Orbuculum – Predicting When Users Intend To Leave Large Public Displays

FLORIAN ALT, Research Institute CODE, Bundeswehr University Munich, Germany
DANIEL BUSCHEK, Research Group HCI + AI, University of Bayreuth, Germany
DAVID HEUSS, LMU Munich, Germany
JÖRG MÜLLER, University of Bayreuth, Germany

We present a system, predicting the point in time when users of a public display are about to leave. The ability to react to users' intention to leave is valuable for researchers and practitioners alike: users can be presented additional content with the goal to maximize interaction times; they can be offered a discount coupon for redemption in a nearby store hence enabling new business models; or feedback can be collected from users right after they have finished interaction without interrupting their task. Our research consists of multiple steps: (1) We identified features that hint at users' intention to leave from observations and video logs. (2) We implemented a system capable of detecting such features from Microsoft Kinect's skeleton data and subsequently make a prediction. (3) We trained and deployed a prediction system with a Quiz game which reacts when users are about to leave (N=249), achieving an accuracy of 78%. The majority of users indeed reacted to the presented intervention.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**; **Displays and imagers**.

Additional Key Words and Phrases: public display, audience behavior, prediction; deployment- based research

ACM Reference Format:
Florian Alt, Daniel Buschek, David Heuss, and Jörg Müller. 2021. Orbuculum – Predicting When Users Intend To Leave Large Public Displays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 1, Article 47 (March 2021), 16 pages. https://doi.org/10.1145/3448075

1 INTRODUCTION

With display and sensing technology becoming widely available, a lot of interactive applications have emerged over the past years: playful applications in store windows attract and engage passersby [18, 20]; interactive directories help users find their favorite stores in large malls [16]; voting applications in public spaces support civic discourse [24, 27]; and maps guide travellers the way through airports [6, 25] and train stations [7]. While a lot of research focused on how to make the user interact in the first place [13, 18, 29], the end of the interaction process only received little attention so far. Yet, such knowledge yields significant potential: operators of store directories can determine what brands users are interested in and, as users are about to leave, provide them a discount coupon to be redeemed at the store they are heading for. Or multi-application displays can point users to further content as they finished their current task. Last but not least, display operators and researchers could

Authors' addresses: Florian Alt, Research Institute CODE, Bundeswehr University Munich, Germany, florian.alt@unibw.de; Daniel Buschek, Research Group HCI + AI, University of Bayreuth, Germany, daniel.buschek@uni-bayreuth.de; David Heuss, LMU Munich, Germany; Jörg Müller, University of Bayreuth, Germany, joerg.mueller@uni-bayreuth.de.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.
2474-9567/2021/3-ART47 $15.00
https://doi.org/10.1145/3448075

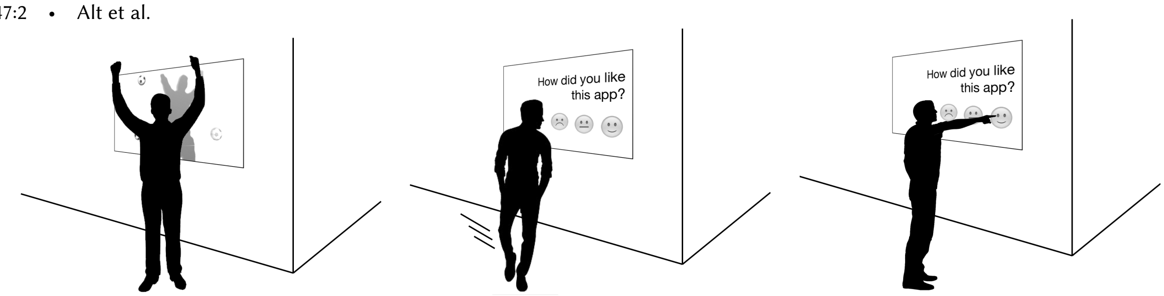

Fig. 1. We present a system, predicting when users are about to leave an interactive public display. The prediction is based on features, such as body orientation, movement, and relaxation. As a result, the display can react by presenting new content or asking the user for feedback (middle) and hence prolong interaction (right). Hence, value can be created both for users and researchers. We report on the design and implementation of a prototype and present findings from a field deployment.

benefit from requesting feedback from the user right after the interaction, for example, whether they found the information they were looking for or whether they enjoyed interacting with a particular application.

Predicting the point in time when users are about to leave can be difficult. On one hand, there are display applications with a clear end of task, for example, purchasing a train ticket or playing a game that lasts for a pre-defined amount of time. In such cases there is a high chance that users leave at this point. On the other hand, a task may not be clearly defined and hence have no clear end, for example, as users are browsing the content of a digital public notice area [1] or as they are playing a continuous game (similar to [18, 29]). In particular, the latter case requires prediction mechanisms that are independent of the content. But also in the former case, users may decide to leave before they finished their task. Users leaving before the end of the task could be a hint for a usability problem that, with our approach, a display could try to address instantly.

In this work we present a concept, implementation, and evaluation of a system, capable of predicting the point in time when users will leave a public display. The concept is depicted in Figure 1. Our work follows three steps:

Feature Identification To inform the system design, we identified suitable features. We conducted real-world observations in the vicinity of public displays (shopping centers, airports, etc.) and reviewed video material from prior public display installations. From these observations and reviews we derived a codebook containing potential features. To identify the most suitable features, we coded 4.5 h of video footage. We found body orientation, walking to the side, walking forward/backwards, as well as taking down the arms during full-body interaction to be good predictors. An interesting finding was that some features strongly depend on deployment-specific characteristics, such as the interaction modality, application, or location.

Design of Prediction System We implemented a Kinect-based system, capable of detecting the most promising features from skeleton data. We explain how the obtained data can be labelled using a ground truth of when users actually left (e.g., derived from video annotation) and be used to build a prediction model.

Proof-of-Concept We integrate the prediction system with a playful display app and deploying it over the course of 11 days. The deployment process also serves as a blueprint for researchers and practitioners who want to employ the approach. In particular, we implemented a Quiz game that prompts users as the system thinks they are about to leave and instantly asks them to confirm the assumption.

During the deployment, 249 users interacted with the game. From these, 175 reacted to the prompt, confirming (a) that the prediction was correct in 78% of all cases and (b) that the overall approach is viable since the point in time when users leave can be predicted with sufficient time for users to notice and respond to a prompt (3.8 s).

Contribution Statement: The contribution of this paper is twofold. Firstly, we present a *set of features* that allow for predicting when users are about to leave a public display and discuss applicability. Secondly, we report on the *design and in-the-wild evaluation of a prediction system*, demonstrating that based on our predictions interventions can be created that are perceived by users and can lead to users changing their interaction behavior.

2 BACKGROUND AND RELATED WORK

Our research draws from several strands of prior work, most importantly modeling interaction with large displays and understanding user behavior in public space. In addition, prior work reported on reasons for which people leave public displays and allowed to make predictions.

2.1 Interaction Models for Public Displays

Previous work investigated how audience behavior in front of public displays can be modelled [2, 9]. On one hand, this includes *spatial models* like Prante's three zones of interaction model [19], the extended model by Vogel et al. [28], and the work on proxemics by Greenberg et al. [11]. In all these works, the space in front of a public display is classified into different interaction zones based on proximity. The immediate vicinity of the display, aka the personal interaction zone, is of particular interest to our work. The aforementioned models assume that as the distance of the user to the display increases, personal interaction ends. Distance may be an interesting feature to be investigated for prediction – yet beyond a certain distance it may already be too late to make the user continue interaction. Thus, this work considers distance just as one possible feature.

On the other hand, there are *temporal models*, describing the interaction process as a movement of the user through different stages, where thresholds need to be overcome to proceed to the next stage. In Brignull and Rogers' public interaction flow model [5], users proceed from the space of peripheral awareness to the space of focal awareness before ultimately entering the space of direct interaction. Before entering the final stage, the so-called participation threshold needs to be overcome. As interaction ends, users transition back to the spaces of focal/peripheral awareness. Depending on the situation, users' dwell time in this area may be very short. Whereas in situations similar to the one where the effect was discovered (a book launch party), users may transition through spaces multiple times. We focus on situations where users are about to finally leave and hence the display may not receive any attention afterwards.

Another temporal model is the audience funnel [17], that describes six stages leading to direct interaction. Users are first passerby, but as the display captures their attention, they may view and react. Once they understand that the display is interactive, they may engage into subtle interaction with the goal to understand how the interaction works. Finally, as they are motivated to interact they may engage into direct interaction. What is interesting about the model is its stage called multiple interactions, referring to users coming back or encountering similar displays on the way. As a display can determine that a user wants to leave and manages to motivate them to continue interaction, multiple interactions may occur at the same display and in a quick temporal order. Also the ultimate stage of the audience funnel, so-called follow-up actions, are of interest. With our approach it becomes possible to trigger these follow-up actions, e.g., by providing a coupon for the user.

2.2 User Behavior

To appropriately identify and react to users' intention to leave, an understanding of how people behave in public space is important. We list a number of effects / terms and explain how they relate to our research.

The *honeypot effect* [5, 18, 30] describes the situation when passersby notice a manipulation of others through a display and approach for closer observation. Often, observers stay for some time and as the person currently interacting leaves, step in to start interacting. Despite being beyond the scope of this research this is interesting, since bystanders may directly impact on the motivation of a user interacting. As a result, social pressure to also allow others to interact may emerge, hence making it more difficult to convince the user to stay and, for example, provide feedback before she leaves.

Ten Koppel et al. showed that the *display configuration* has a strong influence on how people approach and interact with these displays [26]. For example, concave displays lead to rather low numbers of people interacting, while flat displays trigger a strong honeypot effect. From this we learn that predictions may depend on the display configuration / orientation since it may trigger different leaving behavior.

The *sweet spot* is the point in 2D space in front of a display where users usually position themselves for interaction, since here they can optimally perceive the content [4]. This sweet spot depends on the size and position of the displays and may influence the behavior of people at the end of interaction.

What is finally interesting is the notion of *f-formations* [8, 14] that is the possible spatial configurations of people during social interaction. In particular in multi-user situations it may be important to consider these formations since behavior of people may differ with others interacting in parallel.

2.3 Reasons to Leave the Display

While finding a particular piece of information (for example, a store location) or solving a particular task (for example, train ticket purchase) are obvious reasons to end interaction, reasons are less clear in case of open-ended tasks. Why do people end interaction during a continuous game or browsing content?

Previous work reports on several cases where people decided to abandon a display. For example, in the case where a group approaches the display but then only some members of the group interact, other members soon leave the display, creating a tension among those interacting [18]. In many cases people play for some more seconds, before running after and rejoining their group. From this we learn that observing not only the person interacting but also the display vicinity may provide valuable hints as to when users are about to leave the display.

Another reason for people to leave is interrupting interaction. In Strike-A-Pose interrupting gameplay to reveal a hidden feature led to many people leaving the display [29]. The effect of interruption on large display has been further investigated [3]. Findings suggest that so-called secondary tasks need to be carefully used so as not to negatively impact interaction. This highlights the importance to be able and predict the optimal point in time when a user should be approached by the display.

In recent work, Wouters et al. found that reasons for people to dropout include disappointment, discomfort, and completion [30]. Other causes are social reasons (for example, crowdedness, perceived complexity of interactions), external reasons (for example, receiving a call), having discovered all interactive features, or physical reasons (for example, exhaustion).

In conclusion, the reasons for people to leave the display are manifold. A display's ability to prolong interaction strongly depends on those reasons. Our work enables developing means to influence the user. With advances in sensing it may become possible to not only determine a user's intention to leave but also the reason, hence allowing an informed choice for a suitable intervention. For example, as a system detects that a user is about to leave due to exhaustion, it may offer a physically less challenging task.

2.4 Prediction of User Behavior

There exists only little work that tried to predict user behavior on public displays. One of the few examples is the work from Huber et al. [12] who tried to detect intentions of public display users from foot positions. In particular, they tried to predict whether users are in a hurry, i.e. passing-by, or rather in a waiting situation. As a result, users could be presented either quickly-to-perceive information or an entertaining application. As a single feature, this work leverages foot movement. But in contrast to this work we do not focus on determining the user's intention at the outset of interaction but are interested to react to the user's intention to leave.

Of interest in this regard is work by Sahibzada et al. [22]. They found that, for example, the use of a smartphone is mainly reasonable in locations with a younger audience and/or where people have their phone ready anyway. An interesting approach to create display applications that target a broader audience (e.g., younger and older people at the same time) was to try and predict the reason for which people leave a display. For example, do they leave because they are uncomfortable with installing an app on their smartphone or do not have a smartphone? In this case, an intervention could propose an alternative way of interacting with the display (e.g., using touch or mid-air gestures). While interventions are beyond the focus of our work, we believe (a) means to identify *why* people leave and (b) interventions targeted to this reason to be an interesting direction for future work.

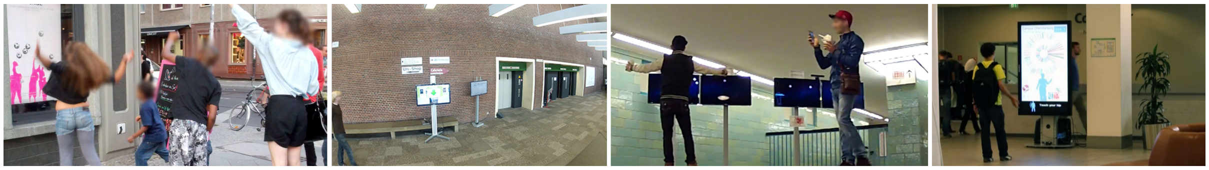

Fig. 2. We analyzed situations in which users are leaving a public display from four different deployments. Deployments differed regarding the physical setting, the deployed application, and the interaction modality.

2.5 Summary

The literature revealed that predicting user behavior is a complex challenge that depends on a large variety of factors, including but not limited to the display location, the display surroundings as well as other audience. We decided to focus our investigation on single user cases to show the feasibility of our idea and leave multi-user scenario for future work.

3 PREPARATION: OBSERVATIONS & VIDEO REVIEWS

We first investigated whether it was possible for a *human* to predict when a person will leave a display and, if so, how accurately. This served to proof the feasibility of our idea and to determine features for later implementation.

We identified a set of promising features by running a number of informative field observations as well as by reviewing videos from previous public display deployments.

In contrast to the common method of examining features with a "black-box" machine learning algorithm (for example, learning on a Kinect dataset), our approach uses human observations first. Hence, we 1) can build directly upon several previous studies and related work, and 2) gain insights into features related to human reasoning, which we expect to be useful for the community beyond this specific piece of research.

3.1 Field Observations

We set out to observe people interacting with displays in a large shopping mall and at a major international airport at different times of the day. A researcher would position himself in different locations from which he could observe the person interacting but was not identifiable as researcher. Locations were changed several times during the observations and lasted for about two hours each. Altogether 42 people interacted with the displays during our observations. One group consisted of four people, two groups consisted of three people and seven groups consisted of two people. Eighteen people interacted alone. Interaction lasted on average for 50 seconds.

The field observations allowed us to gain an initial understanding of how people behave as they are about to leave the display. Important observations were that many people turn their shoulder prior to turning away their head from the display before leaving. Also, many people looked back at the display as they were already walking away. This is promising, since it gives display applications an additional chance to re-catch users' attention and make them aware of additional content. This finding is also interesting because it creates a challenge: since people turn away their attention to look at the way ahead before directing attention back, this may lead to a phenomenon commonly known as change blindness [23], i.e. people may miss that content on the screen changed.

Furthermore, we found that movement parallel or orthogonal to the display seems to be a good indicator. Interestingly, though, people often seemed to not back away from the display but to rather walk towards the display making a turn to the left or to the right. This is equally interesting, since it again gives a display app time to react to the user's behavior.

3.2 Video Reviews

From the field observations we learned that hints were often very subtle and happened very quickly. Hence, we decided to complement our field observations with reviewing videos from prior public display deployments to learn more about the predictive power of different features. We contacted various researchers who had previously run public display deployments to re-evaluate the logging data from these deployments. In total we received 7.3 hours of video material. A detailed description of the deployments can be found in the following section.

We randomly selected scenes that were reviewed by two researchers with the goal of identifying further features. Since the deployments were quite different with regard to the location, application, and interaction technique, we were able to collect additional insights. For example, we found that in applications controlled through mid-air gestures, the moment when people take down their arms (i.e. relaxed) seems to be a good indicator. Since the video allowed the moment of leaving to be played back in slow motion, we were also able to capture more subtle features. For example, in several cases, turning the head and moving the head upwards / downwards seemed to coincide with the intention to leave. This vertical movement might be a result of the display's height: smaller users look upwards while taller users look downwards; as they stop interaction they return to their natural head position. Furthermore, body tension seemed to increase as users prepared to move on.

An interesting observation was that users' behavior differed based on whether they were interacting in groups or alone. For example, as one person left, this increased the chance that others in the group left as well. As previously mentioned we consider this an interesting direction for future work (both from a sensing and an intervention perspective) and did not include this case in our investigation.

During the review, two researchers took notes of the features that were later used to create detailed descriptions and derive a code book. The full list of features can be found below.

4 FEATURE IDENTIFICATION

For the subsequent coding study, videos from the following deployments were used (see Figure 2).

Looking Glass The deployment consisted of three displays that were installed in a retail store in the city center of Berlin, close to a subway station. The application running on the displays was an interactive, continuous ball game, where people could interact with balls using a representation of themselves on the screen (in particular a mirror image or their silhouette). The game neither had a clear goal nor end. People played on average for 31 s but there were many interactions that lasted only for a few seconds. More information on the deployment can be found in Müller et al. [18].

StrikeAPose A large display was deployed in the foyer of the cafeteria at the TU Berlin. It showed a game in which users interacted with virtual cubes on the screen via their mirror image. In particular, users could toss the cubes into a target to collect points. One feature of the game allows them to get a bunny mask as they performed a particular gesture. The gesture was communicated through a hint on the screen. More information can be found in Walter et al. [29].

GazeHorizon GazeHorizon was a display installation at Lancaster University [31]. It allowed people to browse content on a display using gaze as a modality. The display was deployed in front of a lecture hall. The videos covered a diverse set of users, both such that interacted with the game prior to leaving but also cases where people did not understand the interaction technique.

Pauses in Interaction Two screen pairs were deployed in a subway station in the city of Berlin. Displays were situated in a corner along the main trajectory of passengers. The project investigated whether or not people interrupt interaction with a public display if provided the opportunity to do so and why. The game on the display was an adopted version of space invaders controlled by the user silhouette and through mid-air gestures. A detailed description of the game and deployment can be found in Feuchtner et al. [10].

4.1 Approach

From these videos we randomly selected 240 videos, each lasting up to 4 minutes. Each video showed interaction with an application by a single user. We arranged the videos by means of a playlist such that always 15 videos from one deployment were shown in a row (15 videos × 4 deployments × 4 blocks). This was done so that people could familiarize themselves with the deployment while at the same time trying to keep fatigue as low as possible.

In the study, coders assessed (based on time-stamped videos)

- the point in time when the coder thought to know that the person was about to leave the display, and
- the point in time when the person finally left.

From these numbers we derived an estimate on how much time a system would have to respond to the user's behavior and provide an intervention. Furthermore, we investigated whether different features would allow the point in time when a user leaves to be predicted with a larger timeframe.

For each case, coders were asked to record the feature that led to their prediction (*Main*). In addition, they wrote down all features that occurred in this case (*All*). The features were:

Foot Movement People start to subtly move their feet.

Walking to the side People subsequently leave the display to the left or to the right side.

Walking towards the display People walk a few steps towards the display before leaving to the left/right.

Walking backwards People back away from screen.

Head Rotation People rotate their head to the left or to the right when leaving the display.

Shoulder Rotation People rotate their shoulder to the right or to the left prior to leaving the display.

Body Rotation People rotate their hip away from screen.

Upward / downward head movement People move their head up or down prior to leaving.

Raise /Lower Upper Body People raise / lower upper body.

Body Relaxation People take a relaxed body posture, for example, lowering arms and standing relaxed.

Finally we asked coders to record and describe all cases that were not yet covered.

4.2 Codebook

We created a codebook to ensure comparable results. This codebook, firstly, contained a detailed explanation of the task. In particular, we explained them how to use the video playback tool, how to watch the videos, how to record timings, and how to enter the main feature and all other observed features in a spreadsheet we had prepared for them. Therefore, we provided a detailed description of the different columns of the table and some examples. Secondly, we provided coders a detailed explanation of the different features and feature combinations. For example, we specified the approximate angles for which they should code the behavior of users as walking towards the display, walking to the side or walking backward. The description was meant to serve as a reference for coders while watching the videos. The codebook is available as supplementary material.

4.3 Participants and Procedure

We recruited three people as coders who were not part of the research team and not familiar with the project. Recruitment channel was a Facebook group of the computer science study program of our University. Two coders were computer science students, one coder studied Human-Computer Interaction (all male). Based on our experience while creating the codebook we estimated the coding to take approximately 8 hours. Each coder was compensated with a 100 Euro gift voucher.

Feature	Total		Looking Glass		StrikeAPose		GazeHorizon		Pauses in Int.	
	Main	All	Main	All	Main	All	Main	All	Main	All
Foot Movement	23.76	35.17	7.74	9.52	30.14	47.95	44.67	56.67	17.76	36.18
Relaxation	19.52	22.65	16.07	18.45	13.70	16.44	2.00	2.67	43.42	50.00
Body rotation	16.39	27.26	19.95	25.00	23.29	32.88	12.67	26.00	13.82	28.29
Walking to the side	9.58	26.70	13.69	25.60	13.70	42.47	5.33	25.33	7.24	21.72
Shoulder rotation	9.58	19.52	20.83	27.38	12.33	15.07	2.00	12.67	3.29	19.74
Head rotation	7.92	14.18	1.19	8.33	23.29	32.88	12.00	18.00	3.95	7.89
Walking backwards	5.52	17.50	7.74	14.88	0.00	5.48	5.33	18.00	5.92	25.66
Walking towards	3.68	6.08	8.93	14.29	1.37	4.11	0.67	1.33	1.97	2.63
Raise upper body	1.29	1.47	0.00	0.00	2.74	2.74	2.00	2.00	1.32	1.97

Table 1. Overall distribution of features (All) and distribution of main features (Main) used for prediction (both in percent; N=542). We only consider features that occurred in at least 1% of all cases. Coders identified a total of 959 feature occurrences.

Coders were separately invited to our lab, i.e. coders did not interact with each other. The workstation consisted of a desktop PC with a 28 inch monitor, keyboard and mouse. A researcher explained each coder the task, the features, and the codebook. All times and features were manually recorded in a spreadsheet. We encouraged them to contact one of the researchers in case of questions.

To assess inter-rater reliability, we calculated Fleiss Kappa for the features coders mainly based their decision on, that are shoulder rotation, foot movement, body relaxation, body rotation, and walking to the side (cf. section 4.6). We found a moderate agreement between all three coders for those features (0.56).

4.4 Data Analysis

Overall, the three coders made 720 predictions. In a first step, we identified all cases, where coders were not able to make a meaningful prediction. On one hand, this included cases where users left before the coders were able to notice any signs for this (32 cases). On the other hand, we excluded cases where the time when users made a prediction and the actual departure was very long. We assume that in these cases users either changed their mind or coders simply misinterpreted their behavior. Therefore, we set the threshold for a meaningful prediction to 10 seconds, i.e. if people did not leave within 10 seconds after the coder made the prediction, we removed this case. Hence, we included 78.8% of all predictions. The results are based on the remaining data (542 cases).

4.5 Results

Analyzing the data obtained from the coding study allowed us to answer the following questions:

- On which features do coders base their prediction?
- How long in advance can we predict that a user will leave a display?
- Are there any differences between the deployments?

4.5.1 Features Used for Prediction. As can be seen from Table 1, the most observed features (*All*) were Foot movement (35.2%), Relaxation (22.7%), Body rotation (27.3%), Walking to the side (26.7%), and Shoulder rotation (19.5%). The most important features on which coders based their prediction (*Main*), were Foot movement (23.8%), Relaxation (19.5%), Body rotation(16.3%), Walking to the side (9.6%), and Shoulder rotation (9.6%).

As described in the previous section, deployments were quite different regarding employed interaction modality, application, and location. Investigating the distribution of features, we made a number of interesting observations.

Feature	Overall	Looking Glass	StrikeAPose	Gaze Horizon	Pauses in Int.
Walking to the side	3.55	3.13	4.50	3.13	3.90
Relaxation	2.94	2.25	2.40	3.33	3.28
Body rotation	2.70	2.22	2.58	2.31	3.23
Foot Movement	2.57	1.30	2.81	2.28	3.70
Head rotation	2.48	5.00	3.58	2.44	1.66
Walking towards	2.40	1.66	2.00	1.00	6.66
Walking backwards	2.36	1.92	—	1.75	3.55
Raise upper body	2.13	—	1.50	2.16	1.00
Shoulder rotation	2.03	1.80	2.55	2.66	2.40

Table 2. Average prediction time per feature overall and per deployment (in seconds). We found that features coders used for their prediction also yielded the longest prediction times.

The occurrence and distribution of some features seems to depend on the employed *interaction modality*. For example, Relaxation was among the most prominent features for gestures-controlled applications (Looking Glass, StrikeAPose, Pauses in Interaction) but was rarely noticed for gaze interaction (Gaze Horizon). Shoulder Rotation was prominent for applications still allowing interaction while already leaving, which was possible for Looking Glass and StrikeAPose but not for GazeHorizon, where people had to stand in the eye tracker's sweet spot.

Also the *application* itself may have an influence. For Pauses in Interaction, people could suspend interaction using a gesture. Most people did so before leaving, hence making Relaxation the most prominent feature.

Summary. From this we learn that certain features occur frequently, independent of deployment characteristics. Yet, there is no single fixed set of best features, so we recommend determining the best feature set per deployment.

4.5.2 Prediction Times. Table 2 summarizes the average prediction times, that is the time from when the coder believes to know a user is leaving until the user indeed leaves. Looking at the overall prediction times, all investigated features provide a display application 2–3 seconds to respond with an intervention that the user can possibly see and react to[1].

Summary. Looking at the single numbers, we learn that the most frequent features – Walking to the side, Relaxation, Body rotation, Foot movement (as identified in the previous paragraph) – allow the point in time when users leave to be determined between 2.2 and 4.5 seconds in advance. Note, that the feature Foot movement is a preceding action to all walking features (Walking to the side, Walking forward, Walking backward). Hence, we decided to, apart from the features Body rotation and Relaxation, focus on all walking features (Walking to the side, Walking forward, Walking backward) including the Foot movement.

5 PROTOTYPE

To prove the concept, we built a prototype that is capable of (1) detecting features that hint at users' intention to leave; (2) based on these features make a prediction; and (3) provide this information to a display app. The app can then respond to this information with an intervention on the screen.

[1]Note that videos from the deployments differ in camera angle and distance of the camera to the user. Since we consider that users left the display as soon as they leave the visual field of the camera, it is difficult to compare prediction times across different deployments.

5.1 Feature Detection

We use a Microsoft Kinect v1 to detect the most promising features, i.e. Body orientation, Walking to the side, Walking forward / backward, and Relaxation. We calculate features based on the Kinect's skeleton data, namely the person's location (i.e. x,y,z body centre), the angle of the shoulder points and the display, as well as heights of hand and arm joints (to assess "relaxation", following the observations of such postures in the video study).

The feature detection is designed to continuously run in the background. Features are detected in real-time but can also be stored in a text file with a time-stamp. The tool also allows for video-recording the area in front of the display to later obtain a ground truth. The tool was implemented in C# (Microsoft .NET 4.5 framework).

5.2 Prediction

For making predictions, skeleton data needs to be collected as well as data that allow for obtaining a ground truth. This can, for example, be done by recording videos. By watching the videos, the timestamp of the user's leaving for each interaction can be identified. These can then be used to manually annotate ground truth times to label the data samples for each interaction as *before* or *after* leaving the display.

The labelled data will then be used to train a classifier. For our implementation, we chose a simple decision tree[2] [21], since it allows us to inspect and interpret the learned rules (in contrast to, for example, an SVM). Note that our goal is not to find the optimal classifier. We are instead interested in understanding the trained machine learning method's decision making, compared to coders' human reasoning.

5.3 Display Game

We implemented a 'What is this?' game, intentionally designed to keep interaction times short. As the display detected a person in front of the screen, it shows a blurred image with four answers on the left side of the screen. Over time, the picture becomes clearer and finally reveals the correct answer by highlighting one option.

5.4 Intervention

We chose a very easy intervention that would simply ask users whether or not they wanted to leave, as soon as the system predicted so. Specifically, the screen would switch from the game to a view that briefly explained users that the system thinks they are about to leave and asks whether or not this is correct. Users could chose an answer by hovering a hand over one of two buttons (yes/no, cf. Figure 3). If the user chose 'no', the display returned to the game, whereas for 'yes' it would thank the participant and return to idle.

6 DEPLOYMENT

We integrated the prediction system with the aforementioned game to investigate its ability to reach users about to leave. This deployment also serves as a blueprint, showcasing how other practitioners and researchers could use the approach. Table 3 describes the deployment, following the suggestions by Memarovic et al. [15].

We deployed a public display in a University building. The building we chose was home to research groups in the humanities. The display was deployed in a long hallway connecting parts of the building, parallel to the main walking directions of people. The hallways was several meters wide, hence allowing people interacting to comfortably position themselves in front of the display at their preferred distance (the so-called sweet spot [4]). People not interacting could easily pass both the installation and the interacting person at a distance. Hence, we assume the influence of passers-by on people interacting was minimal.

In the immediate vicinity of the display were several large lecture halls as well as chairs and tables for students to work on assignments. The rear part of the building contained offices of faculty and research staff who had to pass the area where the display was deployed to get to/from their workplace. Hence, the audience consisted of students, researchers, faculty and people from administration.

[2]We used the MLR package: https://mlr-org.github.io/

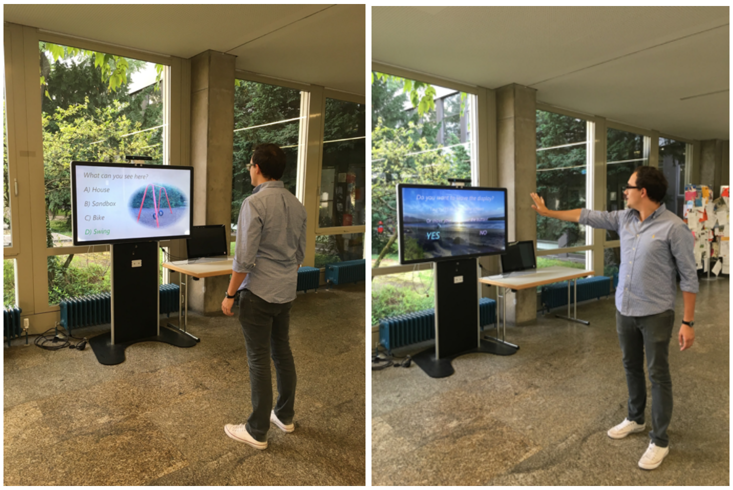

Fig. 3. To prove the concept we deployed the display in a University setting. The display shows a quiz (left). As the system thinks the user is about to leave, users are shown a prompt (right), asking them to confirm whether they are indeed about to leave or not. Users can provide feedback by hovering with their hand over yes/no buttons.

6.1 Process

We briefly describe the single steps to include our prediction system with the display application.

6.1.1 Step 1: Building the Prediction Model. We deployed the game without any changes or interventions, collecting Kinect data to compute the described features (see Section 5.1) over the course of four working days, as described earlier. For obtaining a ground truth, video footage was collected. Participants were notified about the video recording through a sign showing a video camera attached to the display stand. Based on an analysis of the data from 44 participants we trained and implemented a decision tree: Specifically, we used the *classif.rpart* classifier from the mlr R package. The model was trained to predict leaving (manually labelled) on the windowed timeseries of feature values (i.e. mean per feature per window). We used ten-fold stratified cross validation to compare results for different settings and based on this chose 30 frame windows and keeping the default hyperparameters for the deployment. The resulting decision tree was then manually implemented as if-then-else rules in the deployed interactive system (see Step 3).

In contrast to the analyzed deployments, walking backward was more pronounced whereas walking towards the display while leaving happened infrequently. Body rotation was a good indicator. As could be expected, relaxation (i.e. taking down the arms) was a poor indicator, as the app did not require mid-air interaction.

6.1.2 Step 2: Designing and Adding the Intervention. In the second step we added the intervention we described in the previous section. The goal was to test whether the system would be able to catch users' attention as they were about to leave the display. This allowed us to test both the accuracy of the system as well as to gather early insights on user behavior, i.e. whether they would consider re-engaging with the display.

6.1.3 Step 3: Deployment and Analysis. We deployed the game and intervention for 11 days. During that time, 249 people attended the 'What is this?' game. People spent on average 15.8 seconds (std. dev. = 21.8 s) in front of the display. Figure 4 shows the distribution of interaction times. Note, that we deliberately designed the game to support rather short interaction times so that more people could be interacting. Other applications might have a longer or shorter interaction duration or other distributions. A strength of our approach is that it is independent of interaction times. While a naive approach could prompt after the average interaction duration, this would likely annoy users who still want to continue interacting.

Name of deployment:	**Orbuculum**		
Deployment Duration:	The display was deployed over a duration of 11 days and was running 24/7 (as also the building was open without any restrictions). The display was present in the location for several years and is regularly used for deployments. Hence, many passersby can be assumed to be familiar with the display, minimizing novelty effects. The particular application we used we not deployed on the display before.		
Description of Content:	The display showed the interactive 'What is this game' described in section 5. Users were shown blurred images becoming clearer over time and had to select from among 4 options what the image was depicting.		
Description of Location:	The display was deployed in a public University building, in particular, a hallway connecting two parts of the building an in close vicinity to lecture rooms with opportunities for seating nearby. Hence, the audience can be considered to consist of both passersby as well as of people waiting for their courses to begin.		
Screen Properties:	Display: 55" landscape-oriented; Kinect v01 The display was deployed parallel to the walking direction of users but could be observed from the opposite side by people sitting there.		
Numbers			
Average Interaction Duration:	15.8 s		
Overall Interactions:	249		
		$\#_{prediction}$	$\#_{no-prediction}$
$\#_{prediction}$	249 (100%)	233 (93.6%)	16 (6.4%)
		$\#_{responding}$	$\#_{not-responding}$
$\#_{interventions}$	233 (100%)	136 (58.4%)	96 (41.6%)
		$\#_{wanted-to-leave}$	$\#_{wanted-not-to-leave}$
$\#_{user-responses}$	136 (100%)	107 (78.7%)	29 (21.3%)

Table 3. Description of the deployment (adapted from [15]). In 78.7% of cases in which people responded to our intervention, people confirmed the prediction to be correct. Overall, in 58% of cases it was possible to reach people with an intervention, suggesting the approach can support a transition from direct to follow-up interaction (cf. the audience funnel [17]).

- In 233 out of 249 cases the system *predicted* a user was about to leave (93.6%) and the question was displayed.
- In 137 of all cases, people *responded* to the intervention. Note that for the 96 cases where people did not respond we cannot tell whether they did not see the intervention or simply chose not to respond.
- With regard to the *accuracy*, we analyzed in how many cases our prediction was correct, according to the answers of the users. We found that in 107 out of 137 cases (78%), users indeed wanted to leave.

The results are encouraging, since they demonstrate the ability of our approach to motivate about 58% of all users to responded to an intervention, and, hence, transition to what is referred to in the audience funnel [17] as the *follow-up interaction phase*.

We post-hoc reviewed some of the videos in which the system either failed to correctly predict that people were leaving or where people decided not to respond. We found that wrong predictions of people leaving mostly happened for people who, during the interaction, moved too close to the display or too far to the side for the Kinect sensor to still fully capture their skeleton data. With regard to people who decided not to respond to the prompt, we found that this in many cases happened in situations where one person of a group had decided to interact but then, as the others left, did not stay to answer our question on the correctness of the prediction. Overall, these insights indicate gaps between the capabilities of the current system and human interpretation, since we would expect that a human observer would have better handled these cases.

6.2 Limitations

We cannot say how many people responded to our intervention as a result of the novelty effect. Yet, we had previously installed displays in this area and the application was active for quite a while prior to data collection (without this intervention, though). We hence expect the novelty bias to be rather small.

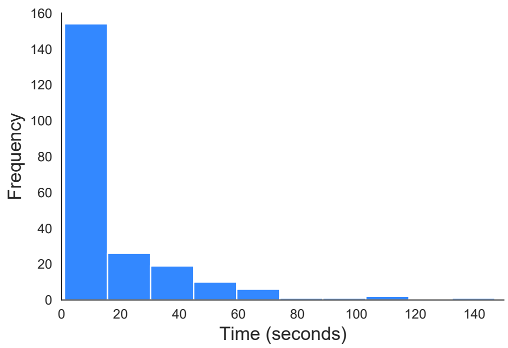

Fig. 4. Distribution of interaction lengths. The majority of interactions with our prototype were short and lasted for less than one minute (avg. = 15.8 s, std. dev. = 21,7 s).

7 DISCUSSION & FUTURE WORK

7.1 Designing the Intervention

We found that more than half of the people were able to be reached with an intervention based on the system's prediction. This suggests that it is worthwhile to further investigate the opportunities provided by the approach.

One interesting direction is the *required effort* for responding to the intervention. While we chose an intervention requiring only little effort, future work could investigate the upper limit. If it was possible to learn more about why the user was leaving – either by explicitly asking or by leveraging information that can be automatically obtained from the user or display context, an appropriate intervention could be shown. For example, in case users get bored, other content or apps could be listed for users to choose from. This information could possibly be obtained from facial expression. If the user was exhausted, a less physically demanding application could be shown. This information could be obtained from analysing user movements. In case a user stopped interaction due to the staging effect (that is, feeling exposed to other people in the display vicinity), the application could respond by switching to other modalities. Or in case the user had to go due to an appointment or the like, displays could provide very short interventions, such as collecting feedback. Whether to ask users about the reason to leave or predict this from context information remains an open question. Prior work showed that interrupting interaction may lead to users leaving [29]. Hence it may be smart to automatically suggest new apps or modalities.

Another aspect regarding the intervention is the *visual design*. First and foremost it needs to be assured, that users see the intervention. As mentioned before, users may take their eyes off the screen but then refocus as they start to leave. In this case it is crucial that users notice the intervention. At the same time, false predictions may be annoying. In our case we made correct predictions in 78% of the cases. However, this means that in 2 out of 10 cases, our prediction was wrong and interrupted users during their task. An interesting approach could be to make interventions more prominent, depending on how confident the system is that users are about to leave.

A third interesting aspect would be the ability to *identify the reason* for which users intend to leave the display and *respond to it*. Common reasons might be that users accomplished their tasks or are not motivated anymore. In these cases, as we stressed before, launching another application or initiating a follow-up action could be suitable interventions. However, reasons might be more complex. For example, Sahibzada et al. [22] identified the use of mobile devices in public to be a major barrier for people to engage in interaction with the display. Reasons for this could not only be that people generally do not want to configure their phone to be able and interact with the display but they might simply not be able because they are holding a cup of coffee in their dominant hand. In either case, the display could suggest a modality switch, for example, to mid-air gestures. Another example for a suggested switch of interaction modality could be cases where people are hesitant to interact with touch displays for hygienic reasons (cf. the current CoVID-19 pandemic) – also here the display could suggest alternatives.

7.2 Towards a Fully Automated Tool

In our current implementation, display owners are required to manually collect a ground truth, i.e. label all cases in which users actually left and then build a prediction model. Future work could look into how both obtaining the ground truth as well as the construction of the prediction model could be automated. The ground truth can be obtained by inferring from the Kinect data when a user left the visual field of view of the camera. Additionally, the prediction accuracy may be further enhanced by building more sophisticated models.

Such an automated tool could provide additional features that allow audience behavior to be analyzed. For example, it could be the case that users always leave at the same point during the interaction. This could hint at a problem with the user interfaces or users not understanding what they are supposed to do. Display owners could then investigate this more closely. In case the reason for such a problem could be identified, the tool could also instantly respond to the user, for example, by showing an intervention in the form of an explanation, tooltip, or even establishing a connection to a help desk. A specific example could be a check-in counter at an airport. As users get stuck and the system detects that the user is about to leave, not having finished the check-in process, the system could notify the passenger, that somebody will be with them shortly to assist them.

7.3 Considering Further Features

In the future we plan to enhance our prediction obtained from audience behavior with knowledge from the application. In particular, we expect the end of task to be a popular point in time when users are about to leave. This may also have an influence on the user behavior itself. Furthermore, the user's context can be taken into account. For example, as other members of the group leave, this may be a good predictor. Other context information could include information on departure times of the user's train/bus but also the weather.

Another way to enhance the prediction could be taking into account body features of the user. As our initial observations suggest, considering the height of the user relative to the display is promising, as this may trigger upwards/downwards head movements prior to the end of interaction. This is also interesting regarding the design of the user interface: in particular for very large displays, the content users are focusing on could be positioned based on their height to elicit the characteristic head movement.

7.4 Designing for Multiple Users

We did not yet look into multi-user cases. Additional features based on multiple audience members, could further increase accuracy. For example, a system could try to determine audience members belonging to a group. The point in time when one group member leaves could be a good feature to predict that the rest will leave, too.

Apart from considering the behavior of the audience as an additional feature for the prediction, multi-user scenarios are also interesting from an intervention perspective. How could multiple audience members be made to stay, for example, by providing content that is targeted towards multiple users?

8 CONCLUSION

In this paper we demonstrated that it is possible to reliably predict when users are about to leave a public display. We furthermore showed that this is possible with sufficient time to prompt users as they are about to leave and that users indeed respond to these prompts.

We believe that this work opens up a lot of opportunities for future research, including but not limited to considering multiple users, designing the intervention, including more features, and building automated tools. Furthermore, we believe our approach to be valuable for evaluation. User feedback could be collected in-situ as soon as users have finished their interaction, hence providing researchers with valuable data. Finally, our approach is also valuable for practitioners. Advertising is the prevailing business model for public displays today. As a result, only few interactive apps can be seen as of today. Our work allows display owners to explore new business models, for example, providing users with coupons and discounts or offering them to reserve a table in a nearby restaurant upon leaving the display. We are curious to see whether our work can contribute to displays receiving a higher uptake in the future as a result of (a) display owners creating new concepts for generating revenue while at the same time providing more attractive, interactive content and (b) displays taking better into account the needs of the user.

Ultimately, an interesting question is to investigate, whether users would indeed stay longer at a display and how users respond to the proposed interventions. In particular, future work could look into how to motivate the user to stay as well as how to minimize annoyance and the feeling of paternalism.

9 ACKNOWLEDGEMENTS

This project is received funding from the Deutsche Forschungsgemeinschaft(DFG) under grant agreement no. 425869382 and from the Bavarian State Ministry of Science and the Arts, coordinated by the Bavarian Research Institute for Digital Transformation (bidt).

REFERENCES

[1] Florian Alt, Thomas Kubitza, Dominik Bial, Firas Zaidan, Markus Ortel, Björn Zurmaar, Tim Lewen, Alireza Sahami Shirazi, and Albrecht Schmidt. 2011. Digifieds: Insights into Deploying Digital Public Notice Areas in the Wild. In *Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia* (Beijing, China) *(MUM '11)*. Association for Computing Machinery, New York, NY, USA, 165–174. https://doi.org/10.1145/2107596.2107618

[2] 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* (Porto, Portugal) *(PerDis'12)*. ACM, New York, NY, USA, Article 17, 6 pages. https://doi.org/10.1145/2307798.2307815

[3] Florian Alt, Sarah Torma, and Daniel Buschek. 2016. Don't Disturb Me: Understanding Secondary Tasks on Public Displays. In *Proceedings of the 5th ACM International Symposium on Pervasive Displays* (Oulu, Finland) *(PerDis '16)*. ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/2914920.2915023

[4] Gilbert Beyer, Florian Alt, Jörg Müller, Albrecht Schmidt, Karsten Isakovic, Stefan Klose, Manuel Schiewe, and Ivo Haulsen. 2011. Audience Behavior Around Large Interactive Cylindrical Screens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. ACM, New York, NY, USA, 1021–1030. https://doi.org/10.1145/1978942.1979095

[5] Harry Brignull and Yvonne Rogers. 2003. Enticing people to interact with large public displays in public spaces. In *Proceedings of the IFIP TC13 International Conference on Human-Computer Interaction (INTERACT '03)*. Springer, Berlin-Heidelberg, 17–24.

[6] Luigina Ciolfi, Mikael Fernström, Liam J Bannon, Parag Deshpande, Paul Gallagher, Colm McGettrick, Nicola Quinn, and Stephen Shirley. 2007. The shannon portal installation: Interaction design for public places. *Computer* 7 (2007), 64–71.

[7] Jorgos Coenen, Niels Wouters, and Andrew Vande Moere. 2016. Synchronized Wayfinding on Multiple Consecutively Situated Public Displays. In *Proceedings of the 5th ACM International Symposium on Pervasive Displays* (Oulu, Finland) *(PerDis '16)*. ACM, New York, NY, USA, 182–196. https://doi.org/10.1145/2914920.2929906

[8] Marco Cristani, Loris Bazzani, Giulia Paggetti, Andrea Fossati, Diego Tosato, Alessio Del Bue, Gloria Menegaz, and Vittorio Murino. 2011. Social interaction discovery by statistical analysis of F-formations.. In *BMVC*, Vol. 2. 4.

[9] Nigel Davies, Sarah Clinch, and Florian Alt. 2014. *Pervasive Displays - Understanding the Future of Digital Signage.* Morgan and Claypool.

[10] Tiare Feuchtner, Robert Walter, and Jörg Müller. 2016. Interruption and Pausing of Public Display Games. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI '16)*. ACM, New York, NY, USA, 306–317. https://doi.org/10.1145/2935334.2935335

[11] Saul Greenberg, Nicolai Marquardt, Till Ballendat, Rob Diaz-Marino, and Miaosen Wang. 2011. Proxemic Interactions: The New Ubicomp? *Interactions* 18, 1 (Jan. 2011), 42–50. https://doi.org/10.1145/1897239.1897250

[12] Bernd Huber, Joong Ho Lee, and Ji-Hyung Park. 2015. Detecting User Intention at Public Displays from Foot Positions. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 3899–3902. https://doi.org/10.1145/2702123.2702148

[13] Hannu Kukka, Heidi Oja, Vassilis Kostakos, Jorge Gonçalves, and Timo Ojala. 2013. What Makes You Click: Exploring Visual Signals to Entice Interaction on Public Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. ACM, New York, NY, USA, 1699–1708. https://doi.org/10.1145/2470654.2466225

[14] Paul Marshall, Yvonne Rogers, and Nadia Pantidi. 2011. Using F-formations to Analyse Spatial Patterns of Interaction in Physical Environments. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work* (Hangzhou, China) *(CSCW '11)*. ACM, New York, NY, USA, 445–454. https://doi.org/10.1145/1958824.1958893

[15] Nemanja Memarovic, Sarah Clinch, and Florian Alt. 2015. Understanding Display Blindness in Future Display Deployments. In *Proceedings of the 4th International Symposium on Pervasive Displays* (Saarbruecken, Germany) *(PerDis '15)*. Association for Computing Machinery, New York, NY, USA, 7–14. https://doi.org/10.1145/2757710.2757719

[16] Alexander Meschtscherjakov, Wolfgang Reitberger, Michael Lankes, and Manfred Tscheligi. 2008. Enhanced Shopping: A Dynamic Map in a Retail Store. In *Proceedings of the 10th International Conference on Ubiquitous Computing* (Seoul, Korea) *(UbiComp '08)*. Association for Computing Machinery, New York, NY, USA, 336–339. https://doi.org/10.1145/1409635.1409680

[17] 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* (Firenze, Italy) *(MM '10)*. ACM, New York, NY, USA, 1285–1294. https://doi.org/10.1145/1873951.1874203

[18] 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* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 297–306. https://doi.org/10.1145/2207676.2207718

[19] Thorsten Prante, Carsten Röcker, Norbert Streitz, Richard Stenzel, Carsten Magerkurth, Daniel Van Alphen, and Daniela Plewe. 2003. Hello.Wall–beyond ambient displays. In *Adjunct Proceedings of Ubicomp*. Citeseer, 277–278.

[20] Fiona Redhead and Margot Brereton. 2009. Designing Interaction for Local Communications: An Urban Screen Study. In *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part II* (Uppsala, Sweden) *(INTERACT'09)*. Springer-Verlag, Berlin, Heidelberg, 457–460.

[21] Lior Rokach and Oded Maimom. 2007. *Data mining with decision trees: theory and applications*. https://doi.org/10.1007/978-0-387-09823-4

[22] Hasibullah Sahibzada, Eva Hornecker, Florian Echtler, and Patrick Tobias Fischer. 2017. Designing Interactive Advertisements for Public Displays. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1518–1529. https://doi.org/10.1145/3025453.3025531

[23] Daniel J Simons and Michael S Ambinder. 2005. Change blindness: Theory and consequences. *Current directions in psychological science* 14, 1 (2005), 44–48.

[24] Fabius Steinberger, Marcus Foth, and Florian Alt. 2014. Vote With Your Feet: Local Community Polling on Urban Screens. In *Proceedings of The International Symposium on Pervasive Displays* (Copenhagen, Denmark) *(PerDis '14)*. Association for Computing Machinery, New York, NY, USA, 44–49. https://doi.org/10.1145/2611009.2611015

[25] Karolina Szymbor. 2015. *The Interactive Touch Wall at the Copenhagen Airport in Human-Computer Interaction Perspective: Evaluation of The User Experience*. Master Thesis at Aalborg University.

[26] Maurice Ten Koppel, Gilles Bailly, Jörg Müller, and Robert Walter. 2012. Chained Displays: Configurations of Public Displays Can Be Used to Influence Actor-, Audience-, and Passer-by Behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. ACM, New York, NY, USA, 317–326. https://doi.org/10.1145/2207676.2207720

[27] Nina Valkanova, Robert Walter, Andrew Vande Moere, and Jörg Müller. 2014. MyPosition: Sparking Civic Discourse by a Public Interactive Poll Visualization. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work Social Computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 1323–1332. https://doi.org/10.1145/2531602.2531639

[28] Daniel Vogel and Ravin Balakrishnan. 2004. Interactive Public Ambient Displays: Transitioning From Implicit to Explicit, Public to Personal, Interaction with Multiple Users. In *Proceedings of the 17th annual ACM Symposium on User Interface Software and Technology* (Santa Fe, NM, USA) *(UIST'04)*. ACM, New York, NY, USA, 137–146. https://doi.org/10.1145/1029632.1029656

[29] Robert Walter, Gilles Bailly, and Jörg Müller. 2013. StrikeAPose: Revealing Mid-Air Gestures on Public Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 841–850. https://doi.org/10.1145/2470654.2470774

[30] Niels Wouters, John Downs, Mitchell Harrop, Travis Cox, Eduardo Oliveira, Sarah Webber, Frank Vetere, and Andrew Vande Moere. 2016. Uncovering the Honeypot Effect: How Audiences Engage with Public Interactive Systems. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (Brisbane, QLD, Australia) *(DIS '16)*. ACM, New York, NY, USA, 5–16. https://doi.org/10.1145/2901790.2901796

[31] Yanxia Zhang, Jörg Müller, Ming Ki Chong, Andreas Bulling, and Hans Gellersen. 2014. GazeHorizon: Enabling Passers-by to Interact with Public Displays by Gaze. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Seattle, Washington) *(UbiComp '14)*. ACM, New York, NY, USA, 559–563. https://doi.org/10.1145/2632048.2636071