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

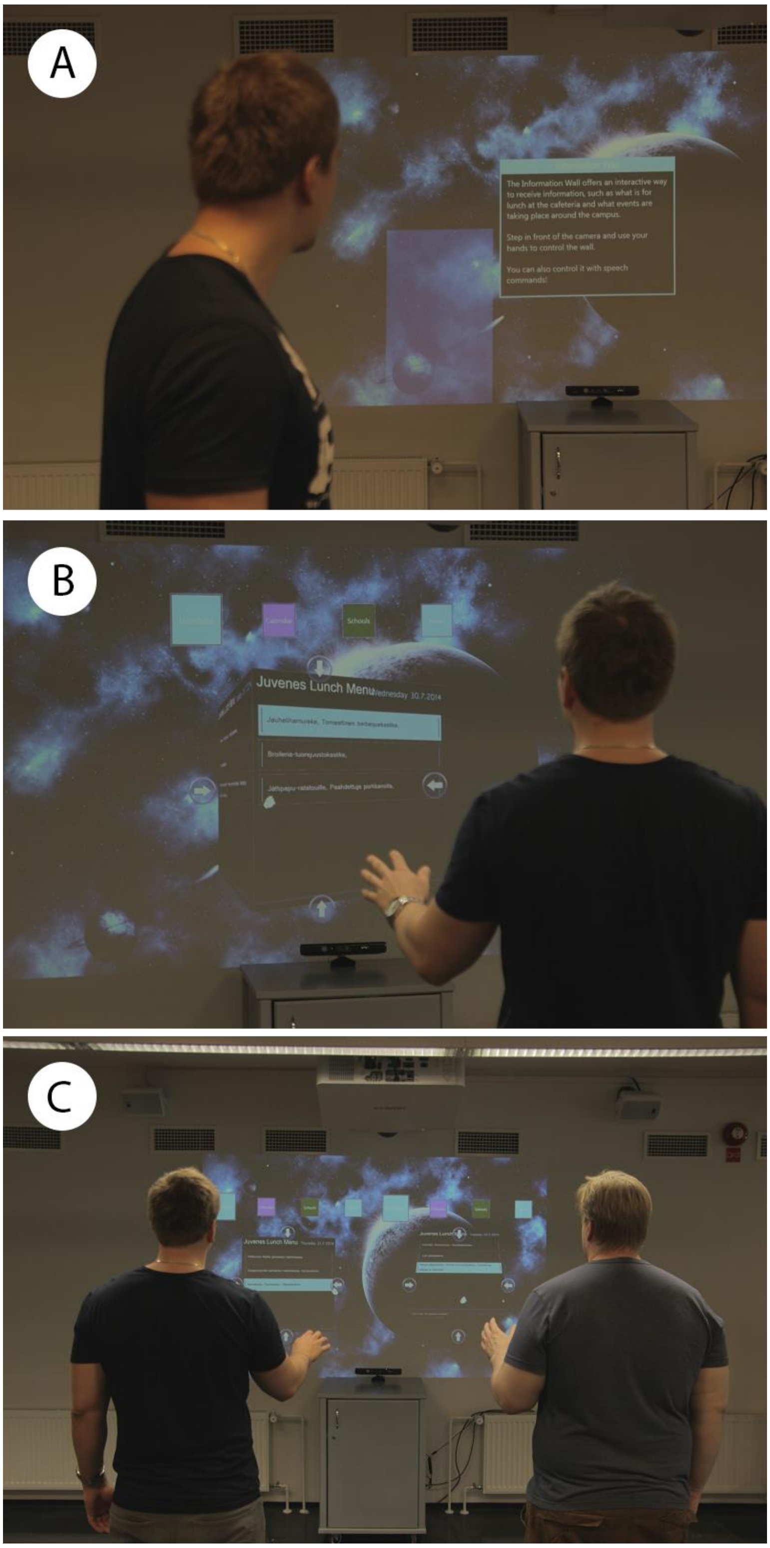

Figure 2. A) A user is tracked with a rectangular shape on the screen. An instruction dialog is displayed. B) A user interacts with an information cube. C) Two users simultaneously interacting with the wall. (Adapted from Mäkelä *et al.* [11])

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/

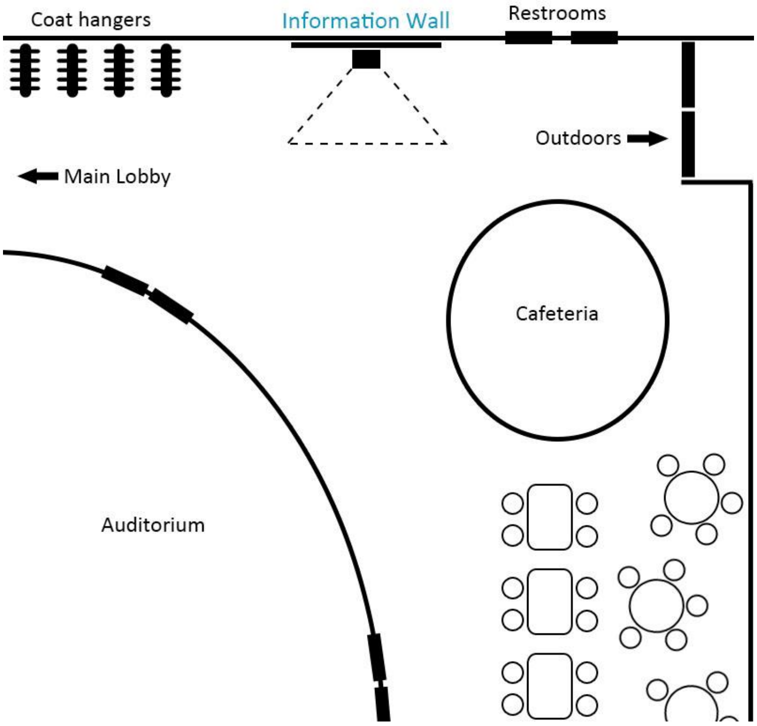

Figure 3. Floor plan of the deployment space.

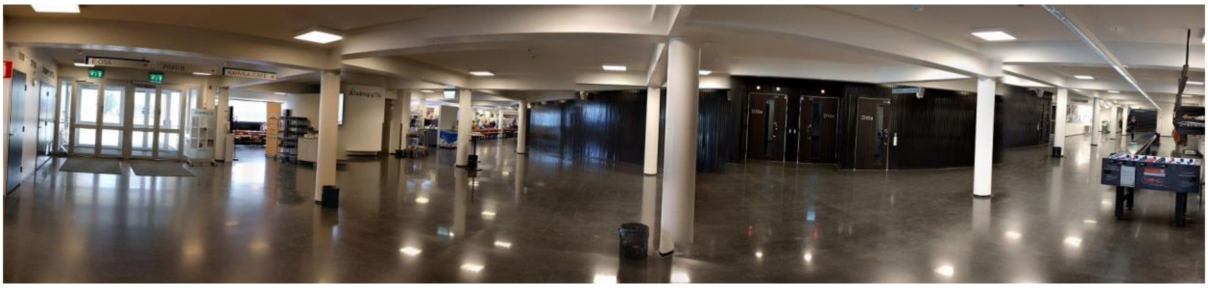

Figure 4. Panorama of the space from the installation's perspective.

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.

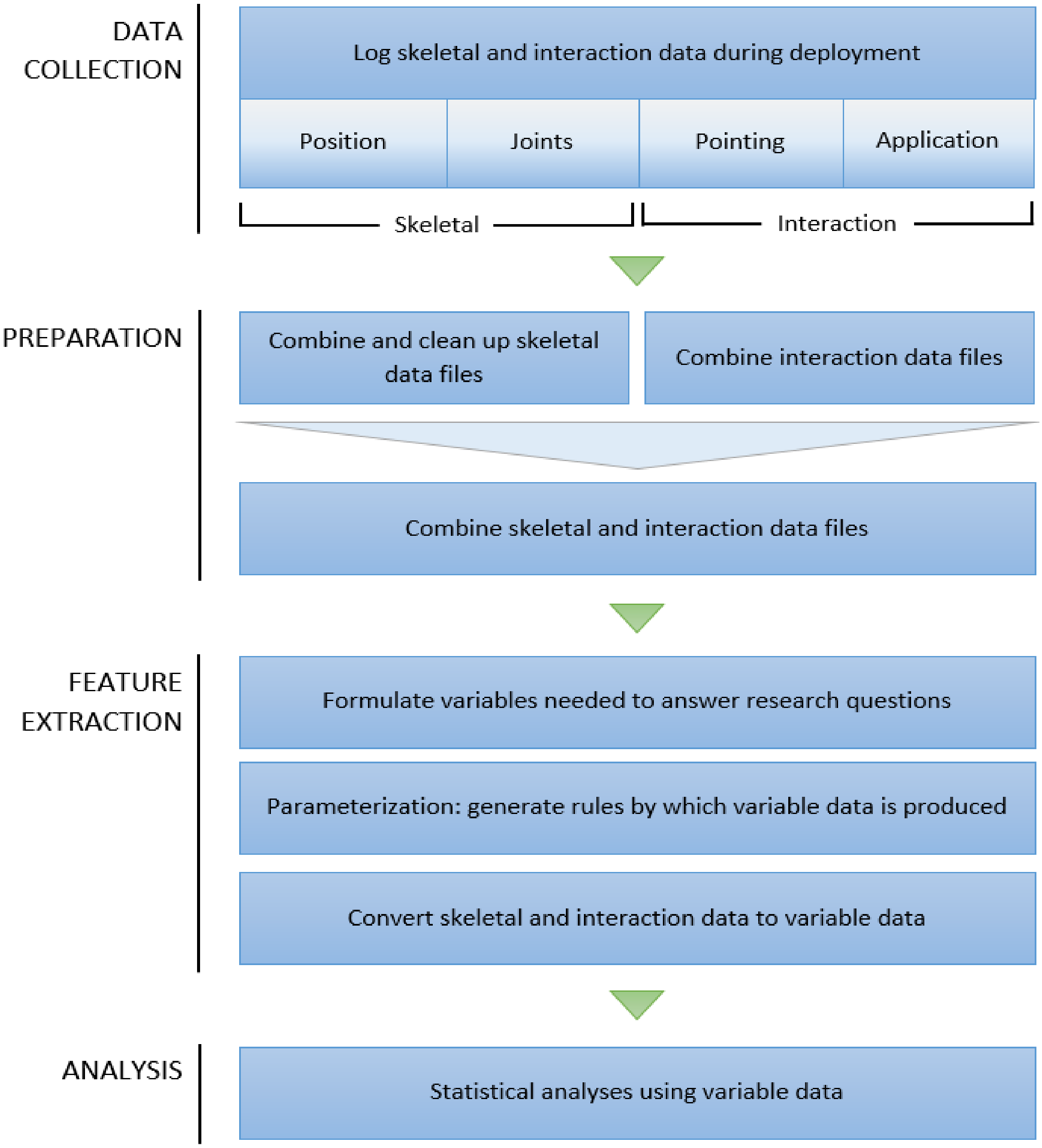

Figure 5. The semi-automated process consists of four primary phases: data collection, preparation, feature extraction, and analysis.

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%.

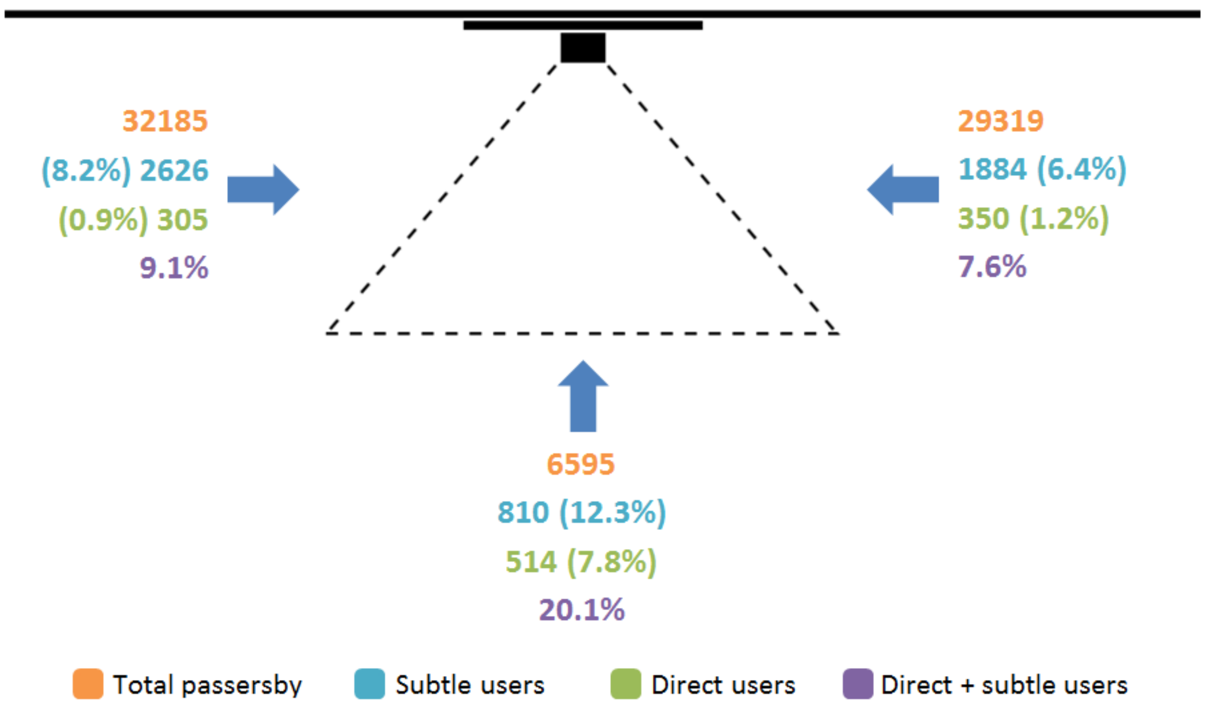

Figure 8. Users' entry directions.

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).

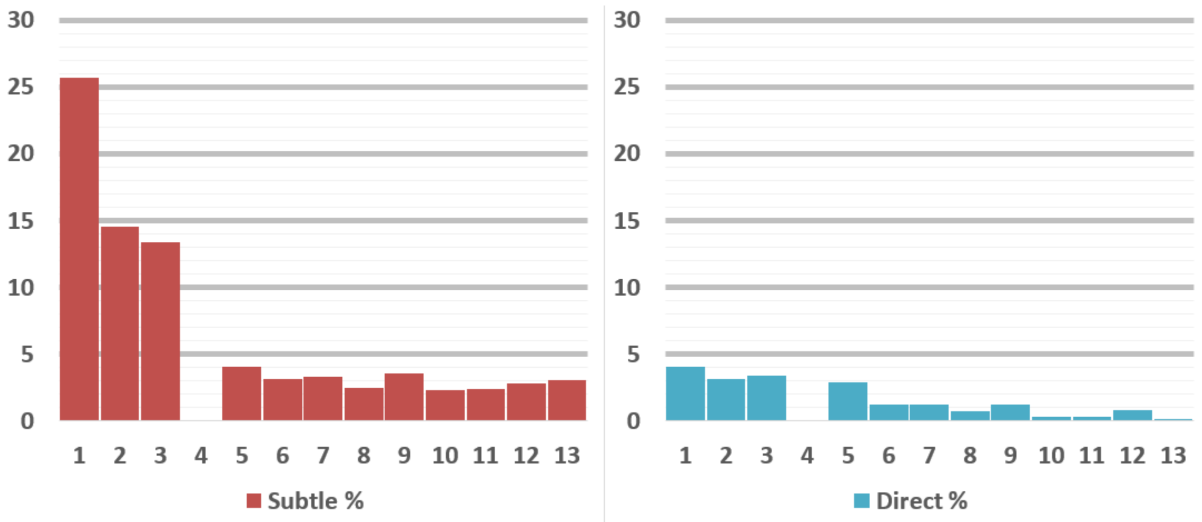

Figure 9. Percentages of subtle users (left) and direct users (right) within each month of deployment.

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.