side corresponding to the new user's location (Figure 2C). A cube is closed whenever the corresponding user leaves the scene, and the leftover cube readjusts itself to make use of the whole screen.

[1] https://youtu.be/YPhIqw5Vrz8

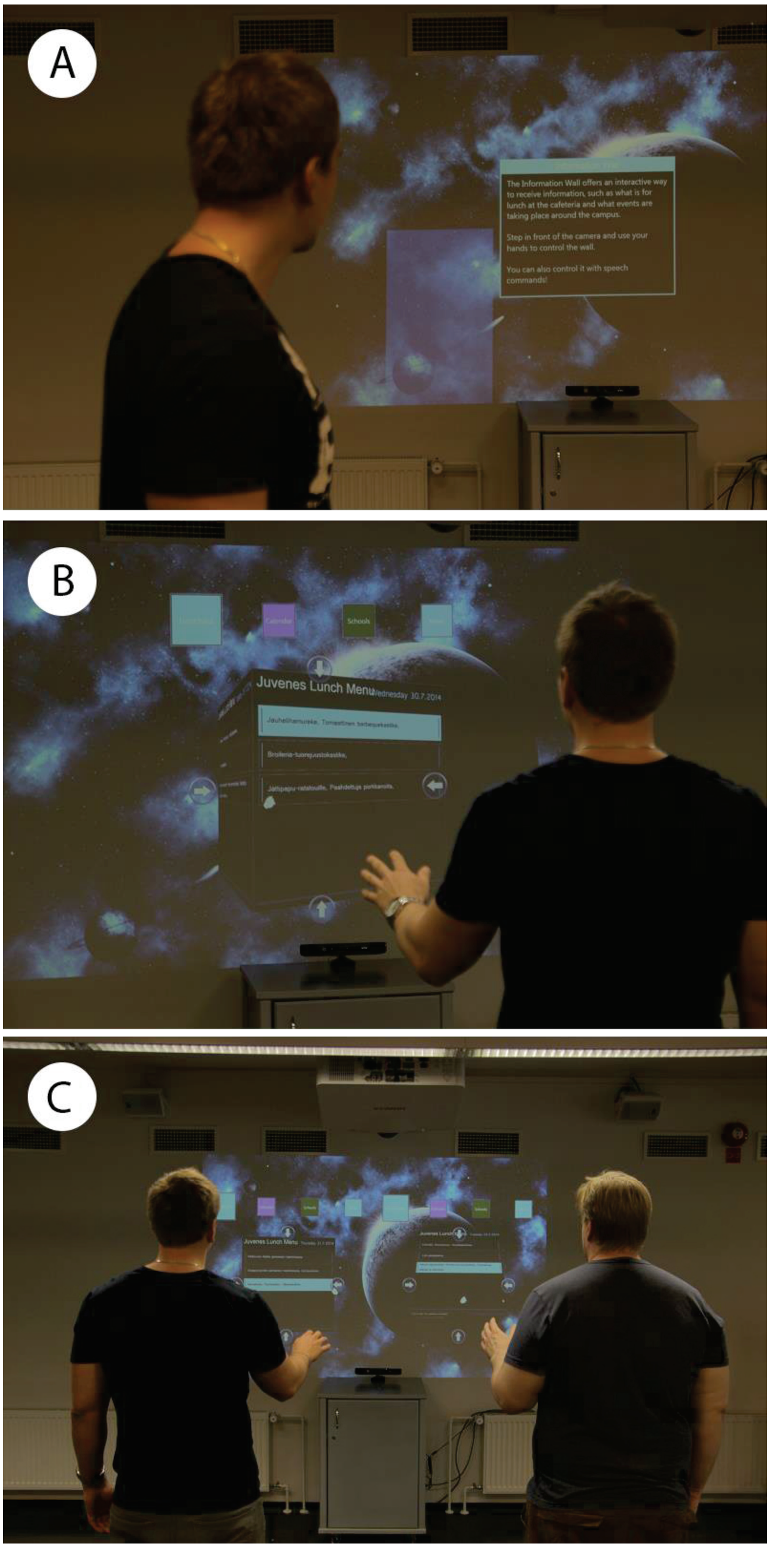


Figure 2. A) A user is tracked with a rectangular shape on the screen. An instruction dialog is displayed. B) A user interacts with an information cube. C) Two users simultaneously interacting with the wall. (Adapted from Mäkelä *et al.* [11])

Gesture Control

Users mainly interact with the display via an on-screen cursor that moves according to where the user is pointing with their hand. Pointing uses the physical interaction zone algorithm provided by the Kinect SDK[2]. All gesture interactions are one-handed interactions, however both hands are tracked and two cursors are displayed on the screen if the user is pointing with both hands. Either hand can be used to interact and the active hand can be changed on the fly.

The content of the display is navigated by rotating the information cube in desired directions. Rotation involves a point-and-dwell [8] selection followed by a swipe gesture. This simulates the rotation of a real-world physical object. The user first triggers a button at the edge of the cube by pointing towards it for a short period of time, after which the users swipes towards the opposite edge of the cube to rotate it. Users can cancel the rotation by swiping back towards the edge that they started from, i.e. away from the cube.

When a rotation button is triggered, an arrow animation is played to point to the direction of the rotation. Rotating to the right and left will change to the next and previous view inside the current section. For example, one can switch from today's lunch menu to tomorrow's lunch menu. Rotating up and down will change the section, for example from the lunch menu to the latest news.

Other functions on the display and launched by utilizing the same point-and-dwell method, but without the additional swipe gesture. During dwelling a circular animation is displayed on the particular button to indicate that it is being triggered.

In addition to rotating the cube, users can use shortcut buttons on top of the cube to quickly access desired sections. Moreover, detailed information, such as the ingredients of a dish or the full story of a news headline, is accessed by triggering the corresponding entry from the face of the cube, which will open a separate popup dialog. The dialog also contains a voting system in the form of two buttons, through which users can give a thumbs up or a thumbs down for a dish or an event. Cast votes are then displayed on the corresponding entry. The dialog is closed by triggering a close button in the bottom corner.

Design Rationale

We aimed to follow the Audience Funnel framework [12] in the interaction design. The rectangles following the passersby were meant to catch the initial attention. During the subtle interaction phase, the user would a) see the information cube appear and b) see the cursor move around the screen when the user tried out something simple, like waving his/her hand.

Despite a growing number of public displays being deployed, few of them utilize mid-air gestures. Considering the novelty of the interaction modality, we decided to work with a cursor-based application. We hypothesized that a cursor would appear as something familiar from other environments, as well as Kinect-based games, and not put off users so easily.

We chose dwelling as the target selection technique as it does not require specific gestures and it has been found to be intuitive [28]. We used a dwell time of 1.5 seconds for all targets expect for the section shortcuts, which were reduced to 1.2 seconds. This was because section shortcuts were on the upper edge of the screen and thus users were less likely to accidentally hover over them or pass through them. In addition, to make target selection with the cursor

[2] https://developer.microsoft.com/en-us/windows/kinect

easier, we utilized the magnetic cursor technique [10], which automatically snaps to a target that is close enough and moves slower while on a target to make accurate pointing easier.

We aimed to fill the system with information that would be either beneficial or interesting to people spending time in or moving through the space. In the context of a university campus, we displayed daily lunch menus for a student cafeteria in the same building as well as information about public events such as talks from guest lecturers. In the third section, we displayed information about the different schools at the university. The purpose of this section was mainly to demonstrate the capabilities of the system, displaying large images along with paragraphs of text. Later during the deployment, we introduced a fourth section in which recent news from a popular Finnish news portal were displayed.

Prior studies have found that users often interact in groups [1] and are more likely to be interested in a public display if there are already people interacting with it (*honeypot effect*) [4]. To support this, we designed the system to support two simultaneous users. However, due to the novelty of gesture-control, we included user-specific information cubes, with the rationale that the cubes would clearly visualize which part of the display a user is controlling.

Implementation and User Tracking

The system is developed on .NET Framework 4.0 using the Windows Presentation Foundation (WPF) library and the 3DTools plugin[3]. The system utilizes the Microsoft Kinect for Windows sensor for tracking users' location and hand coordinates.

Data exchange between the Kinect sensor and the Information Wall application is handled by a dedicated socket-based middleware component that utilizes the Kinect SDK. The data used for tracking users is anonymous, in that we only track the position of the users' upper body joints in 3D space as well as the pointing direction of their hands relative to the screen. This same skeletal data is used for logging and the analysis presented in this article. The format of the data is described in more detail in the next chapter.

The majority of the Information Wall's content is fetched from external sources using public APIs and RSS feeds, although some content is parsed from external websites.

Deployment

We set up the Information Wall in a semi-public location at the university campus to run a longitudinal, in-situ study focused on investigating naturalistic usage. The installation location is a large open space on the lower floor of the main building of a local university.

The layout of the deployment space is presented in Figure 3, and a panorama of the space from the perspective of the installation is presented in Figure 4. The space contains a cafeteria where students and staff have lunch and take breaks between classes. In addition, the space is adjacent to a large auditorium in which lectures and exams are held regularly. Consequently, students often wait around the space for entry into the auditorium. During semesters, areas surrounding the display are accessed by hundreds of students and staff daily.

[3] https://3dtools.codeplex.com/

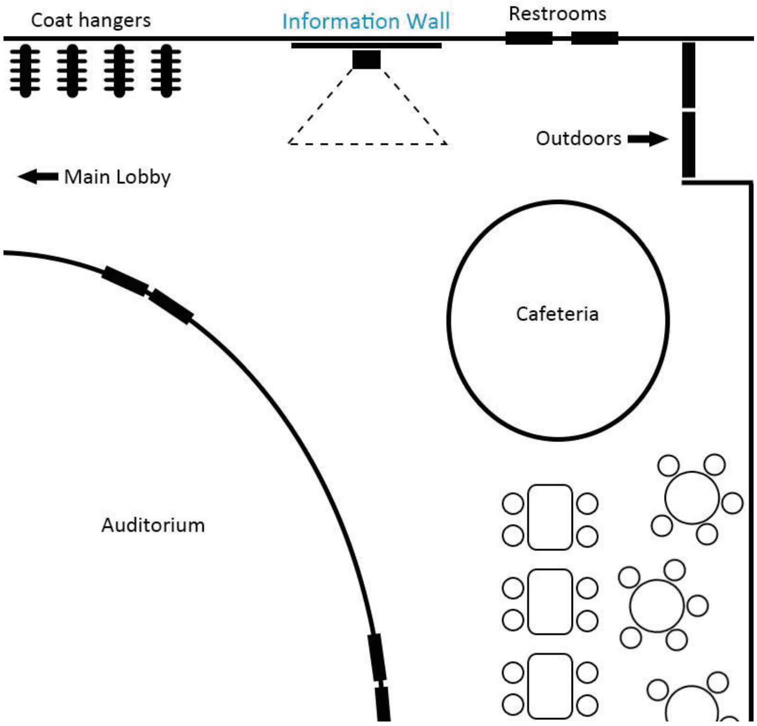


Figure 3. Floor plan of the deployment space.

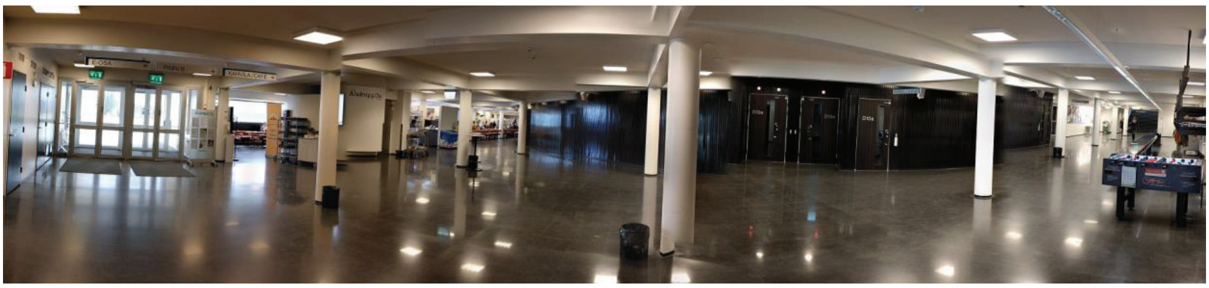


Figure 4. Panorama of the space from the installation's perspective.

The Information Wall was projected on a wall surface with a ceiling mounted projector with a resolution of 1920 x 1080 pixels, which provided approximately 2.5 meters wide display area (see Figure 1). Interface sounds and audio landscape were provided by two active loudspeakers mounted to the sides of the display at ceiling height. The Kinect sensor was positioned at the horizontal midpoint under the projection display, providing a fixed point of reference for the collected data (all coordinates are relative to this point).

The public deployment data set was captured between April 2013 and April 2014. In total, it includes user and interaction traces collected from 210 distinct days, with a total number of passersby being 106,637. The installation was running on most weekdays, but was turned off during weekends, holidays, and occasional maintenance breaks.

SEMI-AUTOMATED, LARGE SCALE EVALUATION PROCESS

In this section, we present the semi-automated evaluation process, which we used to evaluate the Information Wall system. The process is visualized in Figure 5, and consists of four phases: data collection, preparation, feature extraction, and analysis. In the following, we will describe the four phases in detail.

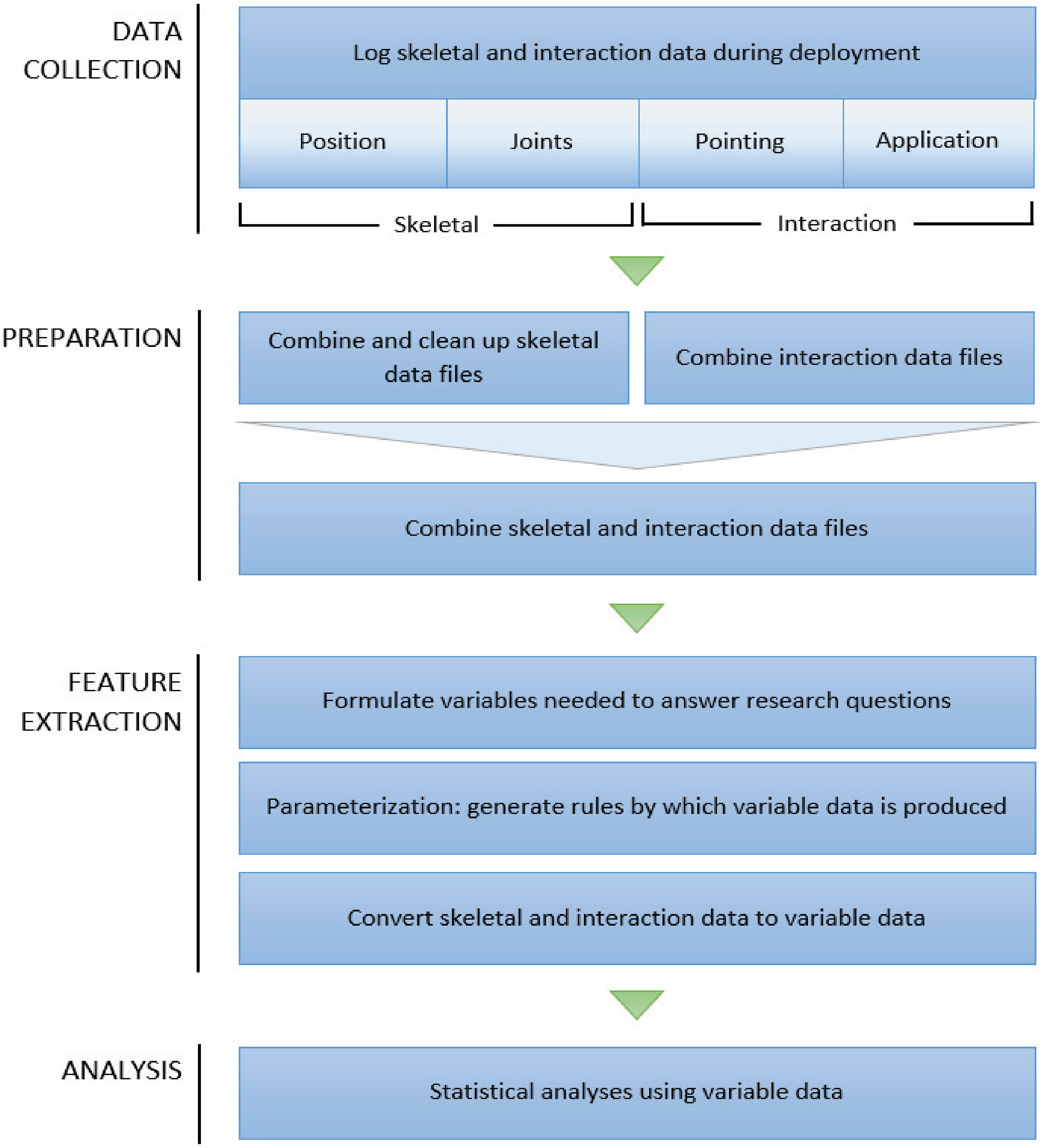


Figure 5. The semi-automated process consists of four primary phases: data collection, preparation, feature extraction, and analysis.

Phase I: Data Collection

The *data collection* phase should be implemented as a part of the deployed application. In our case, the Information Wall system was set to automatically collect data during the deployment. The data can be roughly separated into two categories: skeletal and interaction data (Figure 5). In our case, skeletal data includes the passerby's general location as well as the location of each upper-body joint in 3D space. We did not track nor record the lower body as it was not needed for interaction or the analysis. Interaction-related data consists of the passerby's hand-pointing coordinates relative to the screen, as well as interaction events in the application, including information cube activations, target hovers and triggers, etc. All log entries were saved with both client-side and server-side time stamps.

We aimed to capture data of all passersby within the limitations of the Kinect sensor. The Kinect allowed simultaneous tracking of a maximum of six people. It is possible that some passersby could not be recorded by the sensor if more than six people appeared at the scene, although this was presumably very rare. The two closest passersby were treated as potential users and were tracked in full detail as described above, while for the remaining four passersby only a general location was recorded.

An example log entry of a single skeleton is shown in Figure 6. Different information is separated using a pipe character. Each log entry is started with client side and server side timestamps. The "_U_" communicates the beginning of user information, followed by the user ID and the x, y and z position ("*42929;-156;-279;3223*"), followed by the pointing coordinates for both hands ("*-3,612464;1,684637;6,115403;3,989488*"), followed by upper body joint IDs and x, y, z coordinates. The positions are reported in millimeters from the sensor's location, x being the horizontal axis, y the vertical axis, and z being the distance from the sensor.

```
2014-03-24 12:48:4.123|2014-03-24 12:48:04.127|_U_|42929;-156;-279;3223|-
3,612464;1,684637;6,115403;3,989488|6;-133,7465;175,0641;3285,178|7;-130,5533;-
2,498331;3175,068|9;-93,2767;-345,8763;3177,252|!12;56,86545;-18,49249;3427,463|13;-
207,9531;-4,714899;3464,396|15;90,94656;17,26251;3419,62
```

Figure 6. A log entry for a single user.

During data collection, data was stored in separate files by date, and we further stored skeletal data and system event logs in separate files. This resulted in a collection of hundreds of large log files. This was done for ease of backup and to avoid data loss due to possible system crashes. We note here that the amount of automatically collected data - primarily skeletal data - may be surprising during public deployments. Gathering skeletal data several times per second (every 60 milliseconds in this case) for every person in the scene quickly results in heaps of data, even from a short interaction session.

Phase II: Preparation

In the *preparation* phase (Figure 5), the logged data is processed into a format that can be more easily handled. In our analysis, we wrote C# scripts to transform the large set of collected data files. First, we combined all skeletal data into a single file, which had to be done in several parts due to our machines running out of memory. Similarly, all interaction data was combined into one file. Next, we filtered and cleaned up the skeletal data, as the original data set contained raw skeletal data collected of all detected people at 60 milliseconds interval. To significantly reduce the number of data points for follow-up processing, we converted all skeletal entries into one entry from the period when a passerby or a user was standing still without interacting. Finally, we combined the skeletal and interaction data into one file, connected interactions to specific tracked users (skeletons), and ordered the entries based on their timestamps.

It is worthwhile to note that since the skeletal data is rather generic in nature, the tools created for preparation and analysis should require little to no modification between studies. Interaction data, on the other hand, is likely to be more application-specific. With our deployment, interaction data consisted of the users' pointing coordinates as well as event triggers in the application. In other applications, pointing data could be replaced or accompanied by gesture data.

Phase III: Feature Extraction

The *feature extraction* phase (Figure 5) is most crucial and characteristic part of the proposed process. By feature extraction, we mean the identification and calculation of relevant variables from the data, on which statistical analyses can be performed. To produce meaningful insights from the data, researchers must at this point know what they are looking for, and convert the data into a form that supports relevant analysis procedures.

It is generally good scientific practice to be aware of what is being researched from the very beginning of a study. We agree with this; therefore, we emphasize that in the feature extraction phase, we refer to the definition of low-level variables and parameters, not the formulation of actual research questions. Thus, feature extraction deals with the very specifics of *how* to meaningfully answer the research questions. However, we also note that in some cases the different phases of the analysis process may not be carried out by the same party. For instance, data collected during a public deployment could be published for other researchers to utilize. In such a case, the research questions would not exist during data collection, but would instead have to be formulated during the preparation and feature extraction phases.

The feature extraction phase begins by defining the *variables* needed to answer research questions. Then, one must decide how exactly the variables should be produced or calculated using the data available. We refer to this as *parameterization*: the definition of a set of rules by which a variable is computed.

We provide two examples to better demonstrate variables and parameterization. As the first example, we wanted to see how many people approached the Information Wall from each direction. Consequently, this required that a variable, direction of entry, needed to be calculated for all users by utilizing skeletal data in the data set. For defining the direction of entry, then, we needed to decide how to operationalize the calculation in terms of the available logged data.

We first calculated the angle of movement during the first 0.5 seconds for when the passerby was visible. If the passerby was visible for less than 0.5 seconds, or if the passerby moved less than 10 centimeters during that period, the angle of movement was left undefined. We then distributed passersby to three categories (left, right, and front) based on their angle of movement, each direction forming a 90-degree sector.

For the parameterization of the direction of entry, we had to account for several factors specific to the deployment. Firstly, the Kinect sensor was occasionally slow to recognize passersby who moved sideways. For instance, a passerby entering from the right could be recognized when they were already close to exiting from the left edge. Therefore, the exact location of the passersby could not be a parameter in calculating the direction of entry. Second, given that nothing else that would make passersby stop or change direction was located within the Kinect sensor's field of view, we decided that a simple calculation of the movement angle would be sufficient for our purposes. Moreover, given the small field of view of the sensor relative to the size of the whole space, we deemed it unlikely that the movement patterns of passersby in front of the display would be complicated, unless affected by the display itself.

The second example demonstrates a variable that could not be successfully parameterized. We were interested to see if our data could be used to identify passersby who looked at or reacted to the display, but did not stop or interact with it (i.e., the "viewing and reacting" phase in the Audience Funnel framework [12]). However, during the deployment, we had come to observe that passersby looking at the display would simply turn their head towards the display while walking past it, and the rest of the body remained relatively unaffected by this action. Considering that we only recorded the position of body parts/joints in 3D space (and not their rotation), we concluded that we could not parameterize whether passersby looked at the display. Therefore, we focused on the following phases of the Audience Funnel framework – subtle and direct interaction. This classification will be discussed in the following chapter, where we also show application-specific parameters for the classification.

As an important takeaway from our examples, we note that manual observations play a role in successful parameterization, particularly in identifying factors to account for when calculating variables. This requires observing not only the people directly, but also the physical space, and identifying factors that might affect the variables. For instance, a pillar right next to the display blocking a walking path might affect how a direction of entry should be calculated. It is likely that researchers spend time on-site in the beginning of the deployment when setting up the system and making sure no technical faults are taking place. This phase of the deployment can be utilized for parameterization, and therefore, the features that researchers want to extract should be kept in mind in the early phases of the deployment.

For the feature extraction phase, we wrote another script to analyze the combined log file produced in the preparation phase and extract all the information we wanted into a comma separated (CSV) file. We stored each passerby in separate rows with relevant variables. A subset of the resulting variables is presented in Figure 7.

As we pointed out in the preparation phase, the collected data is partly generic and partly application-specific. During feature extraction phase, additional characteristics, such as those pertaining to the properties of the deployment space, need to be considered. We provided an example of defining a direction of entry for passersby – the exact same parameters could work for some deployments, but not necessarily for others.

	A	B	C	D	E	F	G	H	I	J	K
1	Date	Id	Session	Hovers	Triggers	Entry dir.	Exit dir.	Dur. (total)	Category	Direct users in group	Subtle users in group
2	4/8/2013	1	3	0	0	Unknown	right	3	passing by	0	0
3	4/8/2013	2	3	3	2	Unknown	right	35	direct	0	0
4	4/8/2013	3	3	46	12	left	Unknown	129	direct	1	0
5	4/8/2013	4	150	1	0	Unknown	Unknown	6	subtle	0	0
6	4/8/2013	5	150	0	0	Unknown	right	1	passing by	0	1
7	4/8/2013	6	150	1	0	left	right	27	subtle	0	0
8	4/8/2013	7	150	0	0	left	right	1	passing by	0	1
9	4/8/2013	8	150	0	0	left	right	3	passing by	0	2
10	4/8/2013	9	184	0	0	right	left	3	passing by	0	0
11	4/8/2013	10	187	3	0	Unknown	Unknown	8	subtle	0	0
12	4/8/2013	11	199	0	0	left	right	2	passing by	0	0
13	4/8/2013	12	202	0	0	Unknown	left	1	passing by	0	0
14	4/8/2013	13	206	0	0	right	front	3	passing by	0	0
15	4/8/2013	14	206	1	0	left	right	1	subtle	0	0
16	4/8/2013	15	216	0	0	right	left	3	passing by	0	0
17	4/8/2013	16	222	1	0	right	left	2	subtle	0	0
18	4/8/2013	17	222	0	0	left	right	1	passing by	0	0

Figure 7. Some of the variables produced during feature extraction.

Phase IV: Analysis

Finally, using our CSV file (Figure 7) we were able to run statistical *analyses* (Figure 5) using software such as Microsoft Excel or SPSS. At this point, running any kind of analysis based on the variables defined and generated in the previous phase was relatively straightforward. We expect that this phase of the process will be specific to each particular deployment. Additional data processing may be necessary to accommodate different statistical methods (e.g., switching between wide and long data formats).

Summary

The semi-automated evaluation method consists of four primary phases: data collection, preparation, feature extraction, and analysis. *Data collection* is carried out automatically by the system during the public deployment. The key notions here are that a) skeletal data is collected, i.e., we go beyond simply logging interactions with the display, and b) the amount of data is likely extensive and consequently non-trivial to handle as-is.

The *preparation* phase largely acts as a bridge between data collection and the remaining phases, and is meant to ease the next steps of the process by a) combining the large set of data files into one, and b) filtering out and aggregating the data.

The *feature extraction* phase is the crucial phase in which the potential to answer research questions is defined. Researchers must decide which characteristics of the data they are interested in, define the variables needed, and define the logic by which these variables are produced (i.e., parameterization). We find that such parameters can often be unique to the deployment – researchers must be familiar with the system being evaluated as well as the space in which it is deployed, and decide what makes sense for that particular installation. A valuable insight here is that quick observations play a role in parameterization, in identifying factors to account for when calculating variables.

The *analysis* phase refers to statistical analyses run using the resulting data from the feature extraction phase. We converted the data into a single CSV file, after which running the analyses was relatively trivial. The analysis phase is not unique to this process, but is nonetheless necessary.

In summary, despite the semi-automated nature of our method, it does not automatically provide researchers with answers to their research questions. The first two phases of the process, *data collection* and *preparation*, are relatively straightforward. They can be automated to a high degree, and potentially require only little adaptation between studies. The *feature extraction* phase is where most of the work goes, and which is strongly dependent on the characteristics of the public display deployment as well as the research being conducted. The last phase, *analysis*, involves statistical analysis like with any quantitative data, and is therefore not unique to this process.

INFORMATION WALL ANALYSIS RESULTS

In this section, we present the results of applying our analysis approach to the Information Wall data. First, we divide passersby into three user types, after which we analyze movement and reactions of the passersby. Finally, we present differences in the behavior of individual users and users belonging to a group.

User Classification

Based on the Audience Funnel [12], we aimed to recognize subtle users and direct users within the passersby. The classification parameters are presented in Table 1. We categorized passersby as direct users if they triggered at least one target, as triggering a target required stable pointing towards the screen and was unlikely to happen without deliberate interaction with the display. Subtle users were those who did not trigger any actions, but hovered over (pointed at a target, but not long enough for a trigger) a minimum of two targets. We required more than one hover as we observed that occasionally a person walking past the system would result in the bottom-most element being hovered over briefly, and we wanted to exclude these situations from the analysis. This requirement is relatively strict; as an alternative, we could have filtered out very short single-hovers and still classify the remaining single-hover users as subtle users. This would have likely increased the number of subtle users in our data set. However, we aimed to keep the analysis simple for this case study, while more advanced analyses are certainly possible. Finally, to further ensure that no passersby were accidentally classified as users, we required that both subtle and direct users spent a minimum of two seconds in front of the screen.

All other passersby were defined as passive users. As discussed in the previous section, we excluded the “viewing and reacting” [12] category from our analysis due to some technical limitations as well as limitations in our data set. We discuss overcoming these limitations towards the end of this article.

Table 1. Parameters for user classifications.

	Time in the area	Hovers	Triggers
Direct	>= 2 s	>= 1	>= 1
Subtle	>= 2 s	>= 2	0
Passive	>= 0 s	<= 1	0

Based on this classification scheme, we identified a total of 98907 passive users (92.8% of all passersby), 6241 subtle users (5.9%) and 1489 direct users (1.4%).

In total, users hovered over targets on the screen 68707 times, and triggered 10399 targets. Direct users triggered an average of 6.8 targets (SD = 8.6). Passive users were visible for an average of 2.2 seconds (SD = 9.0), subtle users for 8.0 seconds (SD = 26.6), and direct users for 70.9 seconds (SD = 80.9).

The average amount of target triggers as well as the average duration is surprisingly high. We hypothesized that many direct users would come to check out the next day's lunch menu and then leave the scene, and thus we expected to see most direct users trigger only one or two targets and spend only a short period of time in the area. Of the 1489 direct users, a relatively substantial amount (556 users, 37.3%) did indeed trigger only one or two targets; however, the amount is not as significant as we expected and said users still spent an average of 46 seconds in the area.

The difference between target hovers (68707) and target triggers (10399) is explained by the point-and-dwell mechanism and the placement of targets on the screen. Whenever the user begins pointing at a screen, they are likely to first hover over the button on the bottom edge of the information cube (close to the bottom of the screen). Similarly, users are likely to hover through multiple targets on their way to the final target, especially when moving to targets positioned close to the top edge. Moreover, some playful interaction has likely taken place, wherein users play with the cursors by moving them around the screen without triggering any targets.

With many deployments, it could also be beneficial to reflect upon what actions specifically should be considered interaction. For instance, one simple use case for the Information Wall is that a user on his/her way to lunch would stop by at the installation to check the day's menu as a form of final confirmation on whether or not (s)he wants to go to that particular restaurant. Since the current day's menu is shown to users by default when they activate the information cube by walking close enough to the display, no gestural interaction is required. Therefore, spatial interaction (walking into the scene to activate the cube) is enough to carry out the desired task. While we did not focus on such interaction scenarios in this article, it could certainly warrant further studies.

Movement and User Reactions

We defined passersby's entry and exit directions by calculating the angle of movement during the first and last 0.5 seconds that they were visible. Angles were divided into 90-degree cones, and hence we categorized users' entry and exit direction as left, right, front, or unknown (Figure 8). The direction was defined as unknown if the user was visible for less than 0.5 seconds, if there was backwards movement, or if there was no noticeable movement during that period. This happened when e.g. other people were preventing the sensor from seeing a passerby on entry, and when the passerby was recognized, (s)he had already stopped moving. Consequently, direction of entry could be defined for 63.9% of users; however, exit direction was defined for a significantly larger segment of passersby, 81.5%.

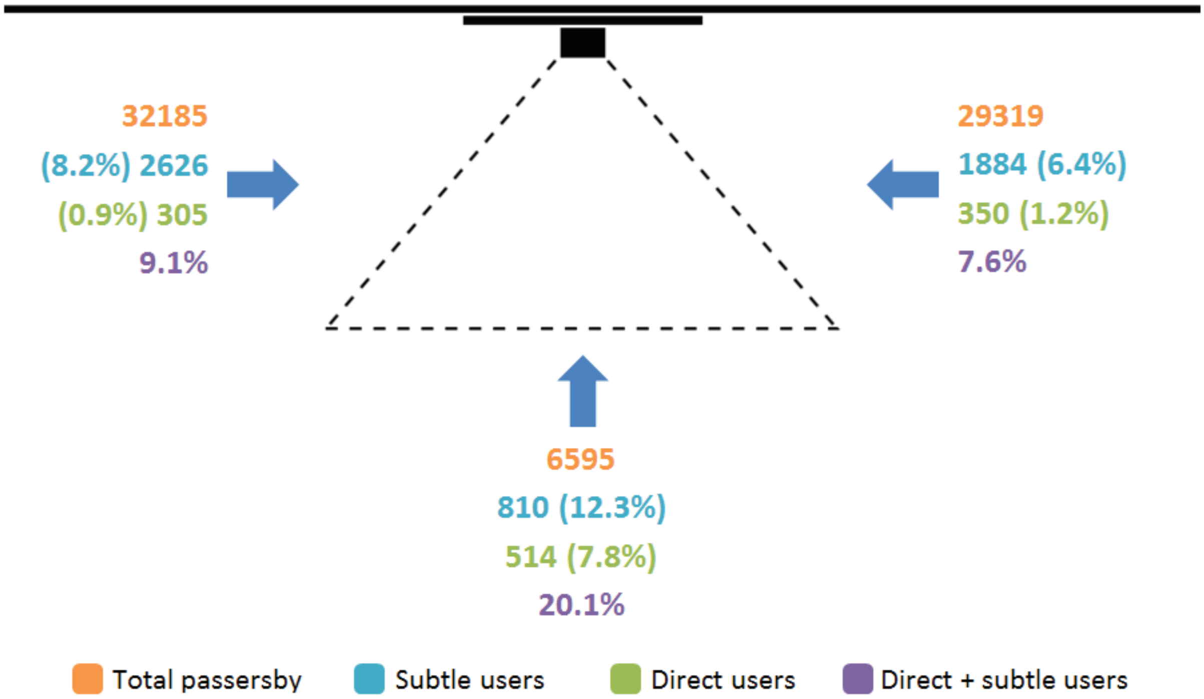


Figure 8. Users' entry directions.

Rates within which passersby converted into subtle and direct users for each direction are also presented in Figure 8. Users entering from the front were significantly more likely to become subtle (12.3%) and direct users (7.8%) than passersby entering from the left and right side (8.2%, 0.9% and 6.4%, 1.2% respectively). This might be due to the display being more visible from the front, and people coming from the corresponding direction could more easily observe the display and its possible users already from relatively far away.

We were interested in investigating the effects of our two-level reaction to passersby. We displayed a rectangular shape on the screen following a user when the user was positioned more than 2.8 meters away from the sensor (subtle reaction). For users positioned less than 2.8 meters away, the information cube was opened (direct reaction). For the analysis, it made sense to exclude users coming from the front or unknown direction, as users arriving from the front would always first trigger the subtle reaction on the display. Hence, we only included passersby who passed by the system sideways. A total of 37952 passersby passed by the system more than 2.8 meters away, of which 843 (2.2%) became subtle users and 189 (0.5%) became direct users. A total of 23552 passersby entered the area less than 2.8 meters away, of which 3667 (15.6%) became subtle users and 466 (2.0%) direct users.

Individual and Group Behavior

Of all passersby, the clear majority of 91306 (85.6%) people were lone passersby, i.e. no one else was seen during their presence. For direct users, 45.8% interacted without anyone else present, and 18.5% were accompanied by passive user(s). The remaining 35.7% were accompanied by subtle or direct users or a combination of user types (passive, subtle, direct).

Lone direct users were present for an average of 55 seconds and triggered an average of 5.8 targets. Direct users with company, however, were present for an average of 103 seconds and triggered 7.8 targets on average. The highest increase in both duration and target triggers was observed in large groups consisting of all three user types (153 seconds, 12.4 targets).

A total of 2552 passive users were present during direct interaction. The majority of these passive users, 2227 (81.7%), were standing far away enough so that a personal information cube was not opened for them.

Usage Over Time

Finally, we looked at how the amount of subtle and direct users develops over time. Figure 9 shows the percentage of subtle and direct users from the total number of passersby within each month. July 2013 (month 4) was left out as the system was not running at the time due to the holiday season.

It is by no means surprising for a public display system that the amount of users drops as time passes. In contrast, we were surprised by the very high usage percentage of the first months. In the first deployment month, 25.7% of all passersby were subtle users, and 4.0% were direct users. For the second month, the figures were 14.5% and 3.1%, and for the third month, 13.4% and 3.4%. This indicates that attention-wise our deployment location was successful and the system managed to make passersby interested enough to approach the display. However, the dramatic drop in usage rates can also suggest that the system failed to make a lasting impression, even with automatically updated content (daily lunch menus and latest news).

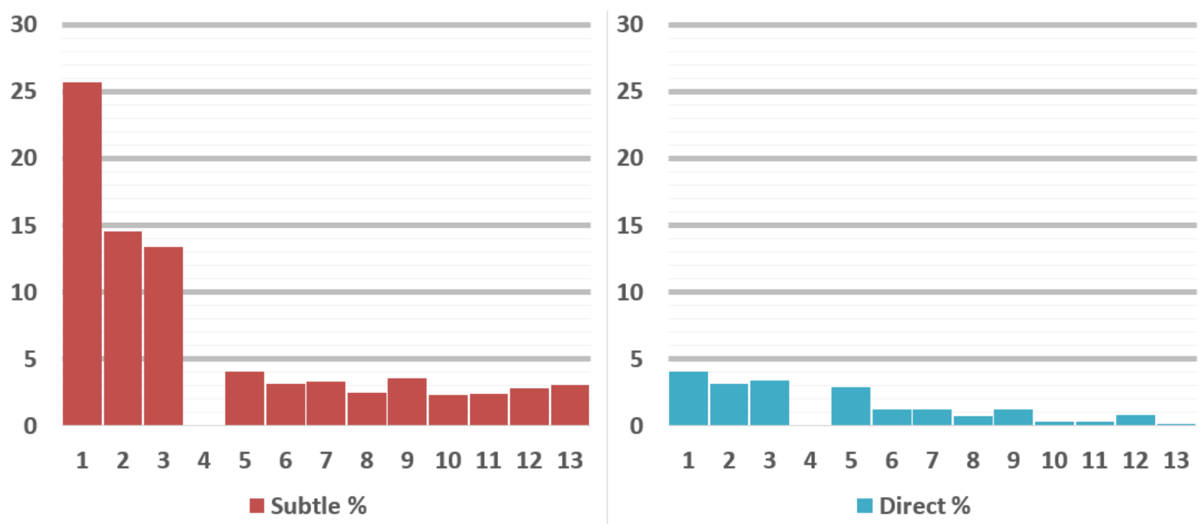


Figure 9. Percentages of subtle users (left) and direct users (right) within each month of deployment.

DISCUSSION

In this section, we first discuss the characteristics of evaluating large displays by semi-automatically analyzing the interaction and skeletal data gathered during a long-term deployment. Then, we discuss the benefits and challenges of the method as well as the interrelated roles of both automated logging and observations. Finally, we discuss our findings from applying our method to the Information Wall system.

Characteristics of Semi-Automated Evaluation

We identified four primary phases in semi-automated evaluation of public displays: data collection, preparation, feature extraction, and analysis. In practice, these phases can overlap somewhat. Moreover, it is certainly possible to further automate the process, and extract desired variables already during the runtime of the deployment. However, this would require that the full process is conducted by the same party from start to finish, which is not always the case.

The most defining phase of the semi-automated method is the *feature extraction* phase (Figure 5), primarily the aspect of variables and parameterization. For instance, in the case of the Information Wall, we wanted to classify users based on their level of interaction. For this, we needed to define parameters for classification, i.e., how exactly a passerby would be classified as a direct, subtle, or passive user. Similarly, we needed to decide how to define entry and exit directions, and when to ignore certain calculations if there was too little data on a user. The need to define parameters is not a drawback per se, but is simply something that researchers and practitioners need to pay attention to.

We argue that these parameters are system and context specific, and are also dependent on what exactly the researchers aim to investigate. It could also be argued that parameterization is present in every evaluation method as well, but in most other cases the process is more implicit. For instance, a researcher observing an installation on-site and counting how many users enter the space from left, right, or front, would similarly use some implicit factors to define the entry direction for each user, even though it would be obvious in most cases. The requirement to explicate the parameters may be beneficial for researchers to better understand the characteristics of the behaviors they are interested in.

One challenge that is inherent to the use of logged data, and indeed any research that seeks to make inferences based on such data, is choosing relevant properties of the phenomenon under study. Often it is possible to record much more data than is practical to store or analyze. In our case, for example, it would have been impractical to record all possible data properties exposed by the depth sensor. This should be considered in the first two phases of the process.

Benefits of Semi-automated Analysis Utilizing Interaction and Skeletal Data

The proposed semi-automatic process lends itself well to a number of applications. Based on our experiences in applying the method in the Information Wall study, we identified the following interrelated situations where it may provide benefits over other methods:

- When data sets need to be collected with minimal effort over the deployment itself.
- When large scale quantitative data collection is of interest.
- When it is needful to study long term usage patterns of the system.
- When the privacy of the users is of concern.
- When studying gesture-based interfaces and proxemic interactions.

One of the most popular methods for gathering data in field studies have been observations, where one or more researchers spend time near the display and observe its use and passerby behavior [3;24]. However, gathering extensive amounts of quantitative data via observations is both time consuming and limited to what a human observer is able to record at a time. For instance, in the case of multiple simultaneous users, multiple observes need to be employed; or an observer needs to divide their attention between the users or ignore the other users to properly focus on one user. With automatic logging of skeletal data, there are no such limitations to how much data can be logged at once, beyond the limitations of the measurement technology.

Our analysis on the collected data in this article was relatively simple, and it could be argued that similar findings could be easily reached with short-term observations. While this may likely be true in some cases, it is important to note that this may not always be the case due to a multitude of factors, such as the nature and purpose of the deployment space. For instance, there were an average of two to three groups per day interacting with the display, which totaled to around only four minutes of daily group use. On some days, there were no group users at all. Therefore, observers wanting to investigate group use could be on-site for hours, even days, without observing a single group session – and to make any reliable conclusions, a relatively large number of groups would have to be observed.

Maintaining a public display with automatic logging requires considerably less effort than conducting in situ observations, resulting in more quantitative data with less effort. During the deployment of the Information Wall, a typical week of running the system only consisted of starting up the system Monday morning, and turning it off Friday afternoon. Our analysis shows the benefits by collecting traces of more than 100,000 people, and also containing all the possible information we could get from both the system itself as well as the Kinect sensor. Other studies have also reached high numbers by utilizing a similar approach. For instance, Müller *et al.* [16] used a similar method for analyzing passersby in several different locations simultaneously, and received data of more than 30,000 passersby in a relatively short period of two weeks. However, in past research it seems that even if extensive depth data is gathered, it is not utilized to full extent. In many cases the data is analyzed manually, and is intended to support observations and interviews.

Furthermore, observation-based deployments are usually relatively short, while one of the most obvious benefits of logged skeletal data is being able to identify how interaction and reactions of passersby develops over time. For instance, our lightweight long-term analysis of the Information Wall revealed very high usage rates during the first three months of deployment, and a relatively rapid drop in the next months.

Another benefit is related to privacy. A semi-automated process utilizing skeletal data does not require recording of video or any kind of material from which users and passersby could be identified. For example, we recorded interactions from the display and anonymous skeletal data of passersby, which consisted of location data for the user's joints only.

Furthermore, we argue that gathering skeletal data is especially useful for studying gesture-based interfaces. First, no separate hardware or software is required as gesture-controlled systems already include motion detection sensors and the logic to interpret motion as interaction or non-interaction. Second, proxemic interactions [6] can be a major part of interaction in gestural interfaces. For instance, the Information Wall incorporated several ways of reacting to passersby based only on their location. Capturing detailed orientation information inherent to the skeletal motion data lends itself well to analyzing interaction proxemics.

Effective Collaboration between Semi-Automated Analysis and Observations

Some benefits of conducting observations are difficult or impossible to match with automated logging. However, as observation is primarily a qualitative method, and automated logging is quantitative, we argue that they are most effective when used together. In past research, logged data has been used in minor roles to support findings from other research methods. In this work, we promote more equal use of both, with the emphasis shifting based on the context of research.

A major benefit of conducting observations is their dynamic nature. For instance, researchers might observe a surprising incident with the display, and decide on the fly to shift their attention towards this phenomenon. Overall, observations are well suited for studying more complex behaviors. With automated logging, it is not often practical to capture the qualitative aspects of interaction, such as facial expressions or verbal comments, without compromising users' privacy. Moreover, researchers in the field can conduct interviews with users as opportunities arise. With automated logging, these benefits are difficult to match. However, as we have shown in this work, observations are resource-intensive. A practical and directly utilizable model would be to use logged data to gain general insight on the phenomena around, and the usage of, the deployed system, and use this information for more effective use of observational resources. For instance, logged data could easily show the time of day when groups of users are generally present, or when a certain interesting phenomenon usually happens. One interesting avenue for future research is to examine the integration of experience sampling into the public display to collect subjective feedback from users when in situ observation and interviewing is not possible.

In a similar manner, observations can also support automated logging. As we noted earlier in this article, observations played a role in the feature extraction phase of our semi-automated process. With observations, parameters for producing desired variables can be identified, e.g., to calculate movement patterns and to create rules for classifying passersby.

Current Limitations and Future Improvements

We experienced some practical and technical limitations during our year-long deployment that can be dealt with in the future, as discussed next.

One limitation is that there are few reliable ways to recognize returning users from new users without relying on techniques that may compromise privacy, such as face recognition. Hence, analyzing multiple interactions or follow-up actions [12] of a user is difficult. Additionally, users in groups often discuss the display while interacting, and recording audio from users in a public setting is both technically challenging as well as problematic in terms of privacy.

Many of the particular challenges in our example deployment were caused by the limitations of technology, which necessitates tradeoffs in the overall process. For example, the Kinect sensor also had trouble recognizing passersby who quickly moved past the display horizontally, i.e. those who were sideways to the display and walked fast. In these cases, we often caught just a glimpse of the passerby before they exited the area and we were not able to determine their walking direction in a trustworthy manner. Newer sensors, such as the Kinect 2, are faster and more reliable, and would likely solve this issue to great extent.

Another related issue was that we were unable to separate passersby who completely ignored the display from passersby who glanced at the display as they walked by. This was in part due to us only collecting simple joint and pointing data, but also due to the aforementioned technical issue of not recognizing passersby early enough. For this reason, we focused our analysis of the Information Wall on subtle and direct users, and passive users who accompanied them. However, this issue could be alleviated with a simple hardware upgrade and gathering head orientation data. This would allow us to e.g. identify an issue of display blindness [13] or interaction blindness [19] with the system.

Another challenge was that some passersby may stand or walk behind other passersby and thus may not be recognized right away or at all. This could be alleviated by setting up the sensor above the display instead of below it, giving it a better view of the space and the people. Such decisions should be informed by a thorough understanding of the measurement technology in use.

Finally, the space in which the Information Wall was deployed is very large, and the Kinect sensor could only cover a small sector of it. Hence, we were not able to analyze movement patterns of passersby on a larger scale, particularly the exact location users were typically coming from. Our approach would be more beneficial in a slightly more confined space, such as a small lobby or a crossroad of two hallways. In our case, relevant pathways such as doors outside and to an auditorium as well as a nearby cafeteria were too far from our system, so detailed analysis of e.g. movement patterns could not be achieved. Again, this issue could be somewhat alleviated with a hardware upgrade, as modern motion detection sensors offer a wider field of view, and thus can cover a larger space. Another option would be to utilize multiple sensors and combine the data streams during the analysis phase.

Usage Rates and User Behavior

The following sections discuss the specific findings of our analysis, with an intent to demonstrate the usefulness of the proposed approach in acquiring insights from large scale usage data. Overall, our analysis revealed that the usage rates of our installation are in line with other studies, although our data set of more than 100,000 passersby is considerably larger than data sets in other studies. We identified a substantial number of subtle users (5.9%), while the number of direct users was 1489 (1.4%). To compare, Müller *et al.* had a usage rate of 4% [16], however they did not distinguish between subtle and direct use, but counted all interactions towards this figure.

The large-scale and long-term data collection allowed us to identify an interesting difference between passersby entering the area from the front and passersby entering from either side. Users entering from the front were significantly more likely to become subtle or direct users (20.1%) than users entering from left or right (9.1% and 7.6%) (Figure 8). The difference is relatively logical in that people passing by the system sideways are more likely to be simply going from A to B and the most direct route is through the installation area, especially considering that the exit from and entry to the building was directly to the right of the installation, and the way to the main lobby directly to the left. On the other hand, people entering from the front had fewer reasons to pass by the installation space unless they were interested in the system. However, this finding could also indicate that people prefer to inspect a display from further away first before making the decision to interact. After all, the Information Wall was highly visible from all angles and from relatively far away. Users coming in from the front likely had time to observe it as they approached it, and possibly saw someone else interact with the display before.

We were surprised to see such a high number of targets triggers per user as well as relatively long interaction times. However, we were able to confirm that users in groups were more active than lone users by spending more time in front of the display and triggering more targets on the screen. It also seems that group size reflects the duration and the number of target triggers, as the most active users belonged to large groups. Ackad *et al.* had similar results, as around half of their users belonged to a group, and interacted longer and performed more gestures [1]. Müller *et al.* also reported similar results [16].

We argue that supporting simultaneous interaction is especially important when utilizing novel and expressive interaction methods such as gestures in public displays. Brignull and Rogers [4] note that people may choose not to interact with a public display due to fear of embarrassment, and we believe interacting with a friend may alleviate this issue. In addition, we were able to identify that while a large percentage of direct users interacted in a group, a large segment of those companions did not interact with the system at all. One possible explanation could be that people were not aware that the system supported simultaneous use. Hence, it could be worth considering that the display should attract passersby even if it is occupied by another user.

The relatively long duration of direct users (even those who triggered one or two targets) suggests that users were exploring the system for the first time. Thus, it could be that very few users returned to use the system. We could investigate this in more detail by e.g. analyzing how much time on average it took for users to trigger the first target, i.e. did they know what they were doing or were they simply exploring. On the other hand, average durations for passive (2.2 seconds) and subtle users (8.0 seconds) are logical, as it takes only a few seconds to walk past the system, and stopping for a while to e.g. wave one's hand should not take much longer either.

Reacting to Passersby in Display Design

Our large display application introduced a two-level reaction to passersby; a dynamic rectangular shape following far-away users (subtle reaction), and a personal information cube that was opened for passersby who were closer to the screen (direct reaction). We found that passersby to whom an information cube was displayed were significantly more likely to interact with the display. This suggests that the subtle reaction was not expressive enough to communicate interactivity and encourage exploration of the system; hence stronger reaction to passersby would be advised. Ojala *et al.* [20] call the issue interaction blindness, which refers to the passerby's inability to notice that a display is interactive. Müller *et al.* [17] studied this phenomenon in detail, and found that displaying a users' mirror image is more effective in conveying interactivity than e.g. silhouettes or abstract representations. On the other hand, the direct reaction seemed to work relatively well.

We observed a phenomenon similar to interaction blindness, which could be called *multi-user interaction blindness*. We found that passive users who accompanied direct users were mostly (81.7%) positioned further away and hence they did not activate an information cube on the screen. This could suggest that people were not aware that the system supported two simultaneous users, and simply stood back to observe the other user. Indeed, this feature was not explicitly communicated, and when one information cube was open, the system did not react to other passersby until they came close enough for another information cube to activate. Our finding suggests that displays supporting simultaneous use should react to and entice passersby even if another user is currently using the display.

CONCLUSION

In this work, we introduced and studied an approach to evaluating public displays by making use of automatically collected anonymous interaction and skeletal data, and analyzing the data programmatically. We deployed a gesture-controlled public information display in a university campus for one year and collected an extensive data set containing traces of more than 100,000 passersby. The data was analyzed to identify how passersby and users react to and interact with the public display. Our starting point was to investigate whether we could programmatically analyze the data and reach findings that we would have likely identified if we had been on-site observing users.

The main benefits of the approach include (1) automatic gathering of large data sets without considerable use of resources (2) privacy-preserving, semi-automated data analysis (3) analyzing the effects of long-term deployment. The approach is not without its limitations: the dynamic nature and interviewing opportunities offered by in situ observations are particularly difficult to match. However, we believe the benefits of the proposed method outweigh the drawbacks when the aim is to analyze public display interactions on a large scale.

To test our approach in a practical setting, we applied our process to the data captured of the Information Wall installation. Our most interesting findings were (1) long duration and high amount of target triggers for users, which could indicate that most users were first-time users exploring the system, and not many returned to use the system again, and (2) many users were accompanied by passive users who observed interaction from further away, which could suggest a case of multi-user interaction blindness.

A defining characteristic of our method is the parameterization of data variables during the analysis process, which is a key factor in producing meaningful, deployment-specific results. Successful parameterization requires knowledge of the system being evaluated, the space the system is deployed in, as well as identification of the factors involved in passerby behavior in relation to the display.

As long-term public display deployments are becoming more frequent, the need for improved evaluation methods is also emphasized. In past research, logged skeletal data has mainly been utilized to support findings from in situ observations, and has often been manually analyzed, resulting in a considerable amount of additional work. As we demonstrated in this work, logged data can be utilized more in-depth through a semi-automated process. We argue that our proposed process can act in a larger role, particularly for long-term deployments, and that observations and automated logging can support each other in a multitude of ways.

REFERENCES

[1] Ackad, C., Clayphan, A., Tomitsch, M., & Kay, J. (2015, September). An in-the-wild study of learning mid-air gestures to browse hierarchical information at a large interactive public display. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (pp. 1227-1238). ACM.

[2] Imeh Akpan, Paul Marshall, Jon Bird, and Daniel Harrison. 2013. Exploring the effects of space and place on engagement with an interactive installation. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13). ACM, New York, NY, USA, 2213-2222.

[3] Florian Alt, Stefan Schneegaß, Albrecht Schmidt, Jörg Müller, and Nemanja Memarovic. 2012. How to evaluate public displays. In Proceedings of the 2012 International Symposium on Pervasive Displays (PerDis '12). ACM, New York, NY, USA, Article 17.

[4] Brignull, H, and Rogers, Y. 2003. Enticing People to Interact with Large Public Displays in Public Spaces. In: Rauterberg, M and Menozzi, M and Wesson, J, (eds.) INTERACT 2003. IOS Press.

[5] Grace, K., Wasinger, R., Ackad, C., Collins, A., Dawson, O., Gluga, R., Kay, J., Tomitsch, M. Conveying Interactivity at an Interactive Public Information Display. In Proceedings of The International Symposium on Pervasive Displays (PerDis'13). ACM, New York, NY, USA.

[6] Greenberg, S., Marquardt, N., Ballendat, T., Diaz-Marino, R., & Wang, M. (2011). Proxemic interactions: the new ubicomp?. interactions, 18(1), 42-50.

[7] Hassenzahl, M., Burmester, M., Koller, F. (2014). AttrakDiffTM. Available at: http://www.attrakdiff.de.

[8] Hespanhol, L., Tomitsch, M., Grace, K., Collins, A., Kay, J. Investigating intuitiveness and effectiveness of gestures for free spatial interaction with large displays. In Proceedings of the 2012 International Symposium on Pervasive Displays. PerDis '12. ACM, New York, NY, Article 6.

[9] Marshall, P., Morris, R., Rogers, Y., Kreitmayer, S., Davies, M. Rethinking 'multi-user': an in-the-wild study of how groups approach a walk-up-and-use tabletop interface. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11). ACM, New York, NY, USA, 3033-3042.

[10] Mäkelä, V., Heimonen, T., Turunen, M. Magnetic Cursor: Improving Target Selection in Freehand Pointing Interfaces. In Proceedings of The International Symposium on Pervasive Displays (PerDis'14). ACM, New York, NY, USA.

[11] Mäkelä, V., Heimonen, T., Luhtala, M., & Turunen, M. (2014, November). Information wall: evaluation of a gesture-controlled public display. In Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia (pp. 228-231). ACM.

[12] Michelis, D., and Müller, J. 2011. The Audience Funnel: observations of Gesture Based Interaction With Multiple Large Displays in a City Center. Int. J. Hum.-Comput. Int. 27(6), 562-579.

[13] Müller, J., Wilmsmann, D., Exeler, J., Buzeck, M., Schmidt, A., Jay, T., & Krüger, A. (2009). Display blindness: The effect of expectations on attention towards digital signage. In Pervasive Computing (pp. 1-8). Springer Berlin Heidelberg.

[14] Jörg Müller, Florian Alt, Daniel Michelis, and Albrecht Schmidt. 2010. Requirements and design space for interactive public displays. In Proceedings of the international conference on Multimedia (MM '10). ACM, New York, NY, USA, 1285-1294.

[15] Müller, J., Bailly, G., Bossuyt, T., Hillgren, N. MirrorTouch: Combining Touch and Mid-air Gestures for Public Displays. In Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '14). ACM, New York, USA.

[16] Müller, J., Eberle, D., Tollmar, K. Communiplay: A Field Study of a Public Display Media Space. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14). ACM, New York, NY, USA, 1415-1424.

[17] Jörg Müller, Robert Walter, Gilles Bailly, Michael Nischt, and Florian Alt. 2012. Looking glass: a field study on noticing interactivity of a shop window. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12). ACM, New York, NY, USA, 297-306.

[18] O'Hara, K., Glancey, M. and Robertshaw, S. Collective play in an urban screen game. Proc. CSCW '08, ACM Press (2008), 67-75.

[19] Ojala, T., Kukka, H., Lindén, T., Heikkinen, T., Jurmu, M., Hosio, S., Kruger, F. UBI-Hotspot 1.0: Large-Scale Long-Term Deployment of Interactive Public Displays in a City Center. In Proceedings of the 2010 Fifth International Conference on Internet and Web Applications and Services, p.285-294, May 09-15, 2010.

[20] Ojala, T., Kostakos, V., Kukka, H., Heikkinen, T., Linden, T., Jurmu, M., ... & Zanni, D. (2012). Multipurpose interactive public displays in the wild: Three years later. Computer, (5), 42-49.

[21] Parra, G., Klerkx, J., Duval, E. Understanding Engagement with Interactive Public Displays: an Awareness Campaign in the Wild. In Proceedings of The International Symposium on Pervasive Displays (PerDis'14). ACM, New York, NY, USA, 180-185.

[22] Peltonen, P., Kurvinen, E., Salovaara, A., Jacucci, G., Ilmonen, T., Evans, J., Oulasvirta, A., Saarikko, P. It's Mine, Don't Touch!: interactions at a large multi-touch display in a city centre. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08). ACM, New York, NY, USA, 1285-1294.

[23] Perry, M., Beckett, S., O'Hara, K., Subramanian, S. WaveWindow: Public, Performative Gestural Interaction. In Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (ITS'10).

[24] Schmidt, C., Müller, J., & Bailly, G. (2013, April). Screenfinity: extending the perception area of content on very large public displays. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1719-1728). ACM.

[25] Tomitsch, M., Ackad, C., Dawson, O., Hespanhol, L., Kay, J. Who cares about the Content? An Analysis of Playful Behaviour at a Public Display. In Proceedings of The International Symposium on Pervasive Displays (PerDis'14). ACM, New York, NY, USA.

[26] Turunen, M., Hakulinen, J., Melto, A., Heimonen, T., Laivo, T., and Hella, J. 2009. SUXES – User Experience Evaluation Method for Spoken and Multimodal Interaction. In Proceedings of Interspeech 2009. ACM, New York, NY, USA, 109-112.

[27] Walter, R., Bailly, G., Müller, J. StrikeAPose: Revealing Mid-Air Gestures on Public Displays. In CHI'13, ACM (2013), 841–850.

[28] Walter, R., Bailly, G., Valkanova, N., Müller, J. Cuenesics: Using Mid-Air Gestures to Select Items on Interactive Public Displays. ACM MobileHCI'14, Toronto, Canada, September 2014.
[29] Williamson, J.R., Williamson, J. Analyzing Pedestrian Traffic Around Public Displays. In Proceedings of The International Symposium on Pervasive Displays (PerDis'14). ACM, New York, NY, USA.

BIOGRAPHIES

All authors are, or have been affiliated with the Tampere Unit for Computer-Human Interaction (TAUCHI) at the University of Tampere, Finland.

Ville Mäkelä is a Ph.D. candidate at TAUCHI. His research interests include cross-device interaction, mid-air gestures, public displays, augmented and virtual reality, and wearables. He received his M.Sc. in interactive technology from the University of Tampere in 2013. Contact him at ville.mi.makela@sis.uta.fi.

Tomi Heimonen, previously with TAUCHI, is currently an assistant professor at the Department of Computing and New Media Technologies at University of Wisconsin-Stevens Point. His research interests include pervasive computing and information visualization. He received his Ph.D. in interactive technology from the University of Tampere in 2012. Contact him at tomi.heimonen@uwsp.edu.

Markku Turunen is a professor of Interactive Technology at TAUCHI. His research interests include e.g. multimodal systems, novel user interfaces, and evaluation methods for mobile and pervasive computing applications. He received his Ph.D. in computer science from the University of Tampere in 2004. Contact him at markku.turunen@sis.uta.fi.

Paper III

Mäkelä, V., Heimonen, T., & Turunen, M. (2014). Magnetic Cursor: Improving Target Selection in Freehand Pointing Interfaces. In *Proceedings of The International Symposium on Pervasive Displays (PerDis '14)*, Sven Gehring (Ed.), 112–117. New York, NY, USA: ACM. doi:10.1145/2611009.2611025

Magnetic Cursor: Improving Target Selection in Freehand Pointing Interfaces

Ville Mäkelä, Tomi Heimonen, Markku Turunen
TAUCHI, School of Information Sciences, University of Tampere
Kanslerinrinne 1, FI-33014 University of Tampere, Finland
(+358) 40 839 4038
{ville.mi.makela, tomi.heimonen, markku.turunen}@sis.uta.fi

ABSTRACT
We present the magnetic cursor, a technique that aims to make distant freehand interaction with targets easier in large-screen public display applications. The magnetic cursor automatically warps to a selectable object that is close by, moves slower while on the object, and shows the relative cursor location visually to the user. Two designs of the magnetic cursor were compared to the snap-to-target technique in a 19 participant user study. Results indicate that the magnetic cursor design with weaker magnetism effect outperforms the other techniques in terms of target selection efficiency. Subjective feedback indicates that snap-to-target and the magnetic cursor design with weaker magnetism effect meet the participants' expectations for freehand pointing and are preferred to unassisted pointing and the magnetic cursor with stronger magnetism. Our findings suggest that the visual feedback of cursor location and short, static activation threshold for the magnetism effect can help users maintain the cursor within the active motor space of a target, especially when several selectable targets are situated in close proximity.

Categories and Subject Descriptors
H.5.2 **[User Interfaces]**: Interaction Styles; User-centered Design; I.3.6 **[Methodology and Techniques]**: Interaction Techniques

General Terms
Design, Experimentation, Human Factors.

Keywords
Freehand pointing, public display applications, target selection, user experience.

1. INTRODUCTION
Despite technological advancements and the unique potential benefits that gesture interfaces can offer, it is still uncommon to encounter gesture-controlled systems in public spaces, as many public display installations are either non-interactive or rely on touch or physical keys for interaction. In addition, real world studies of gesture-based interaction are still relatively rare in the context of public display systems, as many of the in-the-wild studies tend to focus on other aspects, such as the use (or non-use) of such systems. Also, only relatively few studies (e.g., [2, 7, 13, 14]) have investigated freehand pointing and the challenges associated with selecting on-screen objects.

PerDis '14, June 03 - 04 2014, Copenhagen, Denmark

http://dx.doi.org/10.1145/2611009.2611025

The current research on target selection assistance is motivated by our initial usability findings of the use of Information Wall, a public display application that offers easy access through a freehand point-and-gesture interface to simple information like the latest news and events in nearby areas at the university campus. In freehand pointing systems, a simple pointing task requires focused hand-eye coordination and long term pointing based interaction is likely to result in physical fatigue faster than when using e.g. a mouse or handheld controller. These characteristics of freehand pointing, along with jitter and noise inherent in sensor-based detection, make it more difficult to hold the cursor steady than comparable methods utilizing physical input devices. Accurate freehand pointing can be especially challenging when multiple targets are in close proximity. Making the selection process as easy and intuitive as possible therefore aims to help reduce the cognitive and physical load of the users.

Our approach was to make target selection easier utilizing a novel assisting technique that is designed to work together with dwelling selection. This paper presents the results of an experiment where the two designs of the technique were compared to the snap-to-target technique, which has been previously proposed as an assisting method for target selection.

The rest of the paper is organized as follows. We will first introduce the context of our research, the Information Wall public display application. We then review related work on assisting methods for target selection in cursor-based pointing. Next, we introduce the Magnetic Cursor technique and describe the experiment conducted to study its effectiveness. Following, we report the results and discuss their implications for the design of targeting assistance techniques for freehand pointing.

2. INFORMATION WALL
In order to study gesture-based interaction in a public setting, we have developed a public display application called the Information Wall. Interaction with the Information Wall takes place via an on-screen cursor that moves according to where the user is pointing on the physical screen (Figure 1). The system utilizes the Microsoft Kinect sensor for detecting the users' hand coordinates, and maps these to the onscreen cursor location. Users can navigate the content with simple gestures, such as swiping to the cardinal directions to rotate the on-screen information cube in order to reveal new content. The rotation is triggered by selecting an edge of the cube, which involves hovering, or dwelling, the cursor over an active interface element for a short period of time. The same selection paradigm also applies to other objects, such as

buttons and list items. Dwelling was selected as the selection method because not only is it easy to provide immediate visual feedback but it also does not require any additional gestures, and has been found to be more intuitive than other selection methods such as grasping or gestures [7].

Figure 1. The Information Wall.

3. RELATED WORK

Several different targeting assistance methods have been proposed in previous research, although very few of them have been developed or evaluated specifically with distant freehand pointing in mind. These techniques can be broadly divided into ones that reduce the distance from the cursor to the target, either by moving the cursor closer to the target or the target closer to the cursor, or by increasing the size of the cursor or target [11].

Kabbash and Buxton [8] presented the area cursor, which is a static rectangular area, and to interact with a target the cursor only needs to touch it, not completely contain it. The area cursor is problematic with small and dense targets, as several targets may touch the cursor simultaneously. Worden et al [15] improved the area cursor by adding a crosshair to the center of the area. When several targets are contained within the cursor, the one closest to the center is selected. Grossman and Balakrishnan [6] developed the bubble cursor where the ellipse-shaped cursor dynamically resizes itself based on surrounding targets. Resizing is made so that the cursor always touches the closest target, but – thanks to its shape - never touches multiple targets at the same time.

Several targeting assistance approaches are based on expanding the target size in motor space, typically by manipulating the control-display (CD) gain. In semantic pointing [4] the CD gain is adapted according to cursor distance from nearby objects. Objects can have a unique size in motor space as well, and so it would be possible to have for example the most commonly used buttons larger than others in motor space. König et al [9] presented the adaptive pointing technique, which smoothly adjusts the CD gain of the cursor based on the speed and direction of movement, so that pointing appears more precise while still maintaining the feel of absolute pointing, e.g. the cursor appears where the input device is pointing, which makes the technique possibly suitable for distant freehand pointing, too. Ahlström et al [1] presented the force field technique, which works by creating an area around a target, inside which the CD gain for the cursor is lowered whenever the cursor moves towards the target (force point), making the cursor move faster than normal. While moving away from the target, the CD gain is increased again until it reaches the default value. In essence, the user is supposed to feel that the cursor is "attracted" to the target.

With sticky targets (or sticky icons) [1, 5, 15], the cursor moves slower than normal when it is on top of an object. In traditional desktop environments sticky targets have been found to be efficient in simple pointing tasks, but in real-world situations sticky icons might increase acquisition times as users would need to pass through several icons that make the cursor slower to get to the desired target. Some methods combine both at distance and while on the object. For example, Bateman et al [3] proposed a technique called target gravity, which makes the cursor attracted to objects from a distance and makes the cursor move slower while on an object. All gravity-enabled objects affect the cursor's location at the same time.

Relatively few studies have addressed the effectiveness of targeting assistance in non-desktop environments. Bateman et al [3] compared several different techniques with the Nintendo Wiimote as the input device. The target gravity technique was found to be the fastest and most preferred by users. Parker et al [12] compared different targeting assistance methods for stylus-based pointing in tabletop interaction. Their first proposed technique was expand-cursor, in which a circular area surrounds the cursor whenever a selectable object is close enough. The circle grows bigger as it nears an object and shrinks and vanishes when it moves away from the object. The object can be selected when the circle overlaps it, similar to the area cursor. Expand-target works the other way around – the object itself grows bigger when the cursor is moving closer, and shrinks back to its original size when the cursor moves away. In snap-to-target, the cursor "snaps" to the center of an object when it is close by. The cursor moves out of the object when the "real" cursor position moves further away. Snap-to-target was found to be the most efficient, accurate and subjectively preferred. All of the techniques proposed by Parker et al activate when the cursor has travelled 90% of the distance between the starting point and the target.

4. MAGNETIC CURSOR

Based on existing research, we recognized attributes that we wanted to have in a technique that assists in selecting targets in gesture interfaces. First, gesture-based interaction especially in public spaces is a relatively novel domain to most users. For this reason, the cursor should appear familiar from other environments, and thus its visual attributes should not be significantly altered during runtime. Second, distant freehand pointing is an absolute pointing technique, which means that targeting assistance methods that significantly alter the behavior or location of the cursor are not preferable. However, it has been found that small position changes may improve targeting time without being perceptible to the user [10]. Third, it is important to note that in most other types of input devices it is a trivial task to keep the cursor stationary. To achieve the same in distant freehand pointing, users have to focus to maintain their arm stationary. Also, typically interaction with targets happens by gesturing with the same hand that is used for pointing, like forming a fist or making a pushing motion. While performing the gesture, the cursor can easily move around unintentionally. Thus, the assisting technique's task in freehand pointing interfaces is twofold: make *acquiring* the desired target faster and easier, but also make *staying* on the target easier.

As a solution, we developed the Magnetic Cursor technique, which combines the snap-to-target and sticky icons approaches. To achieve the basic requirements outline above, the cursor has two functions; when it moves close enough to a selectable target, it automatically jumps onto that target, and while it is on the target, it moves slower than normal, requiring more movement to jump out of the target. When the cursor jumps onto an object, its position is calculated relative to the actual point where the user is pointing, i.e., pointing at the very edge of the area where magnetism is in effect will move the cursor to the corresponding edge of the object. The dynamic positioning tells the user whether one is actually pointing at the desired target or if one is barely inside the effective area.

The cursor's position is calculated as

$$c \pm a * r / m$$

where c is the (x,y) center of the target, a is the distance between the target center and the point where the user is actually pointing, r is the radius of the target, and m is the radius of the area where magnetism is in effect, calculated from the center of the target. In the case of the cursor being inside several magnetic areas, the closest object is always chosen.

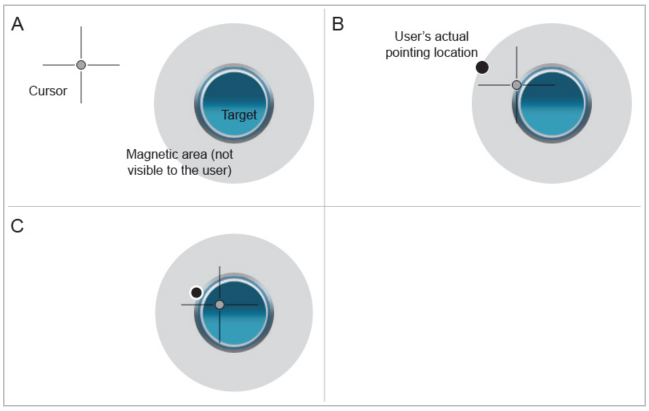


Figure 2. The Magnetic Cursor. A: While outside magnetic areas, the cursor follows the user's pointing location. B: At the edge of a magnetic area, the cursor warps onto the corresponding edge of the target. C: When pointing closer to the target, the cursor moves closer to the target relative to the radius of the magnetic area and the distance between the actual pointing location and the center of the target.

The magnetic cursor's behavior is demonstrated in Figure 2. In Figure 2C, the actual pointing location has travelled 50% of the distance between the edge of the magnetic area and the center of the target, and thus the visual cursor is positioned halfway between the center and the edge of the target.

4.1 Pilot study

A pilot study with seven participants was conducted to study the feasibility of the magnetic cursor technique. Our research questions were: (1) How does the magnetic cursor compare to the normal, unaltered cursor? (2) Does the magnetic cursor increase error rate? (3) What is a suitable activation threshold for the magnetic cursor? The *threshold* refers to the distance from which the magnetic cursor activates and moves onto a target. Thresholds of 100 and 200 pixels were evaluated, while the target radius was a static 100 pixels. In the 2D target acquisition experiment, the participants' task was to select the indicated button while balancing speed and accuracy. An early version of the Information Wall system was used as the interface.

The results showed that both magnetic cursor designs were almost twice as fast as the normal cursor, with the stronger (200-pixel threshold) version being the fastest. Subjectively, 5 out of 7 participants chose the 100-pixel version to be more pleasant to use. The error rates were higher with the magnetic cursors compared to the normal cursor (0.5%). At 2.4% and 3.3%, respectively, they are in line with error rates reported in previous research. Given the low absolute level of errors across conditions, the magnetic cursor can be considered more efficient than a normal cursor in practical terms.

5. EVALUATION

The objective of our user study was to examine both the objective (pointing performance) and subjective (user experience) properties of the proposed Magnetic Cursor technique in comparison to another similar technique, snap-to-target [12]. As discussed earlier, we wanted a cursor that does not visually alter itself or the objects, but instead changes the way it moves based on its and the objects' locations. Also, in the study by Parker et al [12], snap-to-target not only emerged as the most efficient and also the most preferred technique, but was also evaluated on a pen-based distant pointing system with similar characteristics to a freehand pointing interface. The target gravity technique evaluated by Bateman et al [3] shared many of the features, however it was concluded to be too subtle for the experiment as the cursor still requires movement towards the target after crossing the gravity threshold. In addition, given the inconclusive results on the activation threshold in the pilot study, we decided to include both versions of the magnetic cursor in the user study. Considering the results from previous studies and the pilot, the normal cursor was included in the study to serve as an introduction to freehand pointing and as a warm-up session.

Given the characteristics of the assisting techniques, our main hypothesis was that, owing to its visual feedback of pointing location, with magnetic cursor it would be easier for users to maintain focus on the selected element because they get visual feedback of the relative pointing location compared to techniques that do not provide such feedback, such as unassisted pointing or snap-to-target. Further, we were interested in examining the tradeoffs between ease of selection and susceptibility to erroneous selections in the presence of multiple targets. Finally, we were interested in investigating whether the differences in the techniques would be perceptible enough to users to result in differences in subjective feedback.

5.1 Participants

A total of 19 participants took part in the study, 3 women and 16 men. Their age varied from 19 years to 57 years (mean = 24). Most participants had some experience with gesture-based systems, all from the gaming domain using the Nintendo Wii, Microsoft Kinect, or PlayStation Move. Participants were recruited from a basic first-year interactive technology course held at a local university. Participation in the user study counted towards the completion of their course credit.

5.2 Apparatus

The study was carried out in a laboratory setting in October 2013. The interface was displayed on a 1920 x 1080 full HD projection screen. Participants were instructed to stand 2.4 meters away from the screen center.

An experiment application utilizing Microsoft Kinect as the input device was implemented specifically for the study. The default smoothing algorithm of the Kinect SDK was used to filter out minor input jitter. Calculation of the pointing coordinates is based on the Physical Interaction Zone data provided by the SDK.

Snap-to-target was implemented the same way as in the study by Parker et al. Thus, with levels of 400 and 800 pixels we used for distance, snapping would activate when the cursor was either 40 or 80 pixels away from the edge of the target. Similarly to the magnetic cursor, if multiple targets were inside the snapping threshold, the closest target was chosen.

5.3 Task

We used a 2D target acquisition task, where participants pointed at and selected targets positioned around the screen. One highlighted button would appear at a time, surrounded by four, visually different distractors (Figure 3). Distractor targets were added around the target button to investigate the performance effect of different magnetic cursor thresholds on selection time in scenarios where multiple targets appear within the activation threshold. Accidentally selecting a distractor was counted as an error and task progress required the correct button to be selected.

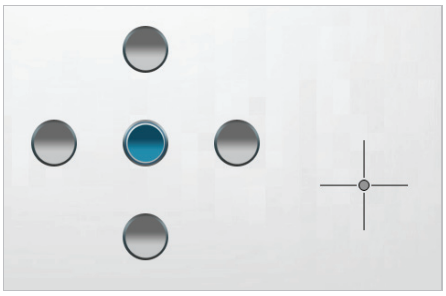


Figure 3. Screenshot of the application.

The buttons were selected using dwelling with the same visual feedback we used in the Information Wall. Dwell time for triggering a selection was set to 1.2 seconds for every condition after a dwell time of 1.5 seconds used in the pilot was reported to have been too slow by the participants.

5.4 Design

The user study was organized as a repeated measures within-subjects experiment with factors:

- **Technique**. Three levels: low-power magnetic cursor (100-pixel activation threshold, referred to as MC1), high-power magnetic cursor (200 pixels, referred to as MC2), and snap-to-target.
- **Distance**. Distance to the target from the previous target with two levels: 400 pixels and 800 pixels
- **Target size**. Size of the target and distractor buttons with two levels: 50-pixel and 100-pixel diameter.
- **Density**. Distance between the target and the distractors with two levels: 50 and 100 pixels.

The participants completed six target selections per each factor level combination, resulting in 2736 total selections.

The dependent variable was target selection time, which was decomposed in the analysis phase into *time to land on target* (first time the user brought the cursor on top of the correct target) and *target selection time* (including dwelling and possible reacquisition of target). In addition, the number of *hovers on target* (i.e., leaving target area and reacquiring it) and *error rate* was calculated for each technique.

5.5 Procedure

The participants started by reading a short introduction and by filling out a background questionnaire. All participants started with the normal cursor to acquaint themselves with the pointing system. The order of the three assisting techniques was counterbalanced and the participants were randomly assigned to one of the counterbalanced orders. Completing each pointing technique block was divided into two sessions, between which participants were able to rest if they desired. Each session contained 25 selections, the first of which always started from the center of the screen and was excluded in the analysis of the study. Values for distance, target size and density were randomized so that all eight possible combinations would appear six times for each technique.

Subjective feedback was collected using 11 questions related to the user of the pointing technique (Table 1). Participants answered the questions four times, once after each technique. Each question was answered on a 7-point bipolar scale.

Table 1. User experience statements.

Q1	Gesture control works very roughly/smoothly
Q2	Pointing requires me to focus too little/much
Q3	Pointing requires physical effort too little/too much
Q4	Pointing accurately is easy/difficult
Q5	Selecting targets by pointing is too fast/too slow
Q6	My arm gets tired when pointing not at all/a lot
Q7	Pointing feels very uncomfortable/ comfortable
Q8	Altogether pointing is very difficult/easy
Q9	I control pointing very poorly/well
Q10	The process of selecting the desired target is slow/fast
Q11	When the cursor is on a target, the target is selected too fast/slowly

Participants were not informed of the actual behavior of the techniques or of the differences between them. This is comparable to a realistic setting – a first-time user of a freehand pointing system will not be aware of the possible assisting techniques. When all the selection sets were completed, participants filled out the last questionnaire comparing the techniques.

6. RESULTS

Outliers were removed before analyzing the results. An entry was considered to be an outlier if total selection time (acquisition time included) differed more than 1.5 times standard deviation from the average selection time. Outliers formed around 9% of the data.

6.1 Time to land on target

A significant main effect of pointing technique was not observed for target acquisition time whereas significant effects were found for pointing distance, target size and target density. This is intuitively clear, given that the assisting techniques operate in close proximity to the targets and the bulk of the acquisition time is made up by the physical pointing motion from the previous to the current target. We did find a three-way interaction pointing technique * distance * density and four-way interaction pointing technique * distance * target size * density. However, interpreting such complex interactions is conceptually challenging within the context of this experiment. Given the experimental setup, the high effect size for pointing distance (partial η^2 = 0.889), and low effect size of the interactions (partial η^2 = 0.293 and 0.161, respectively), it appears that the main contributor to target acquisition time is the distance rather than differences in pointing behavior between the techniques.

6.2 Target selection time

Target selection time was calculated from the time when the user first acquired the target to the time when the target was selected. Repeated measures ANOVA showed a significant main effect of pointing technique on target selection time ($F_{2,36}$ = 11.445, p < 0.001), with MC1 performing significantly faster (1.6 seconds) than MC2 (1.7) and snap-to-target (1.8). Unsurprisingly, also target size and density had a significant effect on selection time. The selection of larger targets was faster than smaller targets and similarly less densely packed targets were faster to select than more densely packed targets.

6.3 Target hovers

Repeated measures ANOVA showed a significant interaction between pointing technique and target density ($F_{2,36}$ = 3.644, p < 0.05), as well as significant main effects for pointing technique, target size, and target density. As Figure 4 shows, when distractors were in close proximity to the target (50 pixels), MC1 had significantly fewer target hovers on average (1.6) than the other techniques (1.8 for MC2 and 1.9 for snap-to-target). At higher distance (100 pixels) the difference was not significant.

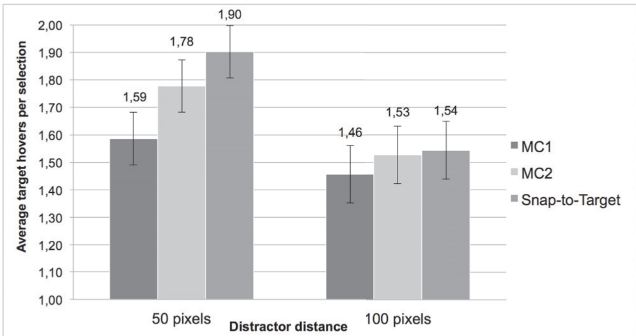


Figure 4. Average target hovers per selection.

6.4 Error rate

According to repeated measures ANOVA, there was also a significant effect of pointing technique on error rate ($F_{2,36}$ = 14.043, p < 0.001), whereby MC2 had a higher error rate (8.6%) than MC1 (2.4%) and snap-to-target (3.2%). Also, an interaction between target size and density was found. Selecting tightly packed smaller targets was significantly more challenging than selecting larger targets, but when target density is lower, the effect of target size effectively disappears. Although the normal cursor condition was not a part of this comparison, it should be noted that similarly to the pilot study, its use was in practice error free.

6.5 User experience

The user experience responses did not show many significant differences between the pointing techniques (Figure 5) based on pairwise comparisons. Normal cursor ratings are shown as a reference. MC2 was found to require more mental effort to operate (Q2) and less controllable (Q9) than its low-power counterpart. Snap-to-target and MC1 were also perceived to be faster to use for selecting the desired target (Q10). Our participants' subjective preferences are in line with their experience ratings. Out of the 19 participants, nine participants chose MC1 as the most pleasant technique, and nine chose the snap-to-target technique. Only one participant preferred MC2, while the normal cursor was not preferred by anyone. Conversely, eight participants chose both the normal cursor and MC2 as the least pleasant technique. The snap-to-target was chosen least favorite by three participants. MC1 did not receive any votes.

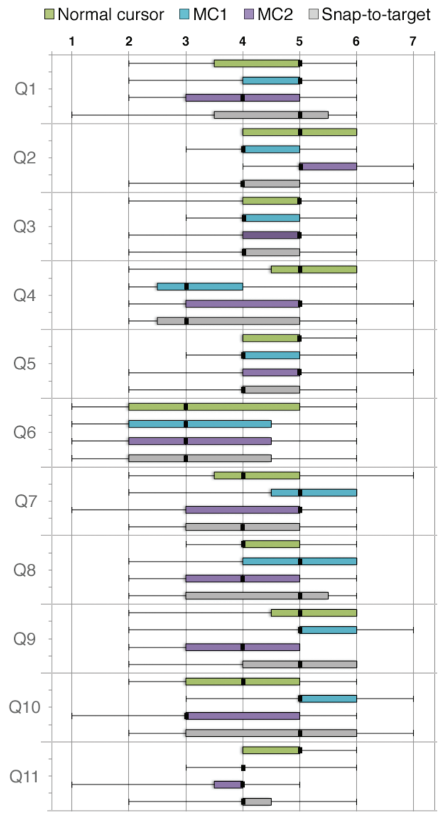


Figure 5. User experience ratings as boxplots (showing median and interquartile range) for each pointing technique.

7. DISCUSSION

Our research questions were concerned with how well the assisting techniques can help users maintain selection on the target during dwelling, what the tradeoffs are between activation threshold and errors in the presence of multiple targets, and how the participants subjectively perceive the different techniques. The results show that the low-power magnetic cursor design provided the best pointing performance in terms of target selection time. This benefit derives primarily from its ability to help the users keep the selection on target in densely packed target arrangements, where it outperformed the other techniques. This can be explained by the visual feedback of relative cursor position within the target, which we believe helped correct the pointing closer to the center to maintain position on the target.

The poor performance of high-powered magnetic cursor, both objectively and in subjective feedback, can be explained by the relationship between the computer cursor location and the strength of the magnetism effect. As the actual pointing location is further from the computed location, inaccuracies with pointing can result in the cursor warping to the wrong target more easily than with the low-power cursor. Participants' difficulties in controlling the cursor at high degree of magnetism are especially evident in the inflated error rate for high-power magnetic cursor.

This study evaluated assisting techniques for freehand pointing interfaces by investigating the tradeoffs between two design variables – visual feedback of cursor location and activation threshold of the warping effect. The magnetic cursor designs utilized a fixed activation threshold whereas the snap-to-target technique was dependent on the pointing distance. Similarly, the magnetic cursor designs showed the relative position of the cursor within the target's expanded motor space whereas snap-to-target placed it into the middle of the target irrespective of location in motor space. Our results suggest that while the visual feedback appears to help the user maintain the cursor on the target, it also needs to be coupled with a relatively low, static activation threshold. While this makes pointing less effective at a distance, it also reduces the chance of sliding off target during dwelling.

The use of dwelling, while practical for novice users, is fairly inefficient. In future studies we are planning to study how different simple gestures, just as grab or pinch, could be used to speed up target selection in conjunction with cursor warping. Increasing the motor space of targets could be beneficial in alleviating targeting issues that can arise as a result of hand gestures affecting the pointing coordinates. Furthermore, it might be beneficial to take the magnetic cursor concept further; instead of the cursor having a predefined magnetic area, targets could have their own magnetic areas with varying sizes and shapes.

8. CONCLUSION

We investigated different means for improving target interaction for freehand pointing in large display environments. We proposed the magnetic cursor technique, which automatically warps the cursor onto a target, makes the cursor move slower on target, and visually shows its relative position in the motor space. We evaluated the performance and user experience of two versions of the magnetic cursor and compared them to the snap-to-target technique. The low-power version of the magnetic cursor was found to be the fastest of all techniques in target selection; however differences between the assisting techniques were mostly small in practical terms. In terms of user experience, all techniques were mostly equal, although slight preference towards the low-power magnetic cursor was found. However, it is clear that too aggressive an activation threshold can negatively affect error rate and user experience, as was observed with the high-power magnetic cursor.

9. ACKNOWLEDGMENTS

This work was funded by Tekes, the Finnish Funding Agency for Innovation, through the RYM Oy Indoor Environment program.

10. REFERENCES

[1] Ahlström, D., Hitz, M., and Leitner, G. 2006. An evaluation of sticky and force enhanced targets in multi target situations. In *Proceedings of the 4th Nordic Conference on Human-Computer Interaction.* NordiCHI '06. ACM, New York, NY, 58-67.

[2] Bailly, G., Walter, R., Müller, J., Ning, T., and Lecolinet, E. 2011. Comparing free hand menu techniques for distant displays using linear, marking and finger-count menus. In *Proceedings of 13th IFIP TC 13 International Conference on Human-Computer Interaction.* INTERACT'11. Springer, Berlin Heidelberg, 248-262.

[3] Bateman, S., Mandryk, R., Gutwin, C., and Xiao, R. 2013. Analysis and comparison of target assistance techniques for relative ray-cast pointing. *Int. J. Hum.-Comput. Stud.* 71, 5, 511-532.

[4] Blanch, R., Guiard, Y., and Beaudouin-Lafon, M. 2004. Semantic pointing: improving target acquisition with control-display ratio adaption. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* CHI '04. ACM, New York, NY, 519–526.

[5] Cockburn, A. and Firth, A. 2003. Improving the acquisition of small targets. In *Proceedings of HCI 2003*. Springer, London, 181-196.

[6] Grossman, T. and Balakrishnan, R. 2005. The bubble cursor: enhancing target acquisition by dynamic resizing of the cursor's activation area. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* CHI '05. ACM, New York, NY, 281–290.

[7] Hespanhol, L., Tomitsch, M., Grace, K., Collins, A., and Kay, J. 2012. Investigating intuitiveness and effectiveness of gestures for free spatial interaction with large displays. In *Proceedings of the 2012 International Symposium on Pervasive Displays*. PerDis '12. ACM, New York, NY, Article 6.

[8] Kabbash, P. and Buxton, W. 1995. The "Prince" technique: Fitts' law and selection using area cursors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* CHI '95. ACM, New York, NY, 273-279.

[9] König, W., Gerken, J., Dierdorf, S., and Reiterer, H. 2009. Adaptive pointing - design and evaluation of a precision enhancing technique for absolute pointing devices. In *Proceedings of 12th IFIP TC 13 International Conference on Human-Computer Interaction.* INTERACT '09. Springer, Berlin Heidelberg, 658-671.

[10] Mandryk, R.L. and Gutwin, C. 2008. Perceptibility and utility of sticky targets. In *Proceedings of Graphics Interface 2008*. GI '08. Canadian Information Processing Soc., 65-72.

[11] McGuffin, M.J. and Balakrishnan, R. 2005. Fitts' law and expanding targets: Experimental studies and designs for user interfaces. *ACM Trans. Comput.-Hum. Interact.* 12, 4, 388-422.

[12] Parker, K., Nunes, M., Mandryk, R., and Inkpen, K. 2005. TractorBeam selection aids: improving target acquisition for pointing input on tabletop displays. In *Proceedings of the 2005 IFIP TC13 International Conference on Human-Computer Interaction*. INTERACT'05. Springer, Berlin Heidelberg, 80-93.

[13] Polacek, O., Klima, M., Sporka, A.J., Zak, P., Hradis, M., Zemcik, P., and Prochazka, V. 2012. A comparative study on distant free-hand pointing. In *Proceedings of the 10th European conference on Interactive tv and video.* EuroiTV '12. ACM, New York, NY, 139-142.

[14] Schick, A., van de Camp, F., Ijsselmuiden, J., and Stiefelhagen, R. 2009. Extending touch: towards interaction with large-scale surfaces. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces.* ITS '09. ACM, New York, NY, 117-124.

[15] Worden, A., Walker, N., Bharat, K., and Hudson S. 1997. Making computers easier for older adults to use: area cursors and sticky icons. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* CHI '97. ACM, New York, NY, 266-271.

Paper IV

Mäkelä, V., Heimonen, T., Luhtala, M., & Turunen, M. (2014). Information wall: evaluation of a gesture-controlled public display. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia (MUM '14, Melbourne, Australia)*, 228–231. New York, NY, USA: ACM. doi:10.1145/2677972.2677998

Information Wall: Evaluation of a Gesture-Controlled Public Display

Ville Mäkelä, Tomi Heimonen, Matti Luhtala, Markku Turunen
TAUCHI, School of Information Sciences, University of Tampere
Kanslerinrinne 1, FI-33014 University of Tampere, Finland
(+358) 040 839 4038
{ville.mi.makela, tomi.heimonen, matti.luhtala, markku.turunen}@sis.uta.fi

ABSTRACT
Public displays that allow users to interact with them through mid-air gestures are still relatively rare, as many applications rely on touch-based interaction. This paper introduces Information Wall, a gesture-controlled public information display that provides multi-user access to contextually relevant local information using remote pointing and mid-air gestures. The application has been studied in two settings: a lab-based user study and several short-term deployments. Based on our results, we present practical guidelines for gesture-controlled public display design.

Categories and Subject Descriptors
H.5.2 **[Information Interfaces and Presentations]**: User interfaces – *evaluation/methodology, interaction styles, graphical user interfaces*.

General Terms
Design, Experimentation, Human Factors.

Keywords
Public displays, gestures, mid-air pointing, pervasive displays, user study.

1. INTRODUCTION
Interactive public information displays have been deployed in malls, train stations and similar spaces over the recent years, and regularly rely on touch input. While touch works well for certain environments, it may create a situation that counters the display's purpose. Users interacting with these displays in very close proximity can block the view from passers-by, which leads to a situation where public displays are not all that public anymore.

Gestures within public information displays (PIDs) remain rare. Instead, displays with motion detection capabilities are often produced in the form of a playful application, such as a simple physics-based game [6]. In this paper we present Information Wall, a gesture-controlled public information display that provides multi-user access to contextually relevant local information. Interaction happens via mid-air pointing and an on-screen cursor, through which targets on the screen are triggered. In our studies of Information Wall we focused on: a) measuring the subjective experience of users and defining the level of quality of the system, and b) producing design guidelines for gesture-controlled public displays.

Next, will first present related work, and our Information Wall prototype. Then, we present the setup and results of the user study and public sessions. Finally, we discuss the results and present design guidelines and suggestions for future work.

MUM '14, November 25 - 28 2014, Melbourne, VIC, Australia
Copyright 2014 ACM 978-1-4503-3304-7/14/11…$15.00
http://dx.doi.org/10.1145/2677972.2677998

2. RELATED WORK
In a public setting most users are likely to be first time users, and people tend to interact with the display only briefly [11]. Thus, it has been argued that immediate usability is more important for a public display than performance [12]. In addition, people trying out public displays tend to be impatient: interaction usually ends if users do not succeed with what they are trying to achieve [3].

Müller *et al*. [5] present three issues that public displays specifically need to address. First, the audience is not necessarily even aware of the public display in the first place, or they might not be aware that the display can be interacted with. Second, users need to be motivated to start interacting with the display. Third, the fact that interaction with the display happens in public should be accounted for. For instance, people may avoid interaction completely or partially because of their role (e.g. police officer) or physical limitations (e.g. an elderly person). Parra *et al*. [7] add a fourth issue that public displays should address: users should reach a goal or "final stage" of interaction with the system.

Very few serious public displays use gestures for direct interaction. The cardiac arrest awareness campaign [7] introduced a system where passersby could touch their heart in front of the display, after which a short video was played. The WaveWindow [8] enabled users to scroll through a list of items by waving in front of the display. Other deployments include a survey tool [12], and the Media Ribbon [9]. All these systems share an aspect in that they introduce only a limited set of simple gestures – most likely a desirable trait when utilizing new ways of interaction such as mid-air gestures in public displays. Further, most systems display instructions for interaction, usually in plain-text form.

3. INFORMATION WALL PROTOTYPE
Information Wall is a public information display that offers its users access to simple information like the lunch menus of nearby restaurants and events taking place around the campus. The Microsoft Kinect sensor is used to detect user movements.

When no users are present, the wall displays a dialog in which users are encouraged to try interacting with the wall by stepping closer and moving their hands in mid-air (Figure 1A). The Information Wall also attracts passersby by displaying rectangular

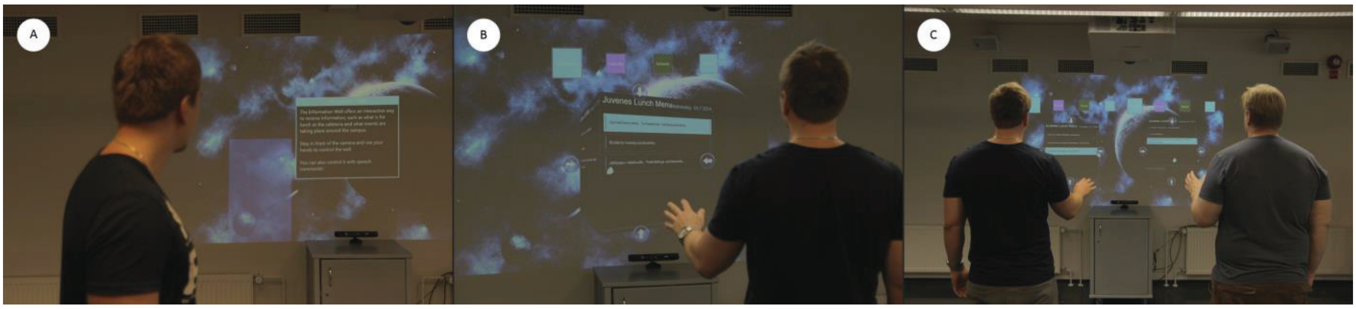


Figure 1. A) A user is tracked with a rectangular shape on the screen. An instruction dialog is displayed. B) A user interacts with an information cube. C) Two users simultaneously interacting with the wall.

elements on the display that follow the movements of the passersby (Figure 1A). Whenever a user steps into the interaction zone of the display (2.8 meters or less away from the Kinect sensor), a three-dimensional information cube is opened on the screen (Figure 1B). The system supports two simultaneous users, in the case of which the interface adjusts to the users' location and opens two information cubes, one for each user (Figure 1C). A cube is closed whenever the corresponding user leaves the scene, and remaining cubes adjust to make use of the whole screen.

Interaction happens via an on-screen cursor which moves according to where the user is pointing. Pointing uses the physical interaction zone algorithm provided by the Kinect SDK. Different functions are launched through buttons, which are triggered with the *dwelling* technique [12], in which hovering over a button with the cursor for a short period of time triggers the corresponding button. During dwelling a circular animation is displayed on the target to indicate that it is being triggered. Target selection is made easier by utilizing the magnetic cursor technique [4].

Figure 2. Selectable elements of the information cube. A) Shortcuts for switching to another section. B) Arrows to trigger the rotation of the cube. C) Entries providing information, which can be triggered to open a detailed information dialog.

The content of the display is navigated by rotating the information cube in desired directions. Rotation happens by triggering a button at the edge of the cube (Figure 2B) and then moving the cursor over to the other side, as if a physical object was being rotated. When a rotation button is triggered, an arrow animation is played to point to the direction of the rotation. Rotating to the right and left will change to the next and previous view inside the current section, e.g., one can switch from today's lunch menu to tomorrow's lunch menu. Rotating up and down will change the section, e.g., from the lunch menu to the latest news. Sections can also be changed via shortcuts (Figure 2A). More detailed information, e.g., the ingredients of a dish or the full story of a news headline, is provided after triggering the corresponding entry from the cube (Figure 2C), which will open a separate dialog in front of the cube. The dialog can be dismissed by triggering a button in the bottom right corner.

4. USER STUDY

We conducted a lab-based user study to evaluate how users perceive and interact with the system in a controlled setting, and to identify the strong and weak points of the system. We recruited 19 participants (16 male, 3 female) for the user study, who were between 19 and 56 years of age (median = 21). The participants were first-year interactive technology students, most with little experience of gesture-controlled systems. The participants received credit towards the completion of an undergraduate course as compensation for their participation.

The Information Wall interface was displayed on a 1920 x 1080 full HD projection screen with physical dimensions of roughly 1.8 (width) by 1.1 (height) meters, and the Kinect sensor was positioned at the horizontal midpoint under the screen. Participants were instructed to stand roughly 2.4 meters away from the screen center.

4.1 Procedure

The participants started the experiment by filling out a short background questionnaire. Afterwards, an introduction to the information wall was presented, however no instructions on how to interact with the system were given. Before the task session, participants were asked to fill out a questionnaire measuring their expectations of the system. The content of the questionnaire (Figure 3) was based on the SUXES user experience evaluation method [10] and consisted of 9 claims to which participants responded on a 7-point scale, where 1 = "strongly disagree", 4 = "neither agree nor disagree" and 7 = "strongly agree".

During the experiment, the moderator gave out tasks for the user, most of which were about finding information or learning a certain interaction. An example of such task would be "There's a seminar this Friday that's open to the public. Which classroom will it be in?" Afterwards, the moderator joined the participant in order to demonstrate how the information wall works with two simultaneous users. During this period the moderator and the participant engaged in a conversation about user experiences and opinions of the system. Lastly, participants filled out the second part of the SUXES method (experiences) as well as the AttrakDiff questionnaire [2]. In addition, we also asked participants what kind of content they would like to see in the system in the future.

4.2 Results

The participants' responses to expectations and experiences as measured with the SUXES statements are presented in Figure 3 as boxplots showing the median and interquartile range. In five out of nine claims the experiences meet the users' expectations, in three claims exceed them and in one claim did not meet the expectations. A Wilcoxon signed-rank test shows that in two claims the experience significantly exceeded the expectation: "The system performs correctly" ($Z = -2.705$, $p < 0.01$) received a median of 5 while the expectation was only rated as 3, and "Using the system is easy to learn" ($Z = -2.070$, $p < 0.05$) received a median of 6 with the expectation receiving a median of 5. No other significant differences were found.

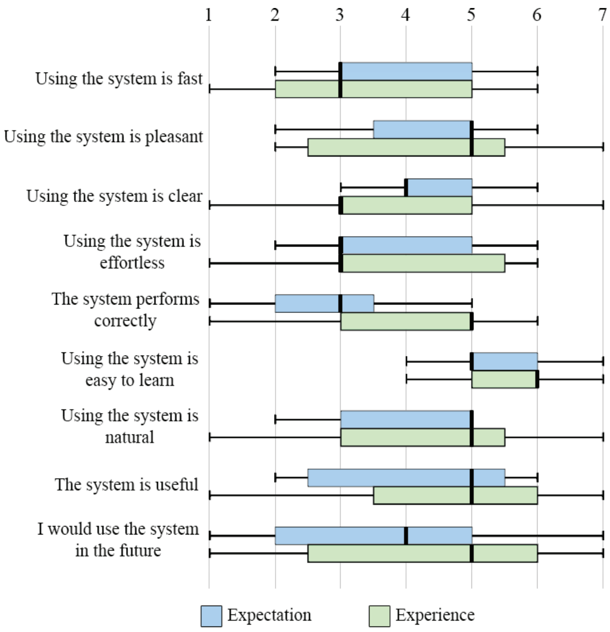


Figure 3. Distribution of users' expectations and experiences.

The clarity, speed and level of effort of use were considered to be slightly negative (median = 3), which suggests that the interaction and interface design had minor usability issues. The system was found fairly pleasant and natural to use (median = 5 for both) and was considered useful (median = 5), which is also supported by indication that the participants would consider using the system in the future (median = 5).

The main results of AttrakDiff are divided into two categories: Pragmatic Quality (PQ), which describes the usability of the system and how successful users are in achieving their goals with it, and Hedonic Quality (HQ), which is concerned with properties such as novelty, how interesting the system appears and what kind of experiences it offers. Both PQ and HQ received slightly higher than average scores.

The mean values for all word pairs are presented in Figure 4. Highest values are found in the hedonic quality – stimulation category (HQ-S), whereas the lowest values are in PQ. This suggests that although the participants found the Information Wall to be neutral in its usability and practical utility, the key elements of experience lie in more hedonic dimensions such as inventiveness, creativeness, presentability and novelty.

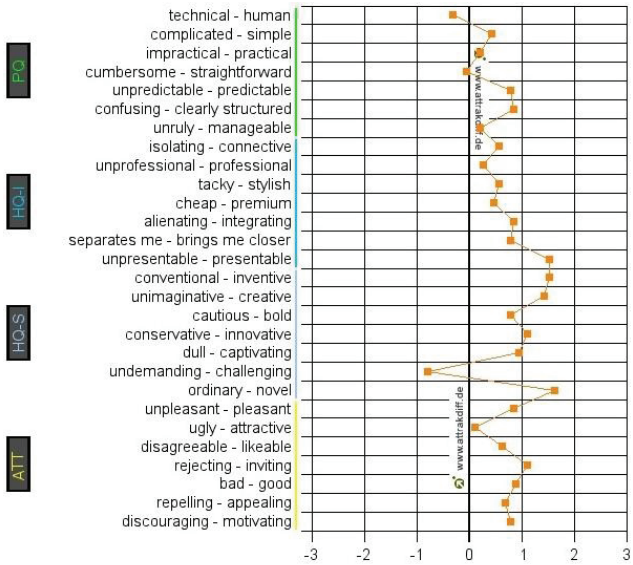


Figure 4. Mean values for AttrakDiff word pairs.

Additionally, participants were asked what kind of content they would hope to see in public information displays such as the Information Wall. 15 out of 19 participants were happy with the current content, most of which explicitly mentioned that they appreciated the lunch menus and the public events. Eight participants hoped to see public events extended to also cover recreational activities such as student parties and festivals. Other mentions were public transportation timetables, latest news and a map of the university campus. We have implemented some of these new content types in the next iteration of Information Wall.

4.3 DEMO SESSIONS

We conducted several short-term deployments at various events at the university campus, involving visitors from information technology companies, other universities and student course groups. The number of participants has varied between 10 and 100 per event. The system setup for these sessions was similar to the user study. However, due to a large open-ended space, the physical size of the screen was larger at around 3 meters in width.

Visitors were encouraged, but not required, to interact with the installation. In all cases, a clear majority of participants interacted with the system once the first few users had the courage to try out the installation in front of an audience. This observation follows the "honeypot" effect [1]. Furthermore, almost all users interacted with the system simultaneously with another user, for example two students who knew each other would interact with the system at the same time. Our observation here is that having another person interact with the system increased the users' confidence and lowered the threshold for interaction in public. Due to mid-air pointing and cursor-based control, it is easy to interact with the system while at the same time observing the other user.

5. DISCUSSION

Our research was focused on measuring the user experience of the Information Wall and producing design guidelines for gesture-controlled public displays. While users' expectations were mostly met in the lab-based study test, it should be noted that the expectations were somewhat low to begin with. Users not only perceived – but also expected – the usability of the system to be rather low. This might suggest that people do not have high hopes for novel ways of interaction. It is worth noting that among the highest values for AttrakDiff word pairs were properties such as novelty, creativeness, and inventiveness. We see this as intuitive and encouraging considering the experimental and playful nature of the system, with which we aimed to offer new and exciting ways of interaction. However, low scores in pragmatic quality along with SUXES results suggest that more attention should be paid to practical usability. Users seem to appreciate quick and effortless interaction over a playful and exciting first-time experience. Based on the user study as well as public session observations, we identified guidelines to consider when designing gesture-controlled public displays:

Support simultaneous interaction: Observations show that even after the initial honeypot effect users prefer interacting in pairs. We encourage systems to support at least two simultaneous users.

Don't hold content back: Most study participants were satisfied with the system's content, however several participants mentioned issues with the way the content was presented. Indeed, the information cubes only display a very limited set of information at a time, while most of the screen space has no informational value whatsoever. We recommend designers to make use of the screen space and display content in a clearly structured manner. Do not hold content back to force interaction – if something can be shown right away, show it.

Aim for relevant content: Subjective feedback seems to indicate that users like to see local and timely content in information displays, e.g., content that is related to nearby areas and relevant right away or in the near future. Among the most popular content were lunch menus, university events and leisure activities.

Display one cursor per user: Information Wall displayed cursors for both hands if they were high enough to appear on the screen. This resulted in two issues. First, while the cursor icons were different for each hand, users had trouble recognizing which cursor was mapped to which hand. Second, due to the use of dwell time for target selection, users often unintentionally selected targets with their inactive hand. This was especially evident when users had items on one hand, such as a coffee cup. The inactive hand was thus high enough for the cursor to accidentally trigger a target. We recommend supporting the use of both hands but displaying a cursor only for the active hand. Furthermore, users had trouble recognizing their own cursors during simultaneous interaction. Designers should consider ways to map a user's cursor to his space on the screen with e.g. similar colors or icons.

Use target selection aids: Freehand pointing quickly results in physical fatigue and minor hand movement often makes trivial tasks such as selecting targets a nuisance. The effects of this issue can be minimized with selection aids such as the magnetic cursor [4], which lowers both the required level of motor accuracy and mental load to select targets and also makes interaction faster.

6. CONCLUSIONS

In this paper we presented the Information Wall, a gesture-controlled public information display that provides simple and relevant local information to users. The application has been studied in a lab-based user study as well as observed in several demo sessions and semi-public events where first-time users explored the system. Based on our findings, we produced several guidelines to consider when designing gesture-controlled public information displays, and also recognized issues with the current version of the Information Wall that will help us improve similar systems. We have also conducted a long-term deployment of our system, during which extensive log data has been gathered. Our future work includes analyzing this data as well as carrying out in-depth observations of Information Wall in situ.

7. ACKNOWLEDGMENTS

This research was carried out as part of the Finnish Strategic Centre for Science, Technology and Innovation on Built Environment (RYM Oy)'s Indoor Environment program.

8. REFERENCES

[1] Brignull, H, and Rogers, Y. 2003. Enticing people to interact with large public displays in public spaces. In Proc. INTERACT 2003. IOS Press.

[2] Hassenzahl, M., Burmester, M., and Koller, F. 2014. AttrakDiff™. Available at: http://www.attrakdiff.de.

[3] Marshall, P., Morris, R., Rogers, Y., Kreitmayer, S., and Davies, M. 2011. Rethinking 'multi-user': an in-the-wild study of how groups approach a walk-up-and-use tabletop interface. In Proc. CHI '11. ACM, New York, NY, USA, 3033-3042.

[4] Mäkelä, V., Heimonen, T., and Turunen, M. 2014. Magnetic cursor: Improving target selection in freehand pointing interfaces. In Proc. PerDis '14. ACM, New York, NY, USA, 112-117.

[5] Müller, J., Alt, F., Michelis, D., and Schmidt, A. 2010. Requirements and design space for interactive public displays. In Proc. MUM '10. ACM, New York, NY, USA, 1285-1294.

[6] Müller, J., Walter, R., Bailly, G., Nischt, M., and Alt, F. 2012. Looking glass: a field study on noticing interactivity of a shop window. In Proc. CHI '12. ACM, New York, NY, USA, 297-306.

[7] Parra, G., Klerkx, J., and Duval, E. 2014. Understanding engagement with interactive public displays: an awareness campaign in the wild. In Proc. PerDis '14. ACM, New York, NY, USA, 180-185.

[8] Perry, M., Beckett, S., O'Hara, K., and Subramanian, S. 2010. WaveWindow: public, performative gestural interaction. In Proc. ITS '10. ACM, New York, NY, USA, 109-112.

[9] Tomitsch, M., Ackad, C., Dawson, O., Hespanhol, L., and Kay, J. 2014. Who cares about the content? An analysis of playful behaviour at a public display. In Proc. PerDis '14. ACM, New York, NY, USA, 160-165.

[10] Turunen, M., Hakulinen, J., Melto, A., Heimonen, T., Laivo, T., and Hella, J. 2009. SUXES – user experience evaluation method for spoken and multimodal interaction. In Proc. Interspeech 2009. ISCA, 109-112.

[11] Walter, R., Bailly, G., and Müller, J. 2013. StrikeAPose: revealing mid-air gestures on public displays. In Proc. CHI'13. ACM, New York, NY, USA, 841–850

[12] Walter, R., Bailly, G., Valkanova, N., and Müller, J. 2014. Cuenesics: using mid-air gestures to select items on interactive public displays. In Proc. MobileHCI '14. ACM, New York, NY, USA, 299-308.

Paper V

Mäkelä, V., Korhonen, H., Ojala, J., Järvi, A., Väänänen, K., Raisamo, R., & Turunen, M. (2016). Investigating mid-air gestures and handhelds in motion tracked environments. In *Proceedings of the 5th ACM International Symposium on Pervasive Displays (PerDis '16)*, 45–51. New York, NY, USA: ACM. doi:10.1145/2914920.2915015

Investigating Mid-Air Gestures and Handhelds in Motion Tracked Environments

[1]Ville Mäkelä, [1]Hannu Korhonen, [2]Jarno Ojala, [2]Antti Järvi,
[2]Kaisa Väänänen, [1]Roope Raisamo, [1]Markku Turunen
[1]University of Tampere, Tampere, Finland
[2]Tampere University of Technology, Finland
{ville.mi.makela, hannu.juhani.korhonen, roope.raisamo, markku.turunen}@sis.uta.fi,
{jarno.ojala, antti.v.jarvi, kaisa.vaananen}@tut.fi

ABSTRACT
Smart spaces with multiple interactive devices and motion tracking capabilities are becoming more common. However, there is little research on how interaction with one device affects the usage of other devices in the space. We investigate the effects of mobile devices and physical interactive devices on gestural interaction in motion-tracked environments. For our user study, we built a smart space consisting of a gesture-controlled large display, an NFC reader and a mobile device, to simulate a system in which users can transfer information between the space and personal devices. The study with 13 participants revealed that (1) the mobile device affects gesturing as well as passive stance; (2) users may stop moving completely when they intend to stop interacting with a display; (3) interactive devices with overlapping interaction space make unintentional interaction significantly more frequent. Our findings give implications for gestural interaction design as well as design of motion-tracked smart spaces.

Author Keywords
Large displays, mid-air gestures, mobile devices, handhelds, smart spaces, motion-tracked environments, ubiquitous computing.

ACM Classification Keywords
H.5.2. Information interfaces and presentation: User Interfaces – *Graphical user interfaces, input devices and strategies, interaction styles, user-centered design*.

INTRODUCTION
Interactive smart spaces have gained increasing interest in the research community over the years. While smart spaces can vary in many aspects such as size, purpose and the number of interactive devices, we foresee two logical future developments. First, device-free *mid-air gestures* will be a core component of smart space interaction. Gestures support the very nature of such spaces, as they enable interaction with devices at will from anywhere in the space. Second, ad-hoc connections between *personal devices* and devices integrated in the space will become more prominent. We argue that this is one of the strong points of smart spaces - e.g. allowing users to transfer information from the space to their mobile devices for later use.

PerDis '16, June 20-22, 2016, Oulu, Finland

DOI: http://dx.doi.org/10.1145/2914920.2915015

Our vision introduces a number of interesting research topics. For instance, what effects do mobile devices have on mid-air gestures? How are gestures affected by other devices in the space that utilize varying modalities? We are not aware of existing studies that investigate gestures or mobile devices from such perspective.

For our study, we built a smart space noticeboard system (originally presented in [6]) containing a) a large display application that allows users to select media content on the screen by utilizing mid-air pointing and dwell time, b) a mobile phone application that enables receiving and sending content items, and c) an NFC reader enabling communication between the large display and the mobile device. Two key design decisions for this infrastructure were to include personal devices, and introduce varying interaction modalities across devices integrated in the space.

Utilizing our infrastructure, we explored how people behave and gesture in a motion-tracked environment with multiple inter-connected devices. In particular, we were interested in studying the following:

- How does a mobile device affect mid-air gestures?
- How is the mobile device held during gesturing and when the user does not intend to interact with any device?
- How is a gesture-controlled large display affected by the presence of another interactive device in the same space?

Our findings can be utilized to design gestural interactions for ubiquitous smart spaces as well as to shape and design the space itself.

The rest of the paper is organized as follows. First, we present related work and the system in detail. Then, we present the study including the setup and procedure. Next, we present the results. Finally, we discuss the results and end with conclusion and future work.

RELATED WORK

Based on our literature review, we can identify a clear research gap. Most related work contains only a part of the infrastructure we aim to investigate. Many existing works deal with mid-air gestures or focus on communication between large displays and mobile devices. However, we are not aware of any studies that combine both, and more specifically, study the effects of gesturing and a mobile device on each other.

Design and Challenges of Gestures

A substantial amount of research is concerned with the design and challenges of mid-air gestures. For instance, *dwelling* is a target selection method for mid-air pointing, which involves pointing at a desired element on a screen and keeping the pointer steady for a short period of time, after which the element is selected. Dwell time has been found intuitive and easy to carry out, and has been utilized in a number of studies and large display applications [5, 8, 9, 15]. Mäkelä et al. showed the importance of targeting assistance techniques for gestural pointing interfaces [8]. Furthermore, Mäkelä et al. [9] and Ackad et al. [1] presented relatively coherent guidelines for gestural design. For instance, both emphasize simple designs and encourage supporting one-handed interactions.

Of the many challenges present in gestural interfaces, the Midas touch issue is one of the most prominent. The Midas touch issue describes the challenge of separating intentional and unintentional interaction, as any kind of gesturing can potentially be understood as input by a system [7]. The issue is also prominent in other modalities sharing the same always-on nature as gestures, such as speech.

Mobile Devices in Smart Spaces

Interactions between displays and handhelds have been studied actively. A common theme in this area has been taking control of the large display via the mobile device. Muta et al. [10] presented the WallSHOP advertisement display. Users could connect to the display by accessing a web service using their mobile device's browser. After connecting, users could browse items on the large screen by using their mobile device as a controller for a cursor displayed on the screen. Baldauf et al. [2] presented display pointing, where a handheld's camera is used to scan for a display, and after a display is recognized, connection between the two is established. Ruan et al. [12] used a mobile device's touch screen to act as a mouse and a keyboard for a computer, thus enabling control of a large display from a distance.

The aforementioned research focuses on establishing a continuous connection between the two devices and controlling a large screen directly via the mobile device. In addition, some studies aim to achieve quick and easy transferring of content between displays and mobile devices. Turner [14] proposes an interesting method where users can look at interesting content on the display and "pull" the content from the display into the phone by swiping down on their mobile touch-screens. However, Turner's approach requires external eye-tracking hardware as well as software.

A common method for establishing data transfer between a display and a mobile device has been the use of Near Field Communication (NFC) technology. Hardy & Rukzio [4] attached a grid of NFC tags on a large screen to allow users to move content, e.g. pictures, to a handheld device by touching the corresponding content on the screen. They found this interaction to be significantly faster and more intuitive than e.g. remotely controlling the display with a handheld. Broll et al. [3] presented a similar setup, in which they compared a set of simple interactions to more complex interactions, such as selecting multiple items at once. The results indicated that users preferred the simple touch-to-select interactions. Similarly, Rekimoto's pick-and-drop technique is used to pick up and drop items from and to a screen by touching NFC tags [11].

In summary, research involving mobile devices in smart spaces has largely focused on studying specific interaction techniques or solutions for connecting mobile devices and public systems such as large displays. We, on the other hand, primarily seek to investigate the high-level effects of a mobile devices's presence on interactive devices with motion-tracking capabilities.

INTERACTIVE NOTICEBOARD

With our smart space system, users can move content from the large display to a nearby NFC reader, and pick up the selected content to an NFC-enabled mobile device by touching the reader (Figure 1).

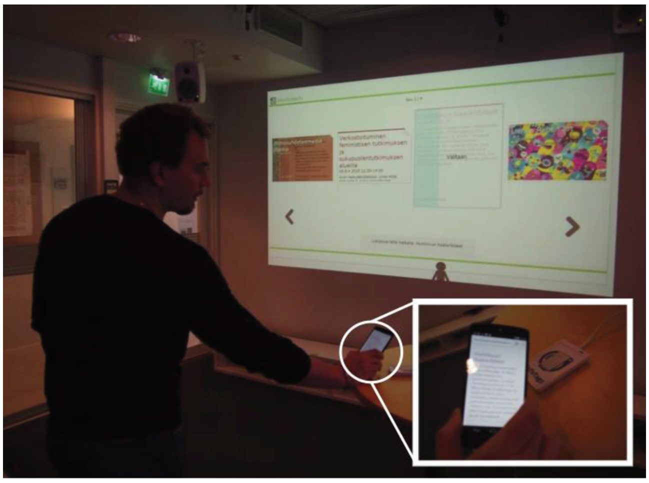


Figure 1. User has selected an item on the large screen and has transferred the content to a mobile phone by touching the NFC reader.

The selected content is then opened on the mobile device's default browser. Similarly, users can create new items for the large display by using a mobile application, and then touching the NFC reader to finalize the upload to the display.

Items on the screen are selected by using the *point-and-dwell* method, i.e. the user points at a desired item, and after an animation representing the selection has finished, the item is selected and is automatically moved to the reader. Simultaneously, the selected item on the screen is enlarged and the item that is currently in the reader is displayed at the bottom of the screen. The user then simply touches the reader with the mobile device, and the content is opened on the mobile device's default browser right away.

We designed the selectable items to be relatively large for easier target acquisition. Additionally, we used a simple targeting assistance method where the nearest target would always be active when a user is pointing towards the screen. For pointing, we used the Physical Interaction Zone method provided by the Kinect SDK. A dwell time of 1.5 seconds was used for target selections based on our previous system [9]. Furthermore, based on earlier work [1, 9], we designed interactions so that they could be carried out with one hand, and the interactive hand could be changed at any time.

The content on the display is shown as separate items which resemble post-it notes. Content is divided into pages, each page displaying 4 items in a horizontal layout. Additionally, two buttons are statically placed on the bottom corners of the screen to enable navigation between pages. The buttons are triggered with the same point-and-dwell method that is used for item selections. When no users are present, the page automatically changes every 15 seconds.

By using the mobile application, users can add their own content to the screen. The application is a simple form-filling application, which allows the user to include basic information such as a title and a description. Additionally, users can add an image and a date and a time.

When the user is finished with the form, the item is uploaded to the server by touching the NFC reader, and it is immediately updated to the large screen. The appearance of the notice is affected by its content, e.g. items with dates are visually different from others, and items with an image attachment are displayed with said image in the background.

Technical Details

The physical setup consists of a PC, a wall-projected full HD display, the Microsoft Kinect® sensor, and an ACS ACR122 NFC reader. Moreover, for the purpose of this study, participants were provided with an Android phone (Nexus 5), however outside test conditions the mobile devices would be the users' personal ones.

Our software consists of the large display client, a web service, an Android client, and two middleware components to handle data between hardware.

The large display client is a web-based application running in a browser, built with HTML5 and JavaScript. A separate middleware component (previously used in [9]) is used for getting the motion data from the Kinect sensor into the client. A web service takes care of storing and providing the large display content. Another middleware component handles messages between the NFC reader and the web service. The large display client fetches the content directly from the web service. The web service allows moderation of content via a separate admin component, and would also allow linking several installations together; however in this study we focused on one system.

The Android application is a standard Android activity containing a form for the creation of items. The application is not mandatory for taking content from the screen. If the app is installed, the content will be opened inside the app. Otherwise the content will be provided directly by the web service and will open in the device's default web browser.

STUDY

We ran a study with 13 participants (7 males, 6 females) aged 20 to 38 (mean = 28 years). Participants were recruited from around a university campus by promising a movie ticket as a reward for participation. The study was conducted in a research unit's semi-public space. The space is used for various activities such as lectures and meetings, but also serves as an open space for students and staff, and it is frequented by e.g. students doing group projects. The space is regularly used to run interactive public displays for people to use freely. The study was controlled, however the space itself represents a typical installation location for such infrastructures - a large room where people can come and go freely.

We used a high-definition projector with physical dimensions of roughly 1.3 meters in height and 2.2 meters in width. The software was run on a high-end Windows PC. The Kinect sensor was placed right under the screen at horizontal midpoint, and the NFC reader was placed on a small table near the middle point of the Kinect's field of vision for easy access.

We recorded the sessions with a full HD video camera for later analysis. The camera was positioned behind the primary interaction space so that it had a good view of the user as well as the display and the NFC reader, and it could also record all dialogue during the session.

The content on the noticeboard consisted of 15 items of different real-world university events, and also coupons for recreational activities, casual messages and photographs, and promotional material of a fictional graphic artist.

Procedure

Participants were first presented with a background questionnaire as well as a consent form to allow video recording of the session. Then, we handed a mobile phone (Nexus 5) with our application installed to the user, but gave no further instructions on how to use it or how to hold the phone during the tasks.

Participants were presented with a total of six tasks based on typical usage scenarios. The tasks were divided into three *information retrieval tasks* and three *content creation tasks*. In information retrieval tasks, participants were asked to find a specific notice on the board, explore it and move the notice to the mobile device for further exploration. For instance, participants were asked to retrieve information about an upcoming social event and how to sign up for the event. In content creation tasks, participants used the mobile application to create a notice fit for a given scenario, touched the NFC reader to upload it to the screen, and used gesture-controls to find the uploaded notice on the screen to confirm that it had been correctly uploaded. In these tasks, participants e.g. promoted a rally against animal abuse (used the board to organize an event), and created another notice to promote a new study space for students (used the board to provide contextual information).

Finally, participants were interviewed, after which they answered statements on a 7-point Likert scale. The statements were primarily about the ease of use of the system, naturalness of gestures, and naturalness of transferring content from one device to another. Each session lasted around 60 minutes in total and participants were rewarded with movie tickets (worth 12€).

RESULTS

We analyzed users by utilizing recorded video, interviews, and questionnaires. The recordings were watched twice by a single researcher, during which observations and notes were made. In the second round, we focused more heavily on behavioral similarities between users.

Interaction with the system's components can be divided into three categories based on both interaction modality and nature of the device: *large display interaction* (gestures, public), *handheld interaction* (touch, personal device), and *NFC interaction* (proxemics, hub). In the following, we present our findings in each category.

Large Display Phase

Participants understood the point-and-dwell interaction well, and although some took a short while to adjust to the sensitivity of pointing, participants were seemingly fluent with gesturing in general. Participants reported that mid-air gestures were natural (Mdn = 6, mode = 6) and relatively easy to carry out (Mdn = 5, mode = 5).

During interaction with the display, most participants preferred to use one hand for gesturing and the other hand for holding the mobile phone (8 out of 13). Two participants left the phone on a table while they were gesturing and only picked it up when they needed to interact with the NFC reader or the phone itself. However, this may have been due to the study setting, as users would possibly not leave their own phones on a table in a public or semi-public setting. Finally, only two participants gestured with the phone hand, and one user mixed all three techniques.

We expected there to be a clear distinction between the active and inactive hand's positioning. However, participants who held the phone in their inactive hand, held it very high, usually against their chest or in mid-air at chest or even shoulder level (Figure 2). This occasionally resulted in gestures not working properly, as the system was not able to determine which hand should be defined as active. However, most participants disagreed when asked if holding the mobile device disturbed gesturing (Mdn = 2, mode = 2). This supports our observation of participants not understanding that the phone hand's position was the primary cause for the issues with gesturing. When not interacting with the display, users still held the phone relatively high, which often caused unintentional items selections on the screen.

Figure 2. Demonstration of two common poses for holding a mobile device during gesturing: against the chest (A), in mid-air (B).

We observed an interesting phenomenon during interaction with the display. Six out of 13 users intuitively "froze" when they wanted to pause interaction and simply observe or concentrate on something else for a moment. This happened after triggering a function on the display, or while transitioning from handheld/NFC interaction to interact with the display moment. For instance, a typical scenario for users was to change a page on the screen, and then pause interaction for a period in order to observe the newly displayed items. After they changed the page, they would simply continue pointing towards the page change button on the screen as they were observing the content. This often resulted in a second, unwanted page change or an item selection a moment later. Consequently, when asked if they had trouble selecting correct items from the noticeboard, most participants agreed (Mdn = 6, mode = 7).

Handheld Phase

Participants used a mobile phone to view content they had taken from the display and to create notices to be uploaded to the display. During interaction with the mobile, participants' behavior was relatively coherent. 11 out of 13 participants held the phone with two hands, positioning the phone close to the body around the upper stomach or chest and looking down. Two users held the phone at almost face level, having their heads tilted only slightly downwards.

Users were comfortable interacting with the mobile application, as 12 out of 13 participants reported a positive experience. Two users reported initial confusion due to the unfamiliar Android OS, both being regular iPhone users.

Users were holding the phone relatively close to the face for easier viewing of the screen. This was occasionally interpreted as input by the large display, and we observed a similar freezing effect as during large display interaction. Three users, while interacting with the phone, noticed movement on the screen and lifted their heads towards it. Participants were standing still and staring at the display as if waiting for the display to "understand" that they did not want to interact with it. None of the users tried anything else to stop the interaction, for instance, no one tried lowering their hands or turning away from the screen.

NFC Phase

Participants interacted with the NFC reader when they wanted to move the selected content from the screen over to the mobile, or when they had created a notice on the mobile device and wanted to upload it to the display. Participants reported that overall interacting with the NFC reader was intuitive and straightforward. Data transfer between the mobile device and the NFC reader felt natural (Mdn = 6, mode = 7). Finally, participants disagreed when asked if having to physically touch the NFC reader made content transfer cumbersome (Mdn = 2, mode = 1).

However, reaching towards the NFC reader in front of the screen was often interpreted as gesture input by the system. At worst, this resulted in another notice being selected before the participant could transfer the correct item to the phone. In addition, the reader was relatively specific about which part of the phone needed to touch the reader. Hence, participants had to touch the reader a few times or "rub" the phone against it in order to be successful.

We observed 11 out of 13 participants having issues with unintentional gestures during NFC interaction. Participants reported the issue in the interview by e.g. stating that the system "is hectic", "feels too responsive" or "does things I do not want it to do". We observed the same phenomenon during NFC interaction as with handheld interaction – turning away from the display and more towards the table the NFC device was located on would have solved the issue, but none of the users tried it.

It is worth noting that participants had a positive experience regarding the NFC reader itself, and clearly attributed the experienced issues towards the large display (i.e. the motion sensor). 10 out of 13 participants were familiar with some form of NFC technology, 9 participants being regular users of NFC-enabled public transit cards – cards that need to be read by a terminal, and also occasionally require rubbing for a successful read.

DISCUSSION

In our study, we investigated users and their behavior in a lightweight smart environment consisting of a gesture-controlled display, a mobile device, and an NFC reader. We were primarily interested in a) how the mobile device affects gesturing as well as non-interaction, and b) how another, physical device affects the use of a gesture-controlled display. Based on our literature review, we are not aware of other studies with similar goals. As such, only a part of our findings have relations to previous work. In the following, we discuss our findings:

Mobile device affects gesturing and passive stance. We found that holding a mobile device affects the hand's position, and the majority prefers to use the free hand for gesturing. While gesturing, users held the phone surprisingly high – often around the chest - which occasionally obstructed interaction as the system had trouble recognizing which hand was active. Moreover, when users did not intend to interact with the display, the phone hand – due to its elevation – was often a cause for unintentional interaction.

Users may stop moving when they want to pause interaction. We observed a freezing effect during large display interaction as well as handheld interaction. We recognized a total of 15 instances of the freezing effect from seven different users. The average duration for the effect was 5.5 seconds (SD = 2.93), the minimum being 2 seconds and the maximum 12 seconds. The effect was most apparent when participants interacting with the screen wanted to pause for a moment to observe the contents of the screen. Users would then effectively freeze in their current stance, leaving one or two hands lifted as they were. We believe the effect was due to users intuitively assuming that *motionlessness*, rather than stance, would equal passiveness. One participant made a descriptive comment while facing the display with both hands lifted up:

"Why does the display react to me even though I'm not doing anything?"

- Female, 26

It is possible that some instances of the freezing effect may have been due to confusion or uncertainty, i.e. not knowing what to do in a certain situation.

Unintentional interaction is a challenge. We observed instances of unintentional interaction during all phases. The most prominent phase for unintentional interaction was NFC interaction. The large display often interpreted users' movement as input when users were reaching towards the NFC reader, as users were still prominently facing the screen. While this issue is well known from related work [7, 9, 15], we observed it in much greater magnitude than before due to the involvement of the NFC reader as well as the mobile device.

Implications for Design

Based on our observations of the freezing effect as well as unintentional interaction during passive phases (when users did not move), we propose that interactions in motion-tracked environments should be designed so that *standing still should be a call to stop interaction*. For instance, dwell time may not be an optimal interaction mechanism for smart environments as it is ultimately a static pose and is prone to cause issues, even if by itself it has been considered intuitive [15]. Yoo et al. [16] have also recently proposed that other selection methods, such as the "push", may be preferred by users over the popular point-and-dwell method.

Turning away from a screen to pause interaction did not seem to be intuitive to users. Furthermore, while observing users it became clear that turning away from the screen does not fit the workflow of inter-connected device interaction. For instance, when transferring content with the NFC reader, users were often facing and looking at the screen instead, waiting to see confirmation on the content transfer. Consequently, this finding does give implications to designing layouts of similar spaces. Since users do not necessarily face the object with which they are interacting if it is otherwise within reach, *unrelated devices should be positioned so that their interaction space does not overlap with others'*. However, we still believe that checking a user's general facing direction is useful for high-level target filtering, such as ruling out targets behind the user.

The most serious challenge we identified was unintentional interaction in situations where the above mentioned solutions do not apply. In our case this was during NFC interaction, as reaching towards the reader logically included movement (reaching forward with the phone and touching the reader) as well as facing the display (inter-connectedness of the reader and the display). Unintentional interaction has been discussed in several studies. Walter et al. [15] identified the same issue with their gesture-controlled display – although not quite to the same extent – but they were unable to suggest a solution.

Schwarz et al. [13] propose an interesting method for detecting users' intention to interact with a system. They analyze a *combination of gaze and body pose, and finalize the intent with a gesture* (such as waving or showing an open hand to the camera). Their method is promising, although the gesture, albeit a simple one, needed to engage a system may slightly interfere with the walk-up-and-use design paradigm that users often expect from public systems especially. We believe the fundamental component here is gaze, which could be utilized to differentiate between devices with overlapping interaction spaces. For example, users interacting with the NFC reader were switching between looking at the reader and looking at the display, even if their body's facing remained the same. The resulting noise with gaze should be possible to separate from stable gaze resulting from the intent to interact with the display.

Finally, we note that some design decisions of our system were also factors in the prominence of unintentional interaction. To allow for smoother and faster interaction, we included a relatively strong targeting assistance technique, and also a short dwell time of 1.5 seconds for target selections. This combination, however, ended up working against the system's usability, as targets were easily triggered when users did not want to interact with the display. As we already noted, *we suggest considering more distinct gestures for interaction than point-and-dwell* or - at the very least - we suggest longer dwell durations.

CONCLUSION AND FUTURE WORK

In this paper, we investigated users and their behavior in a motion-tracked environment. The environment consisted of *a large display* application that allows users to select items on the screen by utilizing mid-air hand gestures, *a mobile phone* that enables receiving and sending said items, and an *NFC reader* enabling communication between the large display and mobile devices.

Our main findings are (1) the mobile device affects gesturing as well as passive stance; (2) users may stop moving completely when they intend to stop interacting with a display; (3) interactive devices with overlapping interaction space make unintentional interaction significantly more frequent. Our findings can be utilized in gestural interaction design as well as smart space design, for instance (A) not moving should not result in interaction, (B) distinct gestures, such as 'grab' or 'push', should be considered instead of the point-and-dwell method, and (C) smart positioning of interactive devices in a space can alleviate issues with gestural interaction.

We aim to incorporate our findings in future studies, in which we further investigate the use of gestural interaction for content transfer in motion tracked environments.

ACKNOWLEDGMENTS

We thank Tampere region ELY Centre, European Regional Development Fund (ERDF), and Tampere Region Economic Development Agency (Tredea) for funding this research.

REFERENCES

1. Christopher Ackad, Andrew Clayphan, Martin Tomitsch, and Judy Kay. 2015. An in-the-wild study of learning mid-air gestures to browse hierarchical information at a large interactive public display. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp '15). ACM, New York, NY, USA, 1227-1238. DOI=http://dx.doi.org/10.1145/2750858.2807532
2. Matthias Baldauf, Markus Salo, Stefan Suette, and Peter Fröhlich. 2013. Display pointing: a qualitative

study on a recent screen pairing technique for smartphones. In *Proceedings of the 27th International BCS Human Computer Interaction Conference* (BCS-HCI '13), Steve Love, Kate Hone, and Tom McEwan (Eds.). British Computer Society, Swinton, UK, UK, Article 45, 6 pages.

3. Gregor Broll, Wolfgang Reithmeier, Paul Holleis, and Matthias Wagner. 2010. Design and evaluation of techniques for mobile interaction with dynamic NFC-displays. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction* (TEI '11). ACM, New York, NY, USA, 205-212. DOI=http://dx.doi.org/10.1145/1935701.1935743
4. Robert Hardy, Enrico Rukzio, Paul Holleis, and Matthias Wagner. 2010. Mobile interaction with static and dynamic NFC-based displays. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services* (MobileHCI '10). ACM, New York, NY, USA, 123-132. DOI=http://dx.doi.org/10.1145/1851600.1851623
5. Luke Hespanhol, Martin Tomitsch, Kazjon Grace, Anthony Collins, and Judy Kay. 2012. Investigating intuitiveness and effectiveness of gestures for free spatial interaction with large displays. In *Proceedings of the 2012 International Symposium on Pervasive Displays* (PerDis '12). ACM, New York, NY, USA, Article 6, 6 pages. DOI=http://dx.doi.org/10.1145/2307798.2307804
6. Antti Järvi, A. 2015. *Design and Implementation of an NFC-Based Mobile Interaction Concept for a Smart Bulletin Board*. Tampere University of Technology (TUT), Tampere, Finland.
7. Rick Kjeldsen and Jacob Hartman. 2001. Design issues for vision-based computer interaction systems. In *Proceedings of the 2001 workshop on Perceptive user interfaces* (PUI '01). ACM, New York, NY, USA, 1-8. DOI=http://dx.doi.org/10.1145/971478.971511
8. Ville Mäkelä, Tomi Heimonen, and Markku Turunen. 2014. Magnetic Cursor: Improving Target Selection in Freehand Pointing Interfaces. In *Proceedings of The International Symposium on Pervasive Displays* (PerDis '14), Sven Gehring (Ed.). ACM, New York, NY, USA, Pages 112, 6 pages. DOI=http://dx.doi.org/10.1145/2611009.2611025
9. Ville Mäkelä, Tomi Heimonen, Matti Luhtala, and Markku Turunen. 2014. Information wall: evaluation of a gesture-controlled public display. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia* (MUM '14). ACM, New York, NY, USA, 228-231. DOI=http://dx.doi.org/10.1145/2677972.2677998
10. Masafumi Muta, Soh Masuko, Keiji Shinzato, and Adiyan Mujibiya. 2015. Interactive Study of WallSHOP: Multiuser Connectivity between Public Digital Advertising and Private Devices for Personalized Shopping. In *Proceedings of the 4th International Symposium on Pervasive Displays* (PerDis '15). ACM, New York, NY, USA, 187-193. DOI=http://dx.doi.org/10.1145/2757710.2757732
11. Jun Rekimoto. 1997. Pick-and-drop: a direct manipulation technique for multiple computer environments. In *Proceedings of the 10th annual ACM symposium on User interface software and technology* (UIST '97). ACM, New York, NY, USA, 31-39. DOI=http://dx.doi.org/10.1145/263407.263505
12. Hanqing Ruan, Yi Qian, Yong Zhang, and Min Zhou. 2010. TouchInteract: An Interaction Technique with Large Displays Using Touchscreen-Phone. In *Proceedings of the 2010 Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing* (UIC-ATC '10). IEEE Computer Society, Washington, DC, USA, 262-265. DOI=http://dx.doi.org/10.1109/UIC-ATC.2010.36
13. Julia Schwarz, Charles Claudius Marais, Tommer Leyvand, Scott E. Hudson, and Jennifer Mankoff. 2014. Combining body pose, gaze, and gesture to determine intention to interact in vision-based interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14). ACM, New York, NY, USA, 3443-3452. DOI=http://dx.doi.org/10.1145/2556288.2556989
14. Jayson Turner. 2013. Cross-device eye-based interaction. In *Proceedings of the adjunct publication of the 26th annual ACM symposium on User interface software and technology* (UIST '13 Adjunct). ACM, New York, NY, USA, 37-40. DOI=http://dx.doi.org/10.1145/2508468.2508471
15. Robert Walter, Gilles Bailly, Nina Valkanova, and Jörg Müller. 2014. Cuenesics: using mid-air gestures to select items on interactive public displays. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services* (MobileHCI '14). ACM, New York, NY, USA, 299-308. DOI=http://dx.doi.org/10.1145/2628363.2628368
16. Soojeong Yoo, Callum Parker, Judy Kay, and Martin Tomitsch. 2015. To Dwell or Not to Dwell: An Evaluation of Mid-Air Gestures for Large Information Displays. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction* (OzCHI '15), Bernd Ploderer, Marcus Carter, Martin Gibbs, Wally Smith, and Frank Vetere (Eds.). ACM, New York, NY, USA, 187-191. DOI=http://dx.doi.org/10.1145/2838739.2838819

Paper VI

Mäkelä, V., James, J., Keskinen, T., Hakulinen, J., & Turunen, M. (2017). "It's Natural to Grab and Pull:" Retrieving Content from Large Displays Using Mid-Air Gestures. *Pervasive Computing, 16*(3), 70–77. doi:10.1109/MPRV.2017.2940966

"It's Natural to Grab and Pull:" Retrieving Content from Large Displays Using Mid-Air Gestures

Ville Mäkelä, Jobin James, Tuuli Keskinen, Jaakko Hakulinen, Markku Turunen
University of Tampere, Finland

***Abstract*—Mid-air gestures have been largely overlooked for transferring content between large displays and personal mobile devices. To fully utilize the ubiquitous nature of mid-air gestures for this purpose, we developed SimSense, a smart space system which automatically pairs users with their mobile devices based on location data. Users can then interact with a gesture-controlled large display, and move content onto their handheld devices. We investigated two mid-air gestures for content transfer, *grab-and-pull* and *grab-and-drop*, in a user study. Our results show that i) mid-air gestures are well suited for content retrieval scenarios and offer an impressive user experience, ii) grab-and-pull is preferred for scenarios where content is transferred to the user, whereas grab-and-drop is presumably ideal when the recipient is another person or a device, and iii) distinct gestures can be successfully combined with common point-and-dwell mechanics prominent in many gesture-controlled applications.**

***Index Terms*—content transfer, large displays, mid-air gestures, mobile devices, smart spaces, ubiquitous computing.**

I. Introduction

Exchanging information between large displays and personal devices is a feature of growing importance in smart spaces. Past research has shown great interest in studying different interaction mechanics for such tasks.

Previous research has largely focused on proxemics-based devices, such as NFC (Near field communication) readers, to enable communication between displays and mobile devices [1][2][3]. Some novel solutions have also been proposed, such as one utilizing the camera of the mobile device to enable drag-and-drop interactions [4]. However, while these solutions may work well for a specific purpose, they also require users to spend time establishing a connection, or require users to specifically walk up to the device to carry out the task. Most notably, they require users to interact with their mobile device in order to successfully transfer content, which may, in many cases, require too much effort from the user. This drawback calls for more advanced and seamless approaches for content transfer.

With this need in mind, we envision ubiquitous spaces where a) ad-hoc connections between mobile devices and large displays are established automatically, and b) content can be moved from one device to another by utilizing mid-air gestures. This vision enables users to simply pass by a display and "grab" interesting content onto their mobile device with a simple gesture and without the need to directly manipulate the recipient device.

In this article, we focus on exploring gestural interaction for content transfer. We built a novel smart space system, SimSense, capable of automatically pairing users with their mobile devices, allowing exchange of content between mobiles and the large screen with simple gestures and without the need to manually set up a connection. We report the results of a user study in which participants experienced two different gestures for information retrieval, *grab-and-pull* and *grab-and-drop*.

II. Designing Mid-Air Gestures For Content Transfer

In our previous work, we investigated transferring content between mobile devices and a large display by utilizing mid-air gestures and an NFC reader [5]. Users could select items from the screen with *point-and-dwell*, after which the selected content could be moved to the mobile device by touching the nearby NFC device with the mobile. We discovered that the point-and-dwell mechanic was suboptimal for content transfer. Primarily, when users were doing something other than interacting with the screen, the screen would still interpret the user's movement as interaction. Then, due to point-and-dwell, actions were occasionally triggered on the screen without the user's intent.

Our approach in this article is inherently different, as we aim for a more novel and seamless solution by completely negating the need to interact with the handheld during content transfer. Regardless, we utilize our previous findings by moving away from point-and-dwell mechanics and instead designing distinct gestures for content retrieval. Additionally, this allows us to utilize point-and-dwell for other, more lightweight interactions, such as navigating the content on the large display.

Received June 2016; revised October 2016. This work was supported in part by Tekes (The Finnish Innovation Agency) for UXUS (User experience and usability in complex systems) research programme under grant 1170/10.

V. Mäkelä, J. James, T. Keskinen, J. Hakulinen, and M. Turunen are with Tampere Unit for Computer-Human Interaction (TAUCHI) at University of Tampere, Kalevantie 4, 33100, Tampere Finland. Contact the authors at ville.mi.makela@sis.uta.fi, jobinjamez@gmail.com, tuuli.keskinen@sis.uta.fi, jaakko.hakulinen@sis.uta.fi, markku.turunen@sis.uta.fi

Our literature review did not reveal existing works that involve mid-air gestures specifically in the context of transferring content between devices. However, studies for more general-purpose mid-air gestures are numerous, which in part affected the design of our content retrieval gestures as well as our user study.

Similar to the findings in our previous work, Yoo and colleagues [6] recently found that users preferred gestures such as *pushing* and *grabbing-and-pulling* over the popular point-and-dwell method for interactions with large information displays. Other techniques for interacting with on-screen item from a distance have also been proposed, such as AirTap and Thumb Trigger [7], and the Gunslinger [8]. These techniques are comparable to usual target selection interaction techniques such as the point-and-dwell. With SimSense, we extend *beyond* the selection of on-screen items, and enable transferring those items to one's personal device. We believe such interactions would benefit from gestures that suggest a link between the large display and the recipient device.

Several works present guidelines for designing mid-air gestures. Nielsen and colleagues [9] provide a set of guidelines for user-centered gesture design. They list that gestures should be 1) easy to perform and remember, 2) intuitive, 3) metaphorically and iconically logical towards functionality, and 4) ergonomic. In one of our earlier studies [10], we also listed several guidelines for designing mid-air gestures, which include e.g. designing for one-handed interaction, and allowing changing the gesturing hand on the fly. Ackad and colleagues [11] studied the design of learnable mid-air gestures, and also presented a set of relevant guidelines, which include e.g. that gesture sets should be small, and on-screen feedback for gestures is important. In their user study, Ackad and colleagues found that an on-screen skeleton helps users understand how the system interprets their movement.

III. SIMSENSE

We developed SimSense[1], a smart space system that allows seamless transferring of content from a large information display to mobile devices by utilizing mid-air gestures and automatic connections between the display and mobile devices [Fig. *1*A]. SimSense consists of a large display application, a Kinect sensor, a mobile application, and a set of Bluetooth beacons (the system is presented in detail in [12]).

SimSense is permanently deployed within the premises of the School of Information Sciences at the University of Tampere. The deployment space is occasionally used for lectures and meetings, but is primarily a relaxed, open space filled with tools and displays for students and staff to utilize freely.

The large display application of SimSense runs on a full HD projector screen with physical dimensions of roughly 3.0 meters in width and 1.5 meters in height. The Kinect sensor is positioned at the horizontal midpoint below the screen. Five Bluetooth beacons are positioned on the ceiling of the space above this area, one on each of the four corners and one in the center.

Fig. 1. A) A user is transferring content from the large information display to his mobile device by performing the grab-and-pull gesture. B) An overview of the elements in the large display application. C) "New feed(s) available" notification on the mobile device about received content. D) The SimSense mobile application displaying content that was transferred from the large display.

A. Large Display Interaction

The SimSense large display application serves as a gesture-controlled information display. The primary elements of the display are presented in [Fig. *1*B]. The user's hand movements are visualized with an on-screen hand cursor. Interactive elements are triggered with the point-and-dwell mechanic, i.e., by pointing at an element and waiting for a short period of time. During dwelling, a progress bar is overlaid on the element to visualize that the element is being selected. Dwell time was set to be relatively long, 2.5 seconds, to prevent accidental target selections while performing content retrieval

[1] https://youtu.be/RkpjCsNBu3U

gestures.

Interactive elements include content elements and navigational elements [Fig. 1B]. Triggering a content element will display the content of said element in full in a new page. The page can be closed via a close button displayed on the bottom of the page. Navigational elements can be triggered to cycle between different pages of content.

The user is visualized on the screen in two ways [Fig. 1B]. An avatar is displayed on the bottom of the screen to reflect the user's horizontal position relative to the screen. Furthermore, the same figure is utilized for feedback. Secondly, we visualize the user's joint movements with an animated stick figure on the top-right corner of the screen.

The large information display can be easily set to present any kind of content. For the study, we chose to include news from Yle, a Finnish news portal, which are fetched via a public API.

B. Connecting Mobile Devices and Users with a Large Display

When a user enters the space with a mobile device that has the SimSense mobile application installed, the device is automatically paired with its owner based on location data. In order to determine the location of the mobile device, the area in front of the large display application is divided into 8 blocks of 50cm X 50 cm dimension, and 100 RSSI (Received Signal Strength Indicator) values from five Bluetooth beacons attached to the ceiling of the space were measured at each of these blocks and stored in a database. The location of the device is predicted by comparing the RSSI values to these previously stored set of values. The location of the user is received via a Kinect sensor. If the two locations match, the entities are paired. Consequently, all content that the user selects on the large screen to be transferred, will be sent to the correct device.

We leave the thorough measurement of the accuracy of the pairing process for another paper; however the process has proven reliable during our evaluations. It is possible that the process fails if two unpaired users enter the space at the same time while being very close to each other and without separating before starting interaction, which is unlikely. We have planned a fallback method for such situations, wherein the system would ask users to step away from each other to complete the pairing reliably. Interaction-wise, two simultaneous users are supported, as the Kinect sensor can accurately track two users at the same time.

C. Content Retrieval

Based on the findings of our previous work as well as existing gesture design guidelines, we designed two content retrieval gestures, *grab-and-pull* and *grab-and-drop*.

Grab-and-pull and grab-and-drop gestures are both initiated by first pointing at a transferable element on the screen and performing a grab gesture (forming a fist with the pointing hand). We note that pointing at an element will also begin the dwell time based selection on said element. To prevent accidental selections, dwell time progression is stopped when the element is grabbed, i.e., when a content retrieval gesture is started.

With grab-and-pull, the user then moves the hand back while keeping the grab gesture – as if physically pulling an object from the screen towards the user. The target element gets larger as the pull progresses to visualize that it is being pulled from the screen. After pulling a certain distance, the grab-and-pull is completed and the content is transferred [Fig. 2A]. The pull distance was tested via pilot studies and was set to 150 mm.

With grab-and-drop, a drop area is generated on top of the user's avatar when an element is grabbed. The user then drags the grabbed element into the drop area and releases the grab, after which the content is transferred [Fig. 2B].

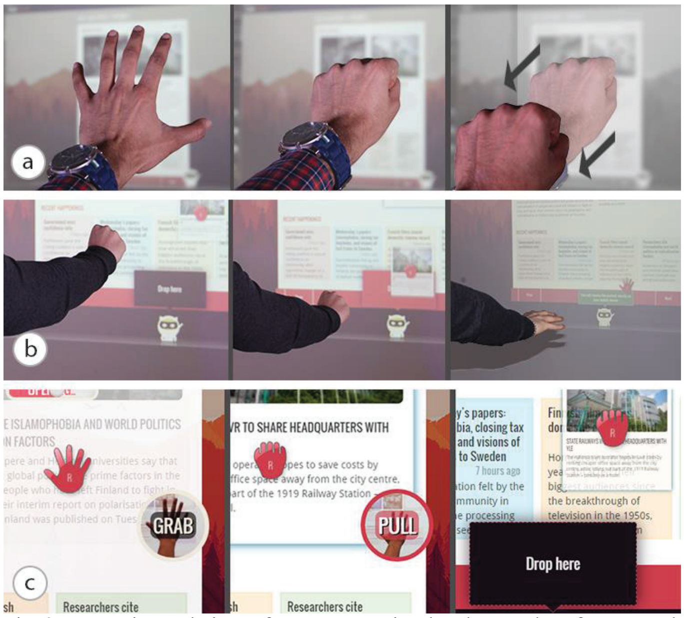


Fig. 2. Interaction techniques for content retrieval and examples of contextual feedback. A) Grab-and-pull: 'grab' and 'pull' an element from the screen. B) Grab-and-drop: 'grab' content and drag and drop it onto a drop area. C) Left: an animated icon visualizing a 'grab' gesture when a transferable element is being hovered. Middle: an animated icon visualizing the 'pull' phase of grab-and-pull after the user has grabbed the corresponding element. Right: with grab-and-drop, a drop area is generated on top of the user's avatar when an element has been grabbed and can be moved around the screen.

Based on the characteristics of both gestures, we produced two minor hypotheses. First, we assumed grab-and-pull to be faster. This was because grab-and-drop requires users to position a dragged element in a specific area, whereas grab-and-pull only requires a one-dimensional pull. Second, we assumed grab-and-drop to appear more familiar and hence result in e.g. increased clarity and confidence. This was due to grab-and-drop's similarity to traditional drag-and-drop mechanics, common in e.g. mouse and touch based interactions.

With both content retrieval gestures, users need to first reach out with their hand (straighten the hand) before elements can be grabbed. This is because i) we wanted to avoid accidental content transfers and ii) the Kinect sensor occasionally triggers false positives and negatives with the grab - reaching out with the hand alleviates the issue somewhat. However, more trivial interactions with the screen, such as navigating the content and reading news, do not require the hand to be reached out.

D. Mobile Application

The SimSense mobile app is an Android application that needs to be installed on the user's mobile device to enable automatic pairing and content transfer. The application triggers a notification on the device whenever new content is received [Fig. *1*C]. The application displays all received content in a scrollable list [Fig. *1*D]. With the prototype, we concentrated on simple textual and image-based content, such as news. There are currently no other functionalities in the mobile application, although we will investigate extending it in the future.

E. Feedback

One of our emphasized themes in the design was multi-sensory feedback to the user, in a similar manner to what Vogel and Balakrishnan [7] utilized in their study of mid-air pointing and clicking techniques. First, we included traditional visual and auditory feedback on all interactions, such as when an element is hovered or grabbed. Second, and more importantly, we included contextually changing feedback [Fig. *2*C]. Finally, the avatar on the screen reacts to the location of the user by moving horizontally on the bottom of the screen, while an animated stick figure reacts to the user's skeletal movements [Fig. *1*B].

Additionally, feedback is provided on the mobile application. Whenever content is received, a simple audio tone is played and the mobile vibrates. Moreover, received content is immediately available on the mobile device to be explored further.

IV. User Study

We conducted a 12-participant user study (six males, six females) following a within-subjects design. The participants were 21–43 years old (M=27, SD=6.4). Most participants were students; however their educational background varied from, e.g., linguistics to business. Their prior experience with mid-air gestures was relatively low: three participants had no prior experience, six had used such systems once or twice, and three reported using such systems a few times a year.

We wrote a script for the user study within the SimSense system that would randomly highlight one news headline from the main page of the application. The participant should then select the corresponding headline and transfer it to the mobile device by utilizing one of the gestures (only one was available at a time). After a successful task, the system would wait for three seconds until highlighting the next item.

In the beginning of each study session, the participant filled in a background questionnaire and signed a consent form for video recording. Then, participants were introduced to the application. Next, a mobile phone with the SimSense application was given to the participant, and all participants put the phone in their pocket.

Each participant experienced both techniques (in a counter-balanced order) according to the following pattern:

- Watching a 5-second video demonstrating the technique once.
- Filling in an *expectations* questionnaire about the demonstrated technique.
- Performing 20 content retrieval tasks, with a short break after the first 10.
- Filling in an *experiences* questionnaire about the technique.

After all tasks, subjects filled in a questionnaire comparing the two techniques. Finally, participants were interviewed.

We used a simplified SUXES method [13] to measure participants' expectations and experiences of both interaction techniques. According to the SUXES method [13], *expectations* can show how significant the different factors are and by themselves already provide insight into how people perceive new types of interactive systems, therefore providing better understanding of the user experience. Moreover, we hypothesized that natural user interfaces seen in, e.g., various science fiction films may have resulted in raised expectations, which could not necessarily be met in reality.

We also included open-ended questions, and had participants choose their preferred technique based on varying categories (described in the following chapter). Finally, objective measures from the content retrieval tasks were logged.

V. Results

Participants' expectations and experiences for both techniques are presented in [Fig. *3*A]. As we hypothesized, expectations were indeed very high in the majority of categories; however we were able to answer with an equally positive user experience.

Grab-and-pull exceeded expectations in five out of eight categories, and met the expectations in three. Grab-and-drop method, on the other hand, failed to meet the expectations for three of eight statements, although it also met the expectations for four conditions, and exceeded them in one.

Additionally, we presented three statements that were only answered after the experience [Fig. *3*B]. Combining statements from both Fig. *3*A and Fig. *3*B, the user experience of grab-and-pull received higher median scores than grab-and-drop in 10 out of 14 categories. Grab-and-drop scored higher in one category (novelty), and for the remaining three, median scores were the same (naturalness, fun element, confidence that content retrieval was successful).

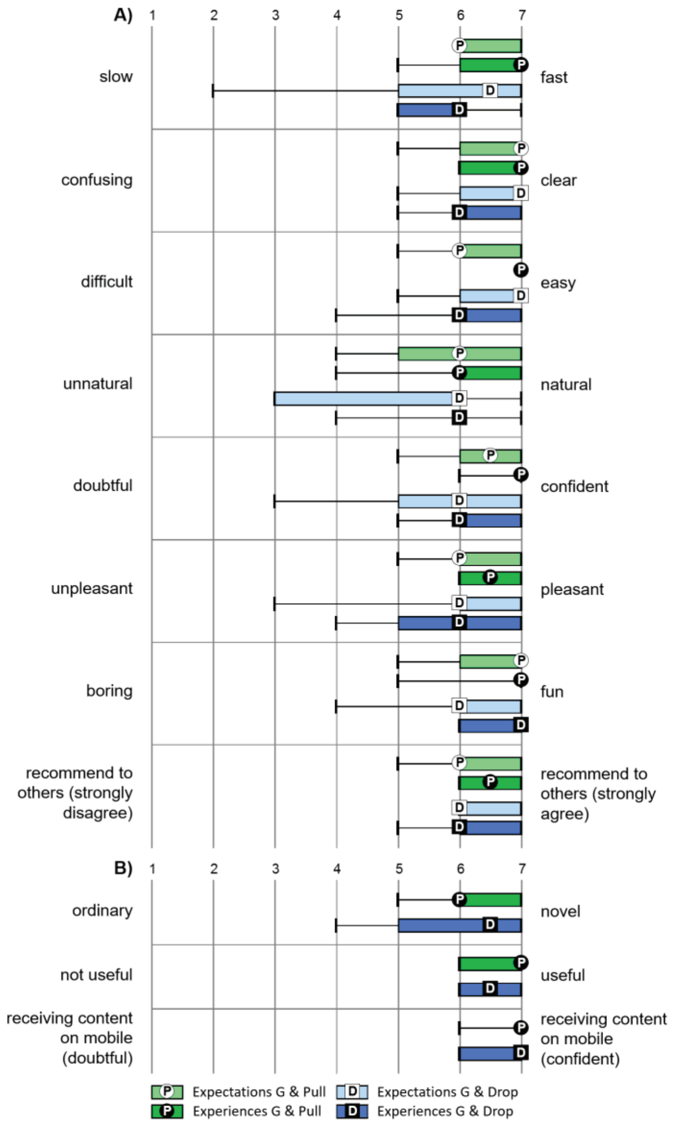


Fig. 3. Boxplots for expectations and experiences for grab-and-pull and grab-and-drop. P and D letters represent the median values.

We assumed that grab-and-pull would be faster than grab-and-drop. Indeed, in addition to users reporting it to be faster [Fig. 3A], a paired samples t test shows that the time to complete the grab-and-pull gesture was significantly shorter than the time for grab-and-drop; $t(11) = -10.002$, $p < 0.001$. Grab-and-pull took an average of 0.65 seconds (SD=0.20), whereas grab-and-drop took an average of 1.99 (SD=0.48) seconds.

In general, both techniques received high scores in all categories, as all median scores landed within the 6-7 spectrum. Presumably, one factor in high UX scores for both techniques was that they worked well and users were able to interact with the system successfully. This is supported by the fact that with grab-and-pull, there was only a single time when incorrect content was moved to the mobile, and with grab-and-drop, there were only eight such instances (from two different users). Moreover, instances where a content retrieval gesture was started but not completed were also rare. Users performed an average of 3.0 unfinished gestures (SD = 3.1) with grab-and-pull, and 2.75 unfinished gestures (SD = 2.59) with grab-and-drop. However, we attribute the majority of these occurrences to the Kinect sensor, as the sensor occasionally reported both false positives and false negatives regarding whether the user's hand was in the 'grab' position. A natural open hand stance, with fingers curved slightly inward, seemed to be in a relatively grey area which could be interpreted as either one by the sensor. Some users reported this by stating that sometimes the object they were dragging would "disappear" and they would need to start over. Grab-and-drop may have suffered more from this issue, as the disappearance of an object being dragged is easier to notice than that of an object being pulled (wherein it is more difficult to attribute the issue to a single factor).

We were also interested to see how content retrieval gestures would work in parallel with standard point-and-dwell selections. Positively, we did not encounter a single instance wherein a participant would have accidentally triggered an element via dwelling while aiming to transfer it to the handheld device. Moreover, no participants reported any issues in regards to dwell selections.

Participants named their preferred technique in four categories, and also named their overall preferred technique for content retrieval (Fig. 4). Nine out of 12 participants reported that grab-and-pull was more natural than grab-and-drop. More balanced results were received when asked which technique they were more confident performing (i.e., how sure they were that they were committing the gesture correctly), however the fun element and pleasantness categories again showed preference towards grab-and-pull. Many participants added that grab-and-pull felt more pleasant because it was slightly easier to perform. Finally, 10 out of 12 participants chose grab-and-pull as the overall preferred content retrieval method, while grab-and-drop was preferred by two subjects.

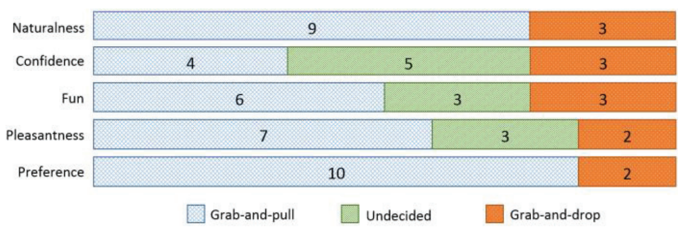


Fig. 4. Subjects' preferred techniques based on naturalness, confidence, fun element, pleasantness, and overall preference.

Many participants explained their preferences for naturalness based on who or what is receiving the content, e.g.:

> *If I want something for myself, it's simply natural to just 'grab and take', and not 'grab and move somewhere'.*
> \- Participant 1

> *Grab-and-pull was more natural and more appropriate for this specific purpose. But if I'm thinking multiple devices, transferring from one device to another, grab-and-drop would be more natural. I could also want to transfer content to another person, and again grab-and-drop would be a better choice.*
> \- Participant 3

Interestingly, four participants explicitly reported that they initially expected to prefer grab-and-drop, but changed their opinion after experiencing both techniques. Presumably, users were initially attracted by the perceived familiarity of grab-and-drop, but ultimately preferred the speed and ease of use of grab-and-pull.

Participants gave positive remarks of the feedback mechanisms. Generally, feedback was reported to be clear and helpful, and that the system was successfully guiding participants through the interaction process. One participant mentioned:

It was good to know that it [content] went to the phone immediately. I had a feeling that I always knew what was happening.
- Participant 5

Finally, participants were seemingly impressed and satisfied with the overall experience. Some commented:

I felt like I was in a movie, in the future.
- Participant 4

It [interacting] feels almost like I'm a superhero.
- Participant 5

[Interacting] is so much fun actually that you don't even care about the content, you just want to do it [interact].
- Participant 6

VI. Discussion

The multifaceted results of our user study suggest that *mid-air gestures are well-suited for content retrieval scenarios*. Both gestures, grab-and-drop and grab-and-pull, received high UX scores and were generally appreciated by participants. In addition, users reported a positive, novel and impressive experience with the system as a whole, and some users made remarks that interacting made them feel like they are superheroes or are a part of a futuristic movie. One participant reported having so much fun with the system that he did not even care about the content, but just wanted to keep interacting.

We successfully mixed point-and-dwell item selections with distinct (content retrieval) gestures. Participants did not commit accidental item selections, not even during content retrieval, nor did they report issues with mixing two different kinds of interactions. This is an encouraging finding and suggests that mid-air gestures can be utilized for a wider range of interactions than many existing works have proposed [9][10][11]. We believe the key factors in this success are that the two mechanics did not overlap beyond the initial act of pointing to an item, and that dwell time was stopped whenever an element was grabbed. Therefore, the dwell time of 2.5 seconds could likely be shortened if needed.

We were able to design two successful gestures with the help of existing guidelines. Based on recent findings, we made the initial decision to move away from dwelling-based mechanics [5][6]. In addition, we followed a number of existing gesture design guidelines [9][10][11] in designing the grab-and-drop and grab-and-pull gestures.

Interestingly, despite highly positive results for both gestures, *grab-and-pull was clearly preferred by users for content retrieval* (10 out of 12 participants). Based on our results, we identified three major factors for the outcome. First, grab-and-pull was reported to feel conceptually more natural, i.e., pulling content towards oneself to transfer it felt more appropriate than dropping it onto an area on the screen. Second, grab-and-pull was faster, and third, it was easier to perform as it required less accuracy. We note here that while the sample size of 12 was relatively limited, participants were recruited from various backgrounds, and the results were consistent across participants.

The reasons for preferring grab-and-pull are logical, however we were surprised by the relatively clear difference, especially considering the positive feedback and experiences for both techniques. We initially expected one of the strong points of grab-and-drop to be its familiarity from other modalities. Considering this, it is interesting to note that users may have had the same expectation: four users explicitly stated that they initially expected to prefer grab-and-drop, but switched to grab-and-pull after experiencing both techniques.

We note that despite grab-and-pull being preferred, said preference may only apply to a specific scenario of transferring content to *oneself*. Many users noted that it simply makes sense that when they want something for themselves, they pull it towards them. On the other hand, users also suggested other interesting scenarios, such as transferring content from one display to another, where they reported that grab-and-drop might work better than a pulling gesture, wherein a drop area on the screen [Fig. 2C] would represent the recipient display instead of the user. Another suggested scenario for grab-and-drop was transferring content to someone else, such as a friend. Therefore, we argue that grab-and-pull is favorable for transferring content to the user, and grab-and-drop is a valid option for scenarios where the recipient is someone or something other than the user. We believe that these techniques can even be mixed into the same system, both for their own purposes.

We note some limitations in our study. We did not measure the intuitiveness of the gestures due to showing participants short demonstrations before the study tasks. This was because we aimed to measure their expectations as opposed to experiences. Regardless, our results support our claim that the designed content retrieval gestures are intuitive as both were regarded as natural and logical. Moreover, Yoo et al. [6] investigated a relatively similar grab-and-pull gesture in a different scenario, and they also make a case for intuitive gestures.

Arm fatigue was also not measured, which is a known issue with mid-air gestures [14]. However, we did account for fatigue by including a break after each set of 10 content retrieval tasks, and no participants mentioned nor seemed to be getting fatigued. Moreover, interactions with large displays are generally short, as are realistic content retrieval scenarios: users would pass by the system, grab a few interesting items with them, and continue on their way.

It is possible that the participants' overall impressed attitudes would somewhat diminish over time when the novelty effect of the system wears of. However, the majority of our results are not directly related to the novelty of the interaction. For instance, the deciding factors in the preference of grab-and-pull were *naturalness*, *ease of use*, and *speed*.

One interesting question is how people feel about their mobile devices automatically connecting to a public service. No one reported privacy-related issues during the study, and the participants' overall positive attitude towards the system provides encouragement. However, the relaxed and calm deployment setting may play a role in this. Hectic and less familiar environments, such as city centers, would presumably propose privacy and other concerns. Another noteworthy addition to this discussion is the requirement of a separate mobile application. We argue this is a positive, as installing an application serves as a permission for the device to provide its location to the service. Generally, the aspect of user acceptance, privacy, and the effects of the deployment location should be investigated in the future.

VII. Conclusion

Mid-air gestures can offer an impressive user experience in content retrieval scenarios between large displays and mobile devices. We discovered an ideal gesture for personal content transfer interactions, *grab-and-pull*, wherein users "grab" an item from the screen and pull it towards themselves. We also successfully mixed content retrieval gestures with common point-and-dwell mechanics, which we believe encourages designing more diverse gesture-controlled applications.

Our work makes way for designing ubiquitous spaces wherein interaction between public and personal devices is made more seamless and accessible by enabling interaction from anywhere in the space. Possible continuations to our work include transferring content from the mobile device back to the large screen, between several large screens, and between people.

Acknowledgment

Authors gratefully acknowledge the support by Tekes (The Finnish Innovation Agency) for UXUS (User experience and usability in complex systems) research programme.

Biographies

All authors are affiliated with the Tampere Unit for Computer-Human Interaction (TAUCHI) at University of Tampere, Finland.

Ville Mäkelä is a Ph.D. candidate at TAUCHI. His research focuses on mid-air gestures and content transfer between public and personal devices. He received his M.Sc. in interactive technology from the University of Tampere in 2013. Contact him at ville.mi.makela@staff.uta.fi.

Jobin James was a researcher at TAUCHI. His research interests include design of interaction techniques for large public displays, and design of ubiquitous environments with such displays. He received his M.Sc. in human technology interaction from the University of Tampere in 2015. Contact him at jobinjamez@gmail.com.

Tuuli Keskinen is a postdoctoral researcher at TAUCHI. Her research focuses on user experience evaluation. In 2010, she received her M.Sc. degree and in 2015 her Ph.D. degree – both in interactive technology from the University of Tampere, Finland. Contact her at tuuli.keskinen@sis.uta.fi.

Jaakko Hakulinen is a university researcher at TAUCHI. His current research interests include multimodal interaction utilizing large scale of technology, e.g., gesture-based interaction, speech-based interaction, and intelligent environments. He received the M.Sc. degree in computer science from the University of Joensuu, Finland, in 1998, and Ph.D. degrees in interactive technology from University of Tampere, Finland in 2006. Contact him at jaakko.hakulinen@staff.uta.fi.

Markku Turunen is a professor of Interactive Technology at TAUCHI. His research interests include e.g. multimodal systems, novel user interfaces, and evaluation methods for mobile and pervasive computing applications. He received his Ph.D. in computer science from the University of Tampere in 2004. Contact him at markku.turunen@sis.uta.fi.

References

[1] G. Broll, W. Reithmeier, P. Holleis, and M. Wagner, "Design and evaluation of techniques for mobile interaction with dynamic NFC-displays", in *Proc. TEI'11*, 2011, pp. 205-212.

[2] R. Hardy, and E. Rukzio, "Touch & Interact: touch-based interaction of mobile phones with displays", in *Proc. MobileHCI'08*, 2008, pp. 245-254.

[3] K. Seewoonauth, E. Rukzio, R. Hardy, and P. Holleis, "Touch & connect and touch & select: interacting with a computer by touching it with a mobile phone", in *Proc. MobileHCI'09*, 2009, 36.

[4] M. Baldauf, K. Lasinger, and P. Fröhlich, "Real-world drag'n'drop – bidirectional camera-based media transfer between smartphones and large displays", in *Adj. Proceedings of the 10th International Conference on Pervasive Computing*, 2012.

[5] V. Mäkelä, H. Korhonen, J. Ojala, A. Järvi, K. Väänänen, R. Raisamo, and M. Turunen, "Investigating mid-air gestures and handhelds in motion tracked environments", in *Proc. PerDis'16*, 2016, pp. 45-51.

[6] S. Yoo, C. Parker, J. Kay, and M. Tomitsch, "To dwell or not to dwell: an evaluation of mid-air gestures for large information displays", in *Proc. OzCHI'15*, 2015, pp. 187-191.

[7] D. Vogel and R. Balakrishnan, “Distant freehand pointing and clicking on very large, high resolution displays”, in *Proc. UIST’05*, 2005, pp. 33-42.

[8] M. Liu, M. Nancel, and D. Vogel, “Gunslinger: subtle arms-down mid-air interaction”, in *Proc. UIST’15*, 2015, pp. 63-71.

[9] M. Nielsen, M. Störring, T.B. Moeslund, and E. Granum, “A procedure for developing intuitive and ergonomic gesture interfaces for HCI”, in *Gesture-Based Communication in Human-Computer Interaction*, vol. 2915, Springer Berlin Heidelberg, 2003, pp. 409-420.

[10] V. Mäkelä, T. Heimonen, M. Luhtala, and M. Turunen, ”Information wall: evaluation of a gesture-controlled public display”, in *Proc. MUM’14*, 2014, pp. 228-231.

[11] C. Ackad, A. Clayphan, M. Tomitsch, and J. Kay, “An in-the-wild study of learning mid-air gestures to browse hierarchical information at a large interactive public display”, in *Proc. UbiComp’15*, 2015, pp. 1227-1238.

[12] J.M. James, “SimSense - Gestural Interaction Design for Information Exchange between Large Public Displays and Personal Mobile Devices”, M.S. thesis, School of Information Sciences, University of Tampere, Tampere, Finland, 2015.

[13] M. Turunen, J. Hakulinen, A. Melto, T. Heimonen, T. Laivo, and J. Hella, ”SUXES – user experience evaluation method for spoken and multimodal interaction”, in *Proc. ISCA*’09*,* 2009, pp. 109-112.

[14] J.D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani, “Consumed endurance: a metric to quantify arm fatigue of mid-air interactions”, in *Proc. CHI’14*, 2014, pp. 1063-1072.

Details of the dissertations are available at
http://www.uta.fi/sis/tauchi/dissertations.html.

1. **Timo Partala:** Affective Information in Human-Computer Interaction
2. **Mika Käki:** Enhancing Web Search Result Access with Automatic Categorization
3. **Anne Aula:** Studying User Strategies and Characteristics for Developing Web Search Interfaces
4. **Aulikki Hyrskykari:** Eyes in Attentive Interfaces: Experiences from Creating iDict, a Gaze-Aware Reading Aid
5. **Johanna Höysniemi:** Design and Evaluation of Physically Interactive Games
6. **Jaakko Hakulinen:** Software Tutoring in Speech User Interfaces
7. **Harri Siirtola:** Interactive Visualization of Multidimensional Data
8. **Erno Mäkinen:** Face Analysis Techniques for Human-Computer Interaction
9. **Oleg Špakov:** iComponent – Device-Independent Platform for Analyzing Eye Movement Data and Developing Eye-Based Applications
10. **Yulia Gizatdinova:** Automatic Detection of Face and Facial Features from Images of Neutral and Expressive Faces
11. **Päivi Majaranta:** Text Entry by Eye Gaze
12. **Ying Liu:** Chinese Text Entry with Mobile Phones
13. **Toni Vanhala:** Towards Computer-Assisted Regulation of Emotions
14. **Tomi Heimonen:** Design and Evaluation of User Interfaces for Mobile Web Search
15. **Mirja Ilves:** Human Responses to Machine-Generated Speech with Emotional Content
16. **Outi Tuisku:** Face Interface
17. **Juha Leino:** User Factors in Recommender Systems: Case Studies in e-Commerce, News Recommending, and e-Learning
18. **Joel S. Mtebe:** Acceptance and Use of eLearning Solutions in Higher Education in East Africa
19. **Jussi Rantala:** Spatial Touch in Presenting Information with Mobile Devices
20. **Katri Salminen:** Emotional Responses to Friction-based, Vibrotactile, and Thermal Stimuli
21. **Selina Sharmin:** Eye Movements in Reading of Dynamic On-screen Text in Various Presentation Formats and Contexts
22. **Tuuli Keskinen:** Evaluating the User Experience of Interactive Systems in Challenging Circumstances
23. **Ogunbase Adewunmi Obafemi:** Pedagogical Design and Pedagogocal Usability of Web-Based Learning Environments: Comparative Cultural Implications from Africa and Europe
24. **Hannu Korhonen:** Evaluating Playability of Mobile Games with the Expert Review Method
25. **Jani Lylykangas:** Regulating Human Behavior with Vibrotactile Stimulation
26. **Ahmed Farooq:** Developing technologies to provide haptic feedback for surface based interaction in mobile devices
27. **Ville Mäkelä:** Design, Deployment, and Evaluation of Gesture-Controlled Displays in Ubiquitous Environments