A Laboratory-based Study Methodology to Investigate Attraction Power of Large Public Interactive Displays

Victor Cheung, Stacey D. Scott
Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada
{victor.cheung|stacey.scott}@uwaterloo.ca

ABSTRACT
A known challenge of designing large public interactive displays is to create an interface that attracts a passerby's attention and communicates its interactivity. However, typical "in-the-wild" field study methods of assessing public display design solutions require costly system implementation and deployment, creating challenges for assessing early stage design concepts. Such studies also limit the amount of experimental control researchers have over the environment, limiting the precision of results. To address these issues, we developed a complementary laboratory-based study methodology that employs experimental deception to assess the ability of an interface design solution to attract a passerby's attention. Our methodology enables more rigorous control of confounding factors, study of early-stage prototypes, and requires minimal setup. We used this methodology to assess existing visual design solutions for drawing attention and enticing interaction, compare our results to previous studies, and reflect on the benefits and limitations of this assessment approach.

Author Keywords
Large interactive displays; public space; laboratory study; experimental design.

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

INTRODUCTION
Large interactive displays are becoming increasingly popular for content presentation due to their dynamic and versatile presence. They now appear in various public venues such as airports, information centres, and subway stations, providing up-to-date, relevant, and interactive content to large groups of people. However, recent studies have shown that these displays are still under-utilized, or even being overlooked or ignored [20,30].

UbiComp '15, September 7 - 11, 2015, Osaka, Japan.

ACM 978-1-4503-3574-4/15/09…$15.00.
DOI: http://dx.doi.org/10.1145/2750858.2805842

To understand this phenomenon, and to evaluate potential design strategies, a number of "in-the-wild" field studies have been conducted. These studies typically involve deploying an interactive display to a target environment, while researchers unobtrusively observe passersby's behaviour and reactions to the display and its interface, and document the interaction using computer log and video/audio recordings (e.g., [10,18,19,21,24,33,34]).

While these studies have uncovered a number of key insights on the design of such public systems, they, like all field studies, are limited to directly observable behavior. Such observations often do not reveal users' intentions or rationale for their behaviours, even with the help of interaction logging and video recording. Furthermore, studying a system in the field necessitates building a robust, fully working system, and in many countries one that fulfills various safety, privacy, and legal regulations. This issue introduces significant barriers to assessing interface design concepts early in the design process.

Laboratory-based studies enable assessment of interface design concepts at various levels of fidelity [36], and provide a higher level of precision of study results [25]. Moreover, they enable a greater variety of data to be collected, including subjective and explanatory feedback that help elucidate people's intentions for observed behaviour. Yet, typical human-computer interaction (HCI) laboratory study methodologies assess interface design concepts by asking participants to complete structured or semi-structured tasks with a design prototype, or by eliciting user opinions on specific design aspects. Due to the nature of these experimental tasks, participants are necessarily aware of the existence, and often the purpose, of the interface designs. Yet, interactive systems situated in public settings commonly must first attract and entice potential users to approach and then use the system. Thus, traditional laboratory-based approaches do not accurately emulate the reality of public interactive systems, especially for the early stages of the engagement process with potential users.

To address this issue we developed a novel laboratory study methodology designed to assess the ability of a large interactive display to attract and entice potential users. The novel contributions of our methodology include:

- a deception task explicitly designed to emulate typical distractions in a public setting that may compete with the display for the user's attention,

- design criteria for architecting deception tasks for this purpose, and
- a carefully constructed process for revealing the employed experimental deception to help elicit relevant participant feedback on tested display design concepts.

This paper describes the iterative design of our deception methodology, including identification of specific design criteria for the deception task. We apply our methodology to a comparative laboratory-based study of design concepts that have previously been shown to be effective at capturing people's attention, in the psychology literature (e.g., content motion [7]) and in field studies of deployed large interactive displays (e.g., displaying a passerby's shadow [29]). Our study findings are consistent with reported findings from field studies of these design concepts, suggesting that our methodology effectively emulates relevant in-the-wild conditions for assessing the attraction power of interface concepts for large interactive displays. Our data also provide additional insights, such as comparative data of the effectiveness of different designs, as well as direct feedback of user's perceptions of the display's interactivity.

To provide further context for this work, we first motivate the need to further understand how a passerby's attention is drawn to public displays. We then discuss the commonly used in-the-wild field study methodologies, followed by the design details of our proposed methodology. Then, we describe the application of our methodology to a laboratory study comparing several interface design concepts previously employed to attract potential users in public settings. Finally, we discuss lessons learned from this study, and potential extensions of our study methodology.

INTRIGUING PASSERSBY

Previous research has explored the interaction involved in large public interactive displays, resulting in the development of several models describing the process of gradually increasing the level of user engagement with the displays (e.g., [10,26,35,42]). This previous research highlights that interaction with such displays is often very different from the traditional human-computer interaction paradigm where a user typically has a goal in mind and uses the system to achieve such goal, without much distraction from the environment. In fact, users of public displays commonly do not even know of their existence until just before they begin using them, and thus often have no particular goal, or only loosely formed goals when using the system. Thus the interaction process with these displays begins before the user actively interacts with the system, and necessarily involves a window of time to attract users and present the display's utility. Therefore, interaction models for public displays typically begin with a "passerby" stage where a potential user is just passing-by possibly without any prior knowledge of the display. If engaged, this person would then transition into being a user exploring the display.

However, most models treat this transition as a one-step process and ignore the intricacy within this potentially short

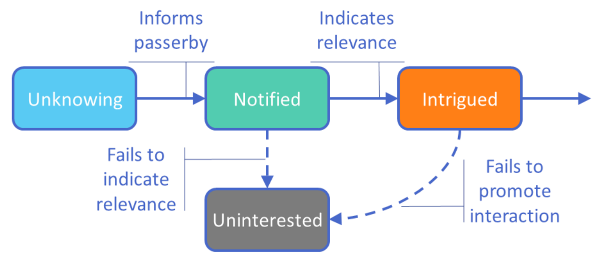

Figure 1. Modelling the early stages of interaction with large public interactive displays from the passerby's point of view. This happens before any active interaction occurs.

but crucial transition step. For example, will the passerby always be intrigued by the content presented? How does the passerby know that the display is responsive, and if so, how can they trigger a response? Inspired by related work in museums focused on the "attraction power" of exhibits (defined as the ability to grasp the attention of a museum visitor [8]), we expanded this step of intriguing passersby into several states and transitions, as illustrated in Figure 1.

In our model of the early stages of the interaction process, a passerby begins in an *Unknowing* state where the display is simply part of the surroundings. The passerby will only become *Notified* if the display captures their attention. The passerby can then either become *Uninterested* if their attention cannot be maintained, or *Intrigued* if their interest is aroused. From this state onward, the display should communicate interactivity, and only then will the passerby become a user and begin to actively interact with the display. This decomposition of the initial stages of the interaction process with public displays helps to suggest places where design can be applied to transition the passerby into a system user, and implies different design strategies depending on the state from which the passerby is transitioning.

Many researchers suggest exploiting the "honey-pot effect", where people are commonly attracted when they see others using the display (e.g., [10,19,21,24,27,34]). This effect has been consistently observed as a strong driving factor for attracting users in many contexts. Yet, this effect only occurs after one or more passersby are initially attracted to the display. As aforementioned, many deployed public displays fail to attract any users for prolonged periods of time [20,30]. To date, no systematic design approach for attracting users is known. Moreover, many displays offer features that may benefit single users also; thus, developing better design solutions for attracting even singleton users is needed to more fully exploit the potential benefit of deployed public displays. We aim to provide a way to rigorous assess the effectiveness of existing and emerging design approaches, and to provide insights on their applications.

RELATED WORK

This section first overviews common methodologies for studying large interactive displays, and then describes the use of deception in human experiments, particularly when investigating participant's attention in response to various types of stimulus.

Recent work by Alt et al. [3] comprehensively overviewed existing study types, paradigms, and methods for evaluating public displays. In this paper, we focus on *paradigms in evaluating a prototype*, and further elaborate the steps involved and trade-offs of each paradigm.

"In-the-Wild" Field Studies

Due to the public nature of the interactive displays, researchers often conduct an in-the-wild field study to investigate passersby's natural behaviour in response to the displays. This methodology, based on an ethnographic approach [6], involves deployment of the displays in the target location with the interface designs under study incorporated into a complete working application running on the system. Researchers then observe and record (via field notes or video/audio capture, and/or system interaction log) how the displays are used, and typically do not interfere during the interaction process. On-site voluntary interviews/surveys may be conducted with an arbitrarily number of people who have interacted with the displays. For example, Peltonen et al. [34] deployed an interactive public display at a store front in a central city location, allowing anyone who passed by to interact. System use was recorded for a month in the form of interaction log, as well as video and audio feed from a web camera. A small number (12) of on-site interviews were conducted to collect immediate user feedback. The video data were then used as the primary content for interaction analysis. Other researchers have used a combination of field notes, video/audio recordings, interaction log, and interviews/surveys to analyse usage pattern, which helped elicit design requirements and identify usability issues (e.g., [10,18,19,21,24]).

The duration of field studies can range from a few hours [10] to days [18] or even years [33], resulting in various context-specific findings. In a paper describing a three-year long-term field study of multipurpose display deployment, Ojala et al. [33] discovered a difference between public's stated information needs and their actual information behaviour. Based on this they argued that there was a need for such longitudinal study for gaining more in-depth knowledge about the real-world use of public interactive displays. However, besides considerable time commitment, as pointed out by McGrath on choosing a research method [25], such field studies gain realism at the price of low generalizability and lack of precision. Such trade-off is inevitable because, to make a direct observation, the displays have to be in a designated location, and the setup cannot be disturbed or controlled during the process.

Moreover, video/audio recordings, while being a major source of data, might not be always easily available. For example, environmental factors (e.g., sunlight, noise, venue constraints) may hinder proper recordings. Also, due to privacy laws, video collection may require prior informed, often written, consent, which is often difficult to collect from passersby and may hinder study participation [19].

Nevertheless, there is no doubt that the naturalness of people's response to the displays is beneficial to the understanding of the interaction process. In designing our methodology we aimed to simulate key aspects of the public context (e.g., creating an environment that induces the "Unknowing" state) by appropriating the use of a deception task, as well as the setup of the experimental venue. We also incorporated the role of silent observer by not interfering during the interaction process, as opposed to experimental techniques that might gain more details on awareness at the expense of interrupting participants [17].

"In-the-Wild" Field Experiments

Due to the lack of precision in the field study methodology, some researchers have introduced controls to the design concepts under investigation during field deployment using a field experiment[1] methodology [25]. This is typically achieved by running different versions of the deployed application at different times or in different locations. For example, Seto et al. [39] investigated different menu invocation designs aimed to promote menu discoverability on a public digital tabletop, using various interface elements and animations. The authors deployed the display in a museum and swapped between four alternative interface design approaches throughout each day of the study. System use was documented through field notes, computer logs, and video recordings. Similarly, Kukka et al. [22] investigated mechanisms for enticing interaction on public displays by developing eight versions of the same application with different visual signals, which were then deployed on eight interactive public displays. Apart from unobtrusive observation, interaction logs, and semi-structured interviews, the authors also collected demographic information and feedback by displaying a questionnaire on the displays upon touches.

Alternating between different versions of an application allowed researchers to control the design features under study, thus allowing for comparative assessment. However, it is possible that the same set of people will be exposed to multiple versions throughout the study, thereby creating a carryover effect. One method is to conduct the study for a longer period of time to have an adequately distinctive set of people exposed to each version. Yet, this approach requires a greater time commitment. To avoid potential carryover effect, our methodology employs a between-subject experimental design to ensure that each participant is only exposed to one design alternative. In addition, we incorporate the use of questionnaire and semi-structured interviews for our participants to gain further insights on their experience with the displays.

---

[1] Regarding terminology, we use McGrath's [25] distinction between *field study* and *field experiment*, where both refer to "in-the-wild" observational studies, but the latter indicates that some control or intervention is made in the environment being studied (e.g., changing the stimulus being studied during the observation period). Using Alt et al.'s [3] terminology, both would be called *field studies*.

Laboratory Study for Public Interactive Displays

In a laboratory study, researchers gain precision by being able to control variables in the environment, thus allowing more rigorous qualitative and quantitative analyses of study data. This is typically done by having the study setup and conducted inside a laboratory, with representative users recruited as participants, who will complete a consistent set of procedures. Under modern ethics protocol, these procedures commonly include learning the purpose of the study, carrying out certain tasks while being recorded, and providing feedback about the task. In addition to improved precision, a laboratory setup poses less demand on application robustness as it is used under controlled conditions, and setup effort, allowing researchers to quickly prototype a system or part of a system for evaluation. For example, Vogel and Balakrishnan [42] developed a prototype system based on their proposed interaction framework for interactive public displays, and conducted an informal user evaluation in a controlled laboratory environment. To explore research questions such as "What techniques could be used to notify and communicate with users in a minimally intrusive, socially acceptable manner?" (p. 137), their participants were deliberately asked to explore the display without any instructions given. Based on the user feedback and direct observations, the authors were able to evaluate the effectiveness of their design solutions and establish future directions. Notably, the authors highlighted that, despite the fact that the system was not technologically feasible for deployment in a real-world study due to hardware limitations, evaluating a prototype solution in a laboratory setting allowed them to evaluate and refine their designs.

Laboratory studies allow researchers to quickly prototype a system, recruit participants, validate interaction models and design approaches, or explore possibilities even before the technology is widely available. However, for many public display researchers, these studies are used as a "pre-study" for empirical measurements, and without much scrutiny in the implications from the results. For example, before carrying out a field study, Müller et al. [29] conducted a laboratory study determining the effectiveness of various visual representations of user embodiments displayed in the interface to indicate display interactivity. This method allowed them to test through eight conditions and quickly determine which representations were more effective. They however only focused on the time it took for each participant to determine whether they believed the display was reacting to their movements, and without any in-depth inquiry of the rationale behind their decisions. Similarly, Schmidt et al. [38] conducted a laboratory study first to validate their model of predicting legibility of content on large displays. The laboratory approach allowed them to have a within-subject study design and accurately compare three display techniques. Again, participants' impression of the content beyond being legible (e.g., does it communicate interactivity? what is the purpose of the display?) was not explored.

We believe that, with a careful and comprehensive design (e.g., questionnaires, structured interviews), a laboratory study has the potential to provide more in-depth, precise, and valuable insights for evaluating design solutions for public displays. Nonetheless, the gain in precision comes with reduced realism, which we address by using deception.

Use of Deception in Human Experiments

The need to use deception in a study typically arises from the necessity to "make sure that the research participants are not aware of what aspect of their psychology is being studied in what way" [9, p.260], and is carried out by withholding the true purpose of (or part of) the study from participants. While its morality has been substantially debated, deception does offer some benefits that are not only valued by researchers, but also by participants [12]. The codes of ethics of the American Psychological Association (APA) [4], the British Psychological Society (BPS) [11], and the Canadian Tri-Council Policy Statement (TCPS2) [41] permit the use of deceptive methods, provided that the experiments fulfil a number of criteria. The notable ones are to allow participants to withdraw from the study at any time, and to debrief them with all relevant information about the true nature of the experiment when it is sensible to do so.

Within Psychology for Attention and Perception

In the psychology domain, deception is often used as a means to retrieve unbiased data, including participants' attention and perception [31,40]. In the widely known "invisible gorilla" study [40], the true purpose of the experiment (unexpected appearance of visual stimuli) was withheld from participants, who were given an unrelated task (pay attention to a team and count). To ensure validity, participants who had heard of the experiment or phenomenon were replaced, and their corresponding results discarded after the first round of probing questions.

Our methodology borrows from the psychological study of attention by first withholding the true purpose of the experiment in the disguise of an unrelated task, and asking probing questions after the unrelated task is completed. However, instead of gradually revealing the existence of the visual stimulus (gorilla), our questions do not explicitly mention the type of visual stimuli used in the experiment to further elicit participants' perception towards them.

Within HCI for Behavioral Impacts

Within the HCI community, deception is also used for collecting unbiased responses. However, in HCI, deception is more typically used as a means to control the behaviour of one or more "group members" in a group task. This is done by the use of a "confederate"; that is, a member of the research team, or paid actor, playing the role of a study participant. The confederate typically follows a script or engages in predefined behaviour, unbeknownst to the real study participant(s) (e.g., [13,15]). Since public displays can be used by individuals as well as groups [34], confederate-based deception was not applicable for our purposes.

Deception has also been used in recent HCI studies of large displays. Beyer et al. [5] used various "distractor" displays in addition to the displays under study, and gave participants little to no task instructions other than to explore the spaces containing the displays and then complete several questionnaires. Alt et al. [2] deceived participants arriving for another (real) study by having them to wait alone, under no instructions, in a hallway (containing the display) under the "cover story" that the experiment room needed some final preparations. After several minutes, participants were led into an adjacent room and asked to participate in the display study by completing a questionnaire. Afterwards, participants were then led to the other study.

The deception strategies used in these studies were effective in withholding the specific purpose of the study from participants, allowing researchers to gather relatively unbiased feedback. However, since participants were given little to no information about how to behave around the displays under study, confounds may have arisen that influenced results. For instance, some participants waiting in the hallway may have checked emails on their mobile phones, potentially affecting their exposure to and perceptions of the display. A desirable benefit of confederate-based deception is the ability to create consistent scenarios that help experimental precision. We believe this benefit can also be achieved by employing a "fake" experimental task, framed under a relevant setup in which participants are led to believe it is the primary (and only) purpose of the study.

In the next section we describe the general procedures of the initial experimental design, followed by the specific deception task and steps we used in a pilot study and its subsequent improved version.

INITIAL EXPERIMENTAL DESIGN

The main objective of the laboratory study was to examine the effectiveness of potential visual design strategies at the early stages of interaction with public displays, namely, drawing attention and communicating interactivity, in order to transition from the *Unknowing* to *Notified* to *Intrigued* stages shown in Figure 1. While there are various suggested design approaches to facilitate these stages, we began with the visual saliency, as it is a prominent aspect in interface design for public displays.

		1-factor within-subject design for deception study	
		Deception task (Condition A)	Deception task (Condition B)
2-factor between-subject design for actual study	Control	Condition 1A	Condition 1B
	Factor 1	Condition 2A	Condition 2B
	Factor 2	Condition 3A	Condition 3B
	Factor 1&2	Condition 4A	Condition 4B

Table 1. Mixed study design of the experiment. The deception task is set as a counter-balanced 1-factor within-subject study. In actuality it is a 2-factor between-subject study.

Saliency and Attention Capture

Modern views of attention capture in psychology describe attention as the first step in perception, which then leads to cognition, and can be driven simultaneously by bottom-up (low-level stimuli such as motion, contrast) and top-down (high-level stimuli such as goals, intentions) processing of the perceiver [7]. These mechanisms can be initialized externally and therefore be incorporated into interface design.

Drawing on previous literature, we focused on visual stimuli (e.g., motion, contrast) as the bottom-up experimental factors, which are more generic and hence more applicable in a wider range of contexts. Context-relevant content (e.g., inspirational quotes, images of known places) was used as a simple adoption of top-down approaches.

Experimental Design

Our laboratory study methodology employs a deception task that omits any mention of the display, and intends to create a consistent public space scenario across participants.

Study Format

To the participants, the study appears to be a one-factor within-subjects experiment, where each participant completes two counter-balanced conditions. In reality, each participant is assigned to only one of the four actual study conditions within a two-factor, between-subject design, as exposure to multiple conditions may take too long and raise suspicions about the true nature of the study. This approach also avoids learning effects across conditions. This mixed-study format is shown in Table 1.

Deception Task and Setup

In order to simulate the desired scenario and keep the participants from discovering the deception prematurely, we established four criteria of choosing the deception task:

1. Being a believable task which participants are already familiar with, and have experience in carrying out in a public setting,
2. Requires a certain degree of attention to compete with the large display,
3. Includes a movement component to simulate passing-by behaviour, and
4. Participants should have opportunities to look around in the environment by having the large display in their line of sight occasionally.

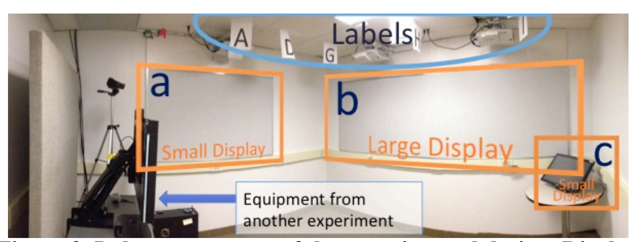

Figure 2. Laboratory setup of the experimental design. Display 'b' was the large display, displays 'a' & 'c' were the additional small displays for distraction. Labels are attached to the ceiling to navigate participants. A camcorder is installed at the left corner to record the deception task.

Such deception task has parallels with "distraction" or "secondary" tasks used in HCI studies focused on attention and interruptions (e.g., [1,14,37]). Indeed, the goal of our deception task is to emulate the type of natural distractions that may be present in a realistic public display setting. Yet methodologically, our use of the deception task has several key differences to previous uses of distraction tasks. First, in previous studies participants were always informed of both the primary and distraction tasks. Second, in those studies participants were required to actively engage with the distraction task. The goal of our deception task is to make participants believe that the deception task is the primary, and only, task they are engaging with during the study. From the experimenter's perspective, the deception task serves as a typical distraction that may exist in a real world setting, and may prevent the large display from capturing (and ultimately holding) the user's attention.

As shown in Figure 2, the experimental room is set up with a 94.2x301.5cm large display (2b) and additional smaller displays (2a, c) next to it, designed to be distractions. The labels on the ceiling are used to facilitate the deception task, which models participants moving about an open space. These labels help the participants navigate without referring to the displays as landmarks. The middle of the room is cleared to emulate the open space in a public setting. A camcorder is installed to record participants' movements.

The Deception Phase

The study begins with the deception phase where each participant undergoes the deception task, while in reality the researcher observes their responses to the display.

What the participant does

Each participant is first briefed with the deception task, including the fabricated motivation and procedures. After filling in a demographic questionnaire (designed to reinforce the deception task), they are led to the experimental room with the displays already running the application. The room is described as a multi-use research space with several on-going studies setup so the participant would feel comfortable with the surrounding displays, and more importantly, not question their presence. At no point are these displays introduced or set up in front of the participant. The participant then completes the deception task under the two counter-balanced conditions (A and B in Table 1).

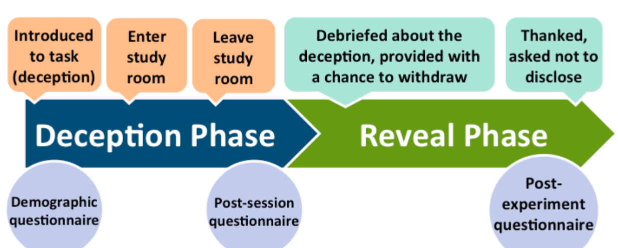

Figure 3. Procedures of the initial experimental design, which consist of a "Deception Phase" and a "Reveal Phase". Above are the steps undertaken by the participant, below are the questionnaires given to the participant at different phases.

What the researcher does

Before the study, the researcher assigns one of the four actual conditions to the participant, and sets up the (large) display accordingly (the small displays always have the same setup as a control). During the deception task, the researcher provides instructions related to the task, while observing how the participant responds to the factor(s) being manipulated. The displays are never mentioned.

Once the task is finished, the participant is led out of the experimental room and asked to complete a post-session questionnaire. The questionnaire is designed to first ask questions closely related to the deception task, and then begins to probe for impressions of the surroundings. These questions are designed based on the transitions described in the model shown in Figure 1, for example, whether the participant noticed the content on the displays, and what they think they can do with the displays.

The Reveal Phase

After the post-session questionnaire is completed, the researcher asks the participant probing questions to see if they have any suspicions about the study. Then the researcher reveals its real purpose, and explains why this purpose is not provided in the beginning. Here the participant can decide if they wish to withdraw from the study, or wish the researcher to proceed with the remainder of the study. They are also provided with contacts of counselling services if they feel uncomfortable in any way. These steps are required as per our university's research ethics policies on use of deception in research studies.

Finally, a second post-experiment questionnaire with more elaborate questions about the displays is given to the participant. Upon finishing, the participant is thanked, paid, and asked to not share their experience, or disclose the real purpose of the study to anyone.

Figure 3 summarizes the overall procedures of the initial experimental design, including the steps undertaken by each participant and the questionnaires they complete at different phases of the study. Listed below is an overview of the questions used in the questionnaires:

Demographic Questionnaire – Gather background information, with relevant questions to reinforce the deception.
Post-session Questionnaire – Inquire impression of carrying out the deception task, with both close- and open-ended questions on whether the participant noticed anything on the displays, and how they thought about the displays.
Post-experiment Questionnaire – Inquire impression of the display's interactivity, and how attention was drawn.

In the next section we describe the pilot study we conducted utilizing these procedures, and some key findings.

PILOT STUDY: MOBILE APP USE

As our pilot study, the deception task was advertised under the disguise of a study titled "Usage Pattern with On-the-go Mobile Applications", aiming at investigating how app (e.g., calendar, clock, settings) usage was affected when a person was walking, and was described as follow:

> *"…an observational study on how different on-the-go mobile scenarios affect the usage patterns of typical tasks (e.g. looking up information, check status updates) on one's own portable device (e.g. a smartphone, a tablet), particularly when the user is moving as opposed to remaining stationary."*

The two counter-balanced conditions were *with* or *without* using apps on the participant's mobile device. During each condition the participant was asked to walk between labeled points in the experimental room, and in the first condition the participant was asked to use a number of apps during the walking sequence. We chose this task because it satisfied all the four criteria we established: 1) use of mobile app in a public setting is commonly seen nowadays, 2) using an app requires attention to the device, 3) participants were asked to walk between labels, and 4) walking instructions, in the form of destination label points, were provided as the task proceeded. The destination label points (see Figure 4) were selected so that the routes involved covered all the possible directions a passerby could take in relation to the display (i.e., away, towards, across). As for the application running on the large display, we prototyped an interface containing many digital Post-it notes with inspirational quotes. The Post-it notes were in constant movement (similar to a motion animation technique used by Seto et al. [39]) and "bounced" off the edges at a slightly randomized angle.

In reality, we were interested in examining the effectiveness of animating the saliency based on user's proximity towards the large display. For the two-factor design we chose two visual attributes: speed and contrast, as they were generic enough to be applicable to a wide range of content. The proximity dependence was inspired by the proxemics model proposed by Marquardt and Greenberg [23], where interaction mechanisms adapt to the proximity of the user to the system. In our study we only used the first three of the four proxemics distance zones in the order of decreasing proximity (*Public*, *Social*, *Personal*, *Intimate*) due to their relevance to the early stages of interaction. For the adaptive speed condition, the speed of the Post-it notes decreased in a stepwise fashion as the participant walked closer to the large display, with the intent of making the content easier to see and interact with at slower speeds. Similarly, for the adaptive contrast condition, the contrast of the Post-it notes increased in a stepwise fashion. To highlight the change, the large display was divided horizontally into three even grids so the change will only appear in the grid closest to the participant. Table 2 summarizes the deception task and the conditions used in the pilot study.

		Mobile App Usage Study (Deception study)	
		Using apps (Condition A)	*Not using apps (Condition B)*
Study on effectiveness of animating saliency (Actual study)	*Control*	Condition 1A	Condition 1B
	Adaptive speed	Condition 2A	Condition 2B
	Adaptive contrast	Condition 3A	Condition 3B
	Adaptive speed & contrast	Condition 4A	Condition 4B

Table 2. Pilot study design, with the deception task being walking with or without using apps on a mobile device.

Key Findings

In the pilot study 16 participants (10 males, 6 females) were recruited from our university. All completed both the deception and reveal phases in less than an hour. No significant results were found in terms of the effectiveness of animated saliency. We believed this was due to the low number of participants (four for each condition), and the saliency being too subtle to notice, as reported by the participants. Nevertheless, based on their feedback and our observation, we have summarized some key findings:

[KF1]: None of the participants suspected the undertaken task was a deception task. Some even believed that the displays were used to simulate traffic or roadside buildings.

[KF2]: After being explained to with the motivation behind the study, no participant expressed concern about the deception involved.

[KF3]: After being led out of the experimental room to answer the questionnaires, most participants were able to recall the setup, elaborate answers, and provide suggestions.

[KF4]: While believable, the deception task of using mobile apps required too much attention from participants during the walking sequence, providing little opportunity to look around the environment. The labels were also too straightforward, allowing participants to glance quickly and walk without looking.

[KF5]: The Post-it notes drew little attention beyond being recognized with some quotes on them, especially when compared to the close-by control display showing photos of the university as the main content.

[KF6]: The adaptive speed change (slowing down when the user was close) and contrast change (higher contrast when the user was close) were generally too subtle to notice and, thus did not evoke a feeling of interactivity.

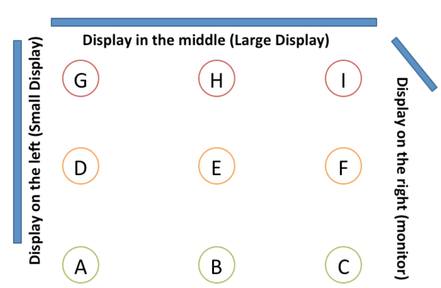

Figure 4. Floor plan of the experimental room. The labels were attached to the ceiling to not interfere with movements and made the participant to look up for the next destination, thereby having the displays in their line of sight occasionally.

Figure 5. Laboratory setup for the improved study design. Labels were attached to the walls to better simulate signage, with a grid layout for a less straight-forward navigation. A Microsoft Kinect was used at the back of the room to capture the silhouette of the participant and measure depth.

Interestingly, when asked about the perceived interactivity, most participants did not consider any of the displays to be interactive, and a few thought the touchscreen monitor (display 'c' in Figure 2), which displayed the same content as the large display, was interactive. This might be due to their familiarity with prevalent consumer touchscreen monitors.

MAIN STUDY: MOBILE NAVIGATION

The main goal of the pilot study was to provide a proof-of-concept to the proposed methodology in terms of the setup and procedures, allowing us to evaluate its feasibility and aspects for improvement. It served its purpose by providing some key findings, which we used as guidelines in re-examining and improving the study design.

The Improved Deception Task

To address KF4, we replaced the Mobile App Use task with a Navigation task using a mobile device, by combining the walking instructions (Condition A: visually on mobile device, Condition B: verbally by researcher) with the task itself, and focused on only one objective (navigate). The task was described in the study recruitment notices as:

> *"…an observational study on how different forms of navigating instructions affect the way a person arrives at a destination …to walk between several marked points while being provided with navigating instructions verbally and/or visually."*

To further address the issue of routes being too straight-forward in KF4, the labels were instead presented in a grid format (A-B-C in one direction, 1-2-3 in the other), as shown in Figure 5. Instructions were given to the participants as either one-step routes (e.g., from C1 to A1) or two-step routes (e.g., from A1 to A3, then A3 to C3). These changes still fulfilled all four established criteria and increased the likelihood of seeing the content on the displays.

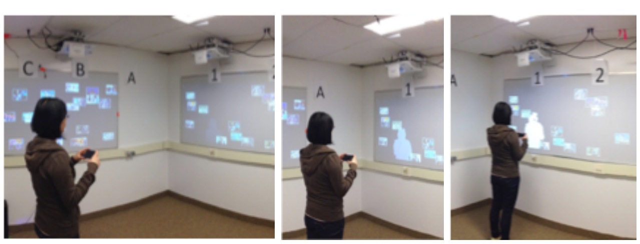

Figure 6. Use of shadow in the Adaptive shadow condition. Contrast of shadow became higher as participant walked closer to the large display.

		Navigation Study (Deception study)	
		Visual (Condition A)	*Verbal (Condition B)*
Study on effectiveness of animating saliency (Actual study)	*Adaptive speed*	Condition 1A	Condition 1B
	Adaptive speed & shadow	Condition 2A	Condition 2B
	Adaptive trajectory & shadow	Condition 3A	Condition 3B

Table 3. Main study design, with the deception task being walking with the navigation instructions provided visually or verbally. The small display was used as a control in all conditions, thus reducing a condition in the actual study.

The Improved Experimental Design

To address KF6, we decided to only keep the adaptive content speed from the pilot study as a factor and compared it to a change in trajectory. We also incorporated a shadow of the participant, which was similar to the silhouette used by Müller et al. [28] to increase the attraction power of the display, and extended it with adaptive contrast in relation to their proximity (see Figure 6). To address KF5 and increase the attractiveness of the content shown on the large display, both displays showed a unified set of photos of the university, with the small display as the control condition at all times (content floating and bouncing off edges at constant speed). This modification allowed us to eliminate the control condition, thus reducing the between-subject conditions from four to three (see Table 3). The touchscreen monitor was removed to avoid additional distraction.

The same procedures from the pilot study were used, except the questionnaires were modified to reflect the new deception task, and added one question in the post-session questionnaire asking specifically if the participant noticed whether the display responded to them (and in what way). A condition-specific questionnaire was also added after the post-experiment questionnaire to further elicit any feedback or comments about the displays[2]. All the questionnaires were still presented outside the experimental room so participants had to recall what they experienced.

MAIN STUDY FINDINGS

We present two types of findings from our main study: effectiveness of the methodology as a means to evaluate design concepts for early stages of public display interaction, and effectiveness of the design concepts themselves.

Methodology Effectiveness

Like the pilot study, the use of the deception task successfully kept the true purpose of our study hidden from the participants. None of the participants showed suspicion during the deception phase. Some again offered suggestions on how the navigation mechanism (the task they were given

[2] The full details of the questionnaires used in this study are provided in the electronic supplemental appendix.

during the deception phase) could be improved. This also indicates that they indeed started off as an *Unknowing* user, as they were not told of the displays during the deception phase. Also, the task appeared to be a favoured one, which is important as having enough participants is crucial to the success of the study. Within one week of sending out recruitment emails, we were able to schedule 30 participants (21 males, 9 females) for a two-week study.

As expected, the entire study took relatively less time and effort to setup compared to a full field deployment. First, the administrative effort of gaining access and permission to setup the displays in a public setting was not necessary. Second, the physical effort of relocating our display hardware was not necessary. This is a significant benefit, as large-format displays, such as the front projected 315.9cm-diagonal display used in our studies, are non-trivial to safely and securely install in a public setting. As the large display equipment was stationary and already setup in our laboratory, safety, privacy, and legal concerns were minimized. These minimal concerns also led to a shorter turnaround time for institutional ethics approval than our previous experiences gaining approval for large display field studies. Finally, our methodology allowed us to focus specifically on an early stage of the interaction process, eliminating the need for a completely robust and working system, as would be expected by users in a public setting.

As typical of a laboratory setting, we had significant control over the study. We were able to have multiple sources of data, including researcher notes, interaction logs, video/audio recordings, and questionnaires, and how they were collected. We were also able to control which condition to experiment with, and apply the same scenarios repeatedly.

The following section also provides evidence that this methodology produces useful insights on potential design solutions for attracting and enticing a passerby's attention.

Design Concept Effectiveness

We present some findings from our data analysis to demonstrate how our proposed methodology helped to evaluate the effectiveness of the tested interface designs in the early stages of interaction, and illustrate its benefits and limitations. In-depth data analysis of all tested designs, while consistent with the results reported here, is beyond the scope of this paper and thus not included.

Confirming Shadow's Effectiveness in Attention Capturing

In our main study, we adapted the use of shadows described in Müller et al.'s interface design for a public interactive display [29] by showing a shadow reciprocating movements of the participant, and relating its contrast to the participant's proximity to the display. Similar to Müller et al.'s findings, our results revealed the majority of participants (18 out of 20 who experienced adaptive shadow conditions) reported noticing a shadow, and more importantly, noticed that it was their own shadow. In addition, six participants explicitly commented on the appearance of their shadows when they "approached" or "were close to the display".

To understand the effectiveness of the shadow design, we conducted a one-way analysis of variance (ANOVA) test across all conditions on the Likert-scale ratings for the question "Please indicate how much [the large] display drew your attention when the instructions were provided vocally"[3]. Then we conducted a post-hoc comparison of Conditions 1B and 2B, representing the adaptive speed condition and adaptive speed plus shadow condition, both in the verbal case (we were primarily interested in the verbal case as participants had more opportunity to see the large display content in these trials). The ANOVA test revealed a main effect of condition, $F(2,27)=3.65$, $MSE=2.89$, $p<.05$, while the post-hoc comparison using Tukey HSD tests revealed the shadow condition (Condition 2B, $M=3.5$, $MSE=.58$) was significantly more effective at capturing participants' attention than the adaptive speed condition (Condition 1B, $M=1.5$, $MSE=.58$), $p<.05$.

In addition, when asked if they thought the display was interactive, 17 out of 20 participants, who were presented with the shadow (Conditions 2 and 3), reported they felt the display was "somewhat interactive" or above, mostly due to movement of the shadow. This finding is again consistent with Müller et al.'s [29] study results, where a shadow was used to communicate interactivity. However, this understanding might not extend to other forms of interactivity such as touch gestures towards contents (only three answered that they thought they could move the photos).

The consistency in the results suggests that our methodology was able to evaluate the visual designs above in a lab environment, while providing a more in-depth understanding of users' behaviour as study participants.

Assessing Potential Attention Capturing Concepts

Another design concept we investigated was reducing content's movement speed as the participant was close to the display (adaptive speed). Albeit reported as an effective saliency property for attention capturing in psychology [7], our observation and participant responses did not reflect this. When asked if they noticed any speed change on the large display, only a few participants responded positively (7 out of 30), with the majority not and some reported that they "did not notice" or even "did not pay attention".

We speculate this discrepancy was due to the way speed change was presented. In most psychology studies speed changes happened in participants' region of focus, whereas in our study participants were never asked to look at the display. This phenomenon of passersby spending a very short period of time looking at the display is supported by several field studies (e.g., [16,20]), and suggests that change of speed may by less effective at capturing some-

[3] Validity of using ANOVA on Likert-scale data was discussed in [32].

one's attention in this context, or that a more dramatic change is required to capitalize one's peripheral vision system in attention capturing on a public display.

As illustrated, the collected data enabled both quantitative and qualitative analyses of participant behaviour. This allowed comparative assessment of display concepts and provided insights on how these designs can be improved or iterated along the design process.

DISCUSSION AND FUTURE WORK

In this section we discuss our experience in designing the study, some considerations when using this methodology, and some directions in extending this work.

Lesson Learned

The deception task selection and the setup are crucial to the study design in terms of validity and believability. We went through numerous iterations to establish a deception task that simulated the public commuting scenario. The experimental room was portrayed as a "multi-purpose experimental room" so participants would not feel suspicious of the presence of seemingly unrelated apparatus.

In particular, the small display served two important purposes: it acted as a distractor to simulate a public space with multiple displays, and reduced the number of study conditions for between-subject tests (compare different design concepts), while allowing within-subject tests in each condition (compare a design concept with control).

The questionnaires were designed so that they not only did not reveal the deception, but also elicited impression towards the display. The use of multiple questionnaires at various stages of the study, together with a combination of close- and open-ended questions allowed us to have a much richer dataset to analyse. It also allowed participants to provide feedback and suggestions in a detailed manner. As mentioned earlier, we intentionally omitted mentioning the specific use of visual stimuli to elicit unbiased participant feedback on the design concepts. This omission, however, inevitably resulted in fewer targeted measurements such as performance and satisfaction for a given design concept.

Limitations

As discussed by Convertino et al. [13] and Müller et al. [29], laboratory-based methodologies typically fall short on ecological validity when compared to in-the-wild methodologies. To mitigate this issue we established a task selection criteria, and experimented with interface designs used in some in-the-wild studies. The consistency between our study findings and from those studies is indicative that our methodology is capable of reproducing and validating these findings. Yet it can still be improved by expanding the demographics of participants, arranging the experimental room to further emulate a public location (e.g., background noise), or using a large room.

Another limitation is the closed study environment, in which the participants tend to have less distractors other than those intentionally installed by the researchers, thus potentially more time to notice the display. To address this issue the deception task was used as an extra layer of distraction to demand both the time and attention of the participants. Also, the task routes were architected to cover all possible directions (albeit all in one session), yet to not be too repetitive (a walking session took 5-10 minutes), thereby created a closer approximation of multiple pathways and had a closer resemblance to a semi-open area like museums. This implies that to create other scenarios, for example, passing by only once, or waiting, requires a different deception task and/or experimental setup (e.g., [2,5]).

The main component of this methodology, the deception task, is necessitated by the assumption of *unknowing* participants and the *distractive* nature of the environment being emulated. Consequently, this methodology is tailored towards the early stages of interaction and less applicable to stages where participant's attention has been captured. Nevertheless, in light of the proliferation of personal devices, advancement of sensing and output technologies in interactive displays, we see ample opportunities to explore new interaction designs to better notify and entice passersby to an interactive display (e.g. [35,43]).

We would like to stress that we recognize the importance of and need for in-the-wild studies of public interactive displays [33], and do not intend to replace this form of study. Instead, we realized the need for a toolbox of techniques to evaluate design concepts for large interactive displays, especially in early stages of design conceptualization, or when constraints for public deployment, such as safety or privacy, require considerable time and effort. The main objective of this laboratory study methodology is to allow researchers and designers to experiment with their ideas, and narrow down promising design concepts before undergoing a more complete deployment and field study. The in-the-wild study methodology can then be used to further validate the design concepts.

CONCLUSION

In this paper, we presented a laboratory-based study methodology to investigate the effectiveness of design concepts in early stages of interaction with large public interactive displays, namely, capturing passersby's attention and intriguing them. Through the use of deception, we were able to simulate a public scenario inside an experimental room, as evidenced by the consistency in findings with previous in-the-wild studies. We have explained the criteria to design an effective deception task, and procedures to prepare for and conduct the study, which require significantly less time and effort to setup. Our methodology complements in-the-wild field studies, particularly for investigating early-stage design concepts, and is intended to enable researchers and practitioners to build large interactive displays that experience greater utilization during public deployment.

REFERENCES

[1] Adamczyk, P.D. and Bailey, B.P.If not now when?: the effects of interruption at different moments within task execution. *SIGCHI Conference on Human Factors in Computing Systems*, (2004), 271–278.

[2] Alt, F., Schneegass, S., Girgis, M., and Schmidt, A.Cognitive effects of interactive public display applications. *International Symposium on Pervasive Displays*, (2013), 13–18.

[3] Alt, F., Schneegass, S., Schmidt, A., Müller, J., and Memarovic, N.How to evaluate public displays. *International Symposium on Pervasive Displays*, ACM (2012), 1–6.

[4] American Psychological Association.Ethical principles of psychologists and code of conduct. 2010. http://www.apa.org/ethics/code/.

[5] Beyer, G., Alt, F., Müller, J., et al.Audience behavior around large interactive cylindrical screens. *SIGCHI Conference on Human Factors in Computing Systems*, (2011), 1021–1030.

[6] Blomberg, J., Giacomi, J., Mosher, A., and Swenton-Wall, P.Ethnographic field methods and their relation to design. In D. Schuler and A. Namioka, eds., *Participatory Design: Principles and Practices*. Lawrence Erlbaum Associates, 1993, 123–155.

[7] Bodenhausen, G. V and Hugenberg, K.Attention, perception, and social cognition. In F. Strack and J. Förste, eds., *Social Cognition: The Basis of Human Interaction*. Psychology Press, 2009, 1–22.

[8] Boisvert, D.L. and Slez, B.J.The relationship between visitor characteristics and learning-associated behaviors in a science museum discovery space. *Science Education 78*, 2 (1994), 137–148.

[9] Bortolotti, L. and Mameli, M.Deception in psychology: moral costs and benefits of unsought self-knowledge. *Accountability in Research: Policies and Quality Assurance 13*, 3 (2006), 259–275.

[10] Brignull, H. and Rogers, Y.Enticing people to interact with large public displays in public spaces. *IFIP International Conference on Human-Computer Interaction*, IOS Press (2003), 17–24.

[11] British Psychological Society.Code of ethics and conduct. 2009. http://www.bps.org.uk/what-we-do/ethics-standards/ethics-standards.

[12] Christensen, L.Deception in psychological research: when is its use justified? *Personality and Social Psychology Bulletin 14*, 4 (1988), 664–675.

[13] Convertino, G., Neale, D.C., Hobby, L., Carroll, J.M., Rosson, M.B., and Tech, V.A Laboratory Method for Studying Activity Awareness. *Nordic Conference on Human-Computer Interaction*, (2004), 313–322.

[14] Cutrell, E., Czerwinski, M., and Horvitz, E.Notification, Disruption, and Memory: Effects of Messaging Interruptions on Memory and Performance. *INTERACT and Conference on Human Factors in Computing Systems*, (2001), 263–269.

[15] Dabbish, L., Kraut, R., Patton, J., Heinz, H.J., and College, I.I.I.Communication and Commitment in an Online Game Team. *SIGCHI Conference on Human Factors in Computing Systems*, (2012), 879–888.

[16] Dalton, N.S., Collins, E., and Marshall, P.Display blindness? Looking again at the visibility of situated displays using eye tracking. *SIGCHI Conference on Human Factors in Computing Systems*, (2009), 3889–3898.

[17] Endsley, M.R.Direct measurement of situation awareness: validity and use of SAGAT. In M.R. Endsley and D.J. Garland, eds., *Situation Awareness and Measurement*. 2000, 131–157.

[18] Hinrichs, U. and Carpendale, S.Gestures in the wild: studying multi-touch gesture sequences on interactive tabletop exhibits. *SIGCHI Conference on Human Factors in Computing Systems*, ACM Press (2011), 3023–3032.

[19] Hornecker, E."I don't understand it either, but it is cool" - Visitor interactions with a multi-touch table in a museum. *IEEE International Workshop on Horizontal Interactive Human Computer Systems*, IEEE (2008), 113–120.

[20] Huang, E.M., Koster, A., and Borchers, J.Overcoming assumptions and uncovering practices: When does the public really look at public displays? *International Conference on Pervasive Computing*, Springer-Verlag (2008), 228–243.

[21] Jacucci, G., Morrison, A., Richard, G.T., et al.Worlds of information: designing for engagement at a public multi-touch display. *SIGCHI Conference on Human Factors in Computing Systems*, ACM Press (2010), 2267–2276.

[22] Kukka, H., Oja, H., Kostakos, V., Goncalves, J., and Ojala, T.What makes you click: exploring visual signals to entice interaction on public displays. *SIGCHI Conference on Human Factors in Computing Systems*, ACM Press (2013), 1699–1708.

[23] Marquardt, N. and Greenberg, S.Informing the Design of proxemic interactions. *IEEE Pervasive Computing 11*, 2 (2012), 14–23.

[24] Marshall, P., Morris, R., Rogers, Y., Kreitmayer, S., and Davies, M.Rethinking 'multi-user': an in-the-wild study of how groups approach a walk-up-and-use tabletop interface. *SIGCHI Conference on Human Factors in Computing Systems*, ACM Press (2011), 3033–3042.

[25] McGrath, J.E.*Groups: interaction and performance.* Prentice Hall, 1984.

[26] Michelis, D. and Müller, J.The Audience Funnel: Observations of Gesture Based Interaction With Multiple Large Displays in a City Center. *International Journal of Human-Computer Interaction 27*, 6 (2011), 562–579.

[27] Müller, J., Alt, F., Schmidt, A., and Michelis, D.Requirements and design space for interactive public displays. *Internation Conference on Multimedia*, ACM Press (2010), 1285–1294.

[28] Müller, J., Eberle, D., and Tollmar, K.Communiplay: a field study of a public display mediaspace. *SIGCHI Conference on Human Factors in Computing Systems*, ACM Press (2014), 1415–1424.

[29] Müller, J., Walter, R., Bailly, G., Nischt, M., and Alt, F.Looking glass: a field study on noticing interactivity of a shop window. *SIGCHI Conference on Human Factors in Computing Systems*, ACM Press (2012), 297–306.

[30] Müller, J., Wilmsmann, D., Exeler, J., et al.Display blindness: the effect of expectations on attention towards digital signage. *Internation Conference on Pervasive Computing*, Springer-Verlag (2009), 1–8.

[31] Neisser, U. and Becklen, R.Selective looking: attending to visually specified events. *Cognitive Psychology 7*, 4 (1975), 480–494.

[32] Norman, G.Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education 15*, (2010), 625–632.

[33] Ojala, T., Kostakos, V., Kukka, H., et al.Multipurpose interactive public displays in the wild: three years later. *Computer 45*, 5 (2012), 42–49.

[34] Peltonen, P., Kurvinen, E., Salovaara, A., et al."It's mine, don't touch!": interactions at a large multi-touch display in a city centre. *SIGCHI Conference on Human Factors in Computing Systems*, ACM (2008), 1285–1294.

[35] Prante, T., Röcker, C., Streitz, N., et al.Hello.Wall - beyond ambient displays. *International Conference on Ubiquitous Computing*, (2003), 277–278.

[36] Preece, J., Rogers, Y., and Sharp, H.*Interaction Design.* John Wiley & Sons, Inc., New York, NY, USA, 2002.

[37] Sasangohar, F., Scott, S.D., and Cummings, M.L.Supervisory-level interruption recovery in time-critical control tasks. *Applied Ergonomics 45*, 4 (2014), 1148–1156.

[38] Schmidt, C., Müller, J., and Bailly, G.Screenfinity: extending the perception area of content on very large public displays. *SIGCHI Conference on Human Factors in Computing Systems*, ACM Press (2013), 1719–1728.

[39] Seto, M.A., Scott, S.D., and Hancock, M.Investigating menu discoverability on a digital tabletop in a public setting. *ACM International Conference on Interactive Tabletops and Surfaces*, ACM Press (2012), 71–80.

[40] Simons, D.J. and Chabris, C.F.Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception 28*, 9 (1999), 1059–1074.

[41] Tri-Council Policy Statement.Ethical conduct for research involving humans. http://www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/Default/.

[42] Vogel, D. and Balakrishnan, R.Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. *ACM Symposium on User Interface Software and Technology*, ACM Press (2004), 137–146.

[43] Zhang, Y., Müller, J., Chong, M.K., Bulling, A., and Gellersen, H.GazeHorizon: enabling passers-by to interacti with public displays by gaze. *International Conference on Ubiquitous Computing*, (2014), 559–563.