ORIGINAL RESEARCH

Spatial and Temporal Audience Behavior of Scrum Practitioners Around Semi-Public Ambient Displays

Jan Schwarzer[a] and Susanne Draheim[a] and Kai von Luck[a]

[a]Department of Computer Science, Hamburg University of Applied Sciences, Berliner Tor 7, 20099 Hamburg, Germany

ARTICLE HISTORY
Submitted: Jan 26, 2022; Revised: May 14, 2022; Accepted: Jul 5, 2022

ABSTRACT
Exploring spatial and temporal audience behavior around ambient displays is an important area of HCI research. It aids in, for instance, understanding better user appropriation in natural environments. However, there are only a few tools to capture said behavior and, simultaneously, little knowledge on the space around ambient display installations exists. In this research, we report on audience behavior observed during an in-the-wild study where we deployed our custom Ambient Surfaces solution in a professional, large-scale agile software development context for circa 5 years. Across 18 weeks in 2017, we collected skeletal data with two Microsoft Kinect v2 cameras resulting in behavior information of more than 30,000 passersby. Our results indicate, among others, that users did show the highest levels of engagement at quite some distance to the Ambient Surfaces and that people engaging in direct interaction did so rather purposely. Ultimately, this article encapsulates our research's originality in four contributions including an approach to separate passersby from real users and an in-depth exploration of skeletal data. With the tools and methods illustrated, we hope demonstrating manifold insights for future research on audience behavior tracking.

KEYWORDS
Audience behavior; ambient displays; skeletal data; longitudinal research; in-the-wild research

1. Introduction

Embedding interfaces *in the wild*, meaning placing them stationary in real-world public and semi-public spaces (Williamson & Williamson, 2017), is one of the key aims of ubiquitous computing (Dalton, Dalton, & Hoelscher, 2015). Nowadays, interfaces have advanced past the desktop metaphor and emphasize on touch and gesture interactions (Stephanidis et al., 2019), while being, in fact, ubiquitous (Boucher, Blumenstein, de Jesus Oliveira, & Seidl, 2021). With the rise of the post-desktop era, the spatial variable in HCI research has become an integral part in the overall interaction process (Dalton et al., 2015). Large and interactive displays, or *ambient displays* as we refer to, are

This is an Accepted Manuscript of an article published by Taylor & Francis in the International Journal of Human–Computer Interaction on 5 Aug 2022, available online:
https://www.tandfonline.com/doi/10.1080/10447318.2022.2099238. Contact: jan.schwarzer@haw-hamburg.de.

no exception to this development in ubiquitous computing. An increasing number of deployments in public (e.g., city settings) and semi-public (e.g., offices) environments can be observed (Ardito, Buono, Costabile, & Desolda, 2015) resulting in a notable contextual diversity of installations (Germany, Speranza, & Anthony, 2019). Commonly, ambient displays are evaluated through understanding their audience behavior (Elhart, Mikusz, Mora, Langheinrich, & Davies, 2017), while *behavior* here means a performance of some kind that people carry out in front of an installation (Williamson & Williamson, 2014). Past research, however, was more concerned with aspects such as a technology's usability (Williamson & Williamson, 2014) leading to a lacking in understanding of how ambient displays affect audience behavior physically, socially, and culturally (Ardito et al., 2015). Unsurprisingly, ambient displays were often found to be irrelevant to the space where they were deployed and, simultaneously, it remains unclear how they are really appropriated in the wild (Parker, Tomitsch, & Kay, 2018). While studies attempted to model spatial and temporal audience behavior in the past (Shi & Alt, 2016), it remains a challenging endeavor (Ardito et al., 2015).

In more recent years, research has been building on the advances in depth-image cameras to scrutinize audience behavior in the wild. It has underlined the importance of understanding better how people move around a display installation (Elhart et al., 2017) to, for instance, learn more about users who wish not to interact (Williamson & Williamson, 2014). The premise is that by investigating the space more holistically, research can come up with more sophisticated tools, methods, frameworks, and theories (Dalton et al., 2015). First tools have already been developed to both increase the automation throughout the research process and reduce the workload resting upon scientists (Mäkelä, Heimonen, & Turunen, 2018). In contrast to manual observations and interviews allowing for the analysis of tens of users, data from up to thousands of users could now be automatically investigated (Williamson & Williamson, 2014). While these tool-based approaches are considered more cost-effective (Mäkelä et al., 2018), they can also be easily used in other deployments and be readily integrated with other methodologies (Williamson & Williamson, 2014). Research on ambient displays in the wild, however, remains a challenge (Mäkelä et al., 2018; Williamson & Williamson, 2017) and warrants methodological development (Schwarzer, Draheim, von Luck, Wang, & Grecos, 2021). Not many tools exist to capture audience behavior (Elhart et al., 2017) and new tools to accurately record and analyze interaction in a wider context are much called for (Dalton et al., 2015). Also, depth-based data, particularly skeleton data, has not been utilized to its full capacity (Mäkelä et al., 2018).

In response, this article draws attention to a recent study (Schwarzer et al., 2021), where we deployed two of our custom ambient display solutions—henceforth referred to as *Ambient Surfaces*—in a professional, large-scale agile software development (ASD) environment for roughly 5 years. Each of the systems was equipped with a Microsoft Kinect v2 sensor operational for 18 weeks in 2017. During this time, skeletal data of more than 30,000 passers-by was gathered. Building on this data set, we elaborate here on the tools utilized and the methods leveraged to expand on both spatial and temporal audience behavior. Ultimately, our research's originality is discussed alongside four individual contributions. First, this study proposes a still missing approach to separate passersby from real users, including their levels of engagement. Second, our research suggests incidents of direct interaction that have not been, so far, considered in existing behavioral models. Third, with behavioral data distilled from a professional context, our study contributes rich nuances to existing work that has been largely informed by research in public environments. Finally, with the illustrated steps of collecting, preprocessing, and analyzing skeletal data, our work fundamentally adds to its exploration.

The article is organized as follows: Section 2 elaborates the related literature with respect to both studies modeling spatial and temporal audience behavior and studies introducing tools to track said behavior. Afterwards, Section 3 introduces the Ambient Surfaces solution and the research context. Subsequently, Section 4 illustrates the methods utilized for the data collection and the analysis process. Further, Section 5 presents the results obtained, whereas Section 6 discusses the originality of our work including research implications, recommendations of future work, and research limitations. Finally, Section 7 concludes the article.

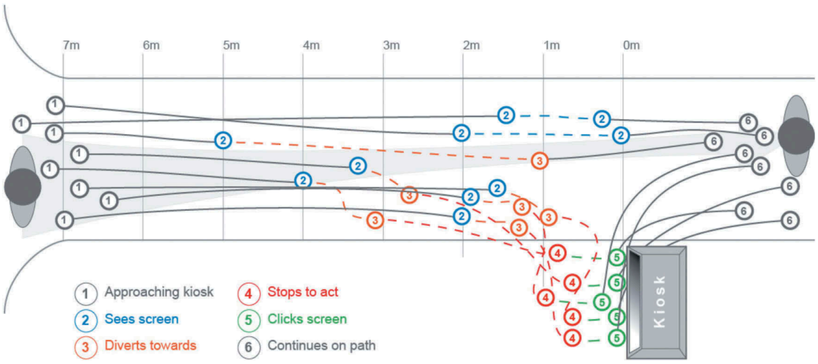
2. Related work

While a vast amount of past research has already used video analysis techniques such as face recognition and eyeball detection (Elhart et al., 2017), we concentrate here on studies taking on the challenge of investigating audience behavior in the wider context (Williamson & Williamson, 2014). Initially, it is emphasized on the fundamentals of spatial and temporal interaction models, including an introduction to seminal work. Then, the focus is on studies that leveraged camera-based approaches to track people in the space of ambient display installations and, in instances, applied examples of said models. Finally, our work is related to existing research.

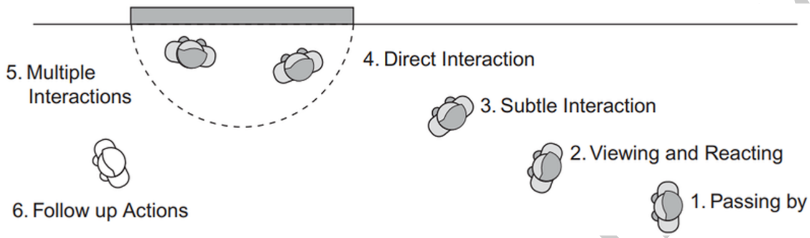
2.1. Spatial and temporal interaction models

In the early 2000s, the first spatial and temporal interaction models appeared. Since then, many models have been suggested to articulate audience behavior (Germany et al., 2019). Two types have ultimately emerged (see Figure 1): *spatial* and *temporal* models (Davies, Clinch, & Alt, 2014). While the former considers how users behave in light of the spatial relationship to a display, the latter considers how the engagement of users with a display changes over time. Whereas existing models share common traits, they are typically derived from different phenomena and contexts (Germany et al., 2019). A similarity of existing models is their emphasis on single-surface installations (Boucher et al., 2021) and their foundation dividing the interaction process in multiple phases (Michelis & Müller, 2011). While the initial phase typically considers cases where people ignore a display entirely, the last phase takes into account instances where people engage in close or personal interaction. Evidently, however, more recent research builds on advances regarding camera-based sensors. While during the early 2000s lab-based research was required to equip users with separate hardware devices to track their spatial behavior (e.g., Streitz, Röcker, Prante, Stenzel, & van Alphen, 2003), newer research uses sensors such as Kinect cameras to do the same in more natural environments (e.g., Elhart et al., 2017).

Fundamentally, spatial models have in common that they define interaction zones based on a user's distance to a display (Alt, Buschek, Heuss, & Müller, 2021). The earliest attempt to model spatial audience behavior was authored by Streitz et al. (2003) and by Prante et al. (2003). Their work, essentially, suggested three zones of interaction, including an active zone directly in front of a display, a notification zone where users can be proactively attracted, and an ambient zone where people are provided with general information. Both solutions *GossipWall* (Streitz et al., 2003) and *Hello.Wall* (Prante et al., 2003) were built as an informative art installation us-



(a) Observed spatial audience behavior with a display showing a text message and an animated effect to attract users (Germany et al., 2019). The installation was located on a university campus.



(b) Captured temporal audience behavior with displays installed behind public storefront windows in a city installation (Michelis & Müller, 2011). The displays visualized, among others, a mirror image of the environment in front of them.

Figure 1.: Two exemplary studies attempting to model spatial (top) and temporal (bottom) audience behavior.

ing light patterns to emit information. Using RFID technology, both systems created distance-dependent semantics. Were people, for example, tracked nearer to the installation, the *GossipWall* and the *Hello.Wall* emitted information such as personal light patterns. Subsequently, said three-zone model was used and extended in later work. For instance, Vogel and Balakrishnan (2004) introduced a spatial model ranging from distant implicit interaction (e.g., subtle cues such as body orientation) to up-close explicit personal interaction (e.g., hand and touch gestures). Their model focuses on fluid transitions between phases (e.g., opportunity to quickly and seamlessly initiate and end interactions) and suggests a wider range of interaction techniques. Vogel and Balakrishnan (2004) leveraged a motion tracking system with sensors attached to a user's body to measure distances. Another example is the work from Greenberg, Marquardt, Ballendat, Diaz-Marino, and Wang (2011) based on the seminal theory of proxemics (Hall, 1966). In ubiquitous computing, as Greenberg et al. (2011) note, proxemics concern inter-entity (i.e., a mix of people, digital devices, and non-digital things) distances. The goal is to in one way or another sense the proximity between these entities. However, the previous models primarily focus on single-user interaction and do not support back-and-forth transitions (Davies et al., 2014). In contrast, the model by Memarovic et al. (2012) allowed to elaborate multi-user interactions and transitions. It builds on a principal human need in public spaces—i.e., to engage passively and actively as well as to discover. Finally, Germany et al. (2019) partially build on the work from Vogel and Balakrishnan (2004) and created models based on different implicit (i.e., an animated effect) and explicit (i.e., a message inviting interaction) screen prompts (see Figure 1a). They investigated how increased attention and awareness affect the engagement with their display. Using an ultrasonic range finder sensor, distances to the display were determined.

Opposed to spatial models, temporal models consider interaction as a movement process of users through different stages (Alt et al., 2021). Each of these stages can only be reached by overcoming a certain threshold. The *public interaction flow* model by Brignull and Rogers (2003) is an early conceptual example. The authors' work was, among others, motivated by the question of how groups of people socialize around ambient displays. The principal idea is that people's interest must be stimulated enough and the system has to provide affordances about what it offers. Simultaneously, people have to be willing to spend their time and effort. Another example is the *Audience Funnel* framework (see Figure 1b) introduced by Michelis and Müller (2011). This framework is considered one of the most influential ones (Mäkelä et al., 2018) and was motivated by the fact that no quantitative data had been collected of how many people pass through the individual thresholds of a given model's phases. The authors pursued a better understanding on how to improve conversion rates between phases by indicating where high numbers of people drop out. At each transition between these phases, certain amounts of dropouts can be observed, hence only a percentage of people proceed to the next phase. Latest research, additionally, took into account the contemporary ubiquity of interactive surfaces. For example, the *Multi-Device Interaction Model* (Boucher et al., 2021) is driven by this development and considers people carrying around their own personal devices and bringing them into public spaces. In a museum context, the authors observed the interaction focus shifting away from single, large display installations and interaction zones melding together.

2.2. Examples of audience behavior tracking

The examples presented in this section follow the notion that some automation of the audience behavior evaluation process can be realized (Elhart et al., 2017). Specifically, to monitor audience behavior in the vicinity of a display installation, these studies leveraged approaches based on anonymous depth data and computer vision algorithms. The cameras deployed were configured in a bird's eye setup or were placed near the display facing passersby. For example, Mäkelä et al. (2018), on a more methodological note, present a semi-automatic evaluation process for longitudinal deployments. In their study, the authors collected interaction and skeletal data recorded with a Kinect v1 sensor. Their gesture-controlled display solution *Information Wall* was deployed for 1 year on a university campus and data of a total of more than 100,000 passersby could be gathered. Mäkelä et al. (2018) applied the Audience Funnel framework and found, among others, that most users were first-time users. Their approach benefits the reduction of resources, the capability of privacy-preserving and semi-automatic analyses, and the study of effects in longitudinal deployments. Another example is the work from Elhart et al. (2017), who applied computer vision algorithms to raw depth frames—also recorded with a Kinect v1 sensor. The authors used and evaluated their tool *Audience Monitor*, an open source tracking tool to capture audience behavior. With this tool, Elhart et al. (2017) were reportedly able to detect passersby with high accuracy on average. Audience Monitor was operated for 52 days (1,248 hours) in a university canteen and detected 40,763 passersby in total. Different kinds of visualizations were created by the tool to prepare the results (e.g., bar chart diagrams), while the overall analysis process was still very labor-intensive (e.g., querying log files). In the study from Williamson and Williamson (2014), computer vision techniques were used as well. The authors introduced *Tracker Tool*, a solution to capture pedestrian traffic, based on motion detection and background subtraction. With the help of this tool, they were able to track passersby and their walking paths. In the context of a field study of two public installations, pedestrian data of 4 hours in total could be gathered. Additionally, 2 hours of baseline data and 1 hour of data with a street musician were collected. It was found, among others, that relocating a display does affect walkways of pedestrians and that passersby changed directions because a display was in close vicinity to their direction of travel. A few years later, Williamson and Williamson (2017) concentrated on the investigation of the experimenter role and how it affects pedestrians. In this study, the authors did not focus on understanding interaction but on comparing different evaluation approaches. To this end, their *Silly Hats Only* solution, a playful gesture-controlled display, was deployed. Silly Hat Only prompted passersby to perform a teapot gesture and displayed their silhouette. Overall, 32 hours of video material were collected including 20 hours of baseline material (12,213 pedestrians in total). A Microsoft Kinect v2 camera was leveraged and data was processed using OpenNI libraries. Williamson and Williamson (2017) revealed that, for example, overt observations resulted in a substantially lowered conversion rate of passersby that spend 5 seconds or more at the display (compared to covert observations). Finally, attention is drawn to the study from Shi and Alt (2016). While no results are presented in this study, the introduced *Anonymous Audience Analyzer* tool shows promising means to track audience behavior. The tool uses virtual reality technology and allows for scrutinizing skeletal data from one to many Microsoft Kinect sensors. With the help of different visualized observational perspectives, Anonymous Audience Analyzer enables researchers to view the scene in front of a display from a user's or bird's eye point of view. It further allows to replay recorded material at arbitrary speed.

2.3. Relating to existing research

Considering the studies presented, the work from Mäkelä et al. (2018) arguably informed our research the most. We similarly placed the Kinect sensors directly facing passersby and our study also heavily builds on skeletal data. Additionally, the aforesaid study encouraged us to adapt the Audience Funnel framework to organize findings regarding user engagement. Methodologically, their semi-automated evaluation process demonstrated insightful ideas on how to process skeletal data and guided us in the early stages of our research. Other studies, however, informed our research as well. Elhart et al. (2017) inspired us to similarly prepare heat map visualizations to investigate data showing spatial conditions. The work from Williamson and Williamson (2014) resulted in different metrics applied in said heat maps such as the distance of passersby and in the idea to analyze curvatures of pathways. Furthermore, as demonstrated in spatial models, the space in front of the Ambient Surfaces was divided into different zones following metrics suggested by Michelis and Müller (2011).

Our research, however, differs from existing work in two fundamental ways. First, it considers behavioral data stemming from a semi-public, professional environment, while the majority of audience behavior tracking research can be associated with more public contexts such as walkways in cities (e.g., Williamson & Williamson, 2014) and university sites (e.g., Elhart et al., 2017). Instead of unveiling behavior of a notable

Figure 2.: Both Ambient Surfaces at their installation location, equipped with two Microsoft Kinect v2 sensors.

amount of first-time users, the audience behavior distilled here shows insights on how ASD professionals appropriate ambient displays in their everyday work life. Results allow for inspecting authentic behavior beyond issues known in ambient display research such as the *novelty effect* (Koch, von Luck, Schwarzer, & Draheim, 2018). Second, differences become apparent in the principles behind the display solutions. In contrast to, for instance, Mäkelä et al. (2018) and Germany et al. (2019), it was fundamentally not aimed at repeatedly attracting passersby to encourage engagement with the Ambient Surfaces but to display information that ASD team members deemed relevant to be conveyed. Patterns obtained were intended not to be a result of somewhat biased actions caused by any attention-grabbing features our solution could offer.

3. Research prototype and context

Our Ambient Surfaces solution (see Figure 2) represents a sub-class of ambient displays that leverages screen-based technology and targets "supporting informal, nonurgent communication, collaboration, and awareness" (Huang, Mynatt, Russell, & Sue, 2006, p. 37). It consists of three components: a custom software application, interactive displays, and compact desktop computers. The software handles touch interactions with the displays and creates the corresponding visualizations. It builds on the *Microsoft .NET framework*[1], specifically on its graphical subsystem *Windows Presentation Foundation (WPF)*[2] to render user interfaces. We used this software suite because we have gathered experiences with it during first, preliminary field deployments and accumulated a notable amount of practical knowledge with it throughout past research (e.g.,

[1] https://docs.microsoft.com/en-us/dotnet/
[2] https://docs.microsoft.com/en-us/dotnet/framework/wpf/

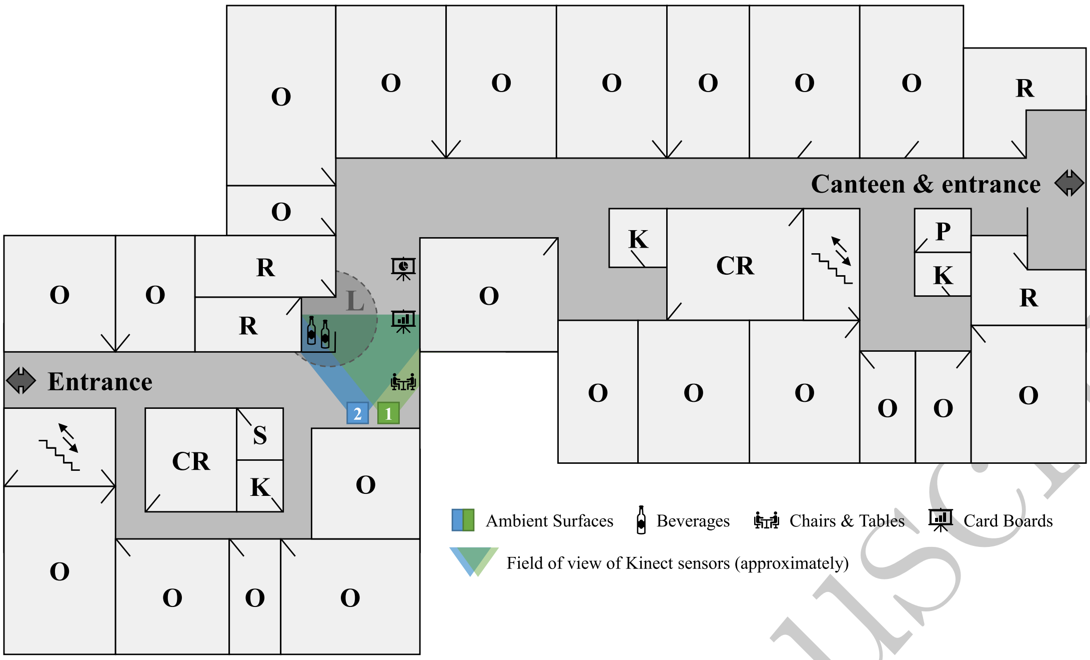


Figure 3.: The ground level floor plan of the department's two-story building. The two squares indicate the locations of both Ambient Surfaces (*System 1* on the right; *System 2* on the left). Abbreviations: CR = Conference Room; K = Kitchen; L = *Lobby*; O = Office; P = Printers; S = Storage Room; R = Restrooms.

Schwarzer et al., 2016). The design process was guided by the related literature. A central puzzle of the design was, as Parker, Tomitsch, Davies, Valkanova, and Kay (2020) highlight, the value that the Ambient Surfaces could potentially deliver to users. For instance, we geared toward creating some kind of ownership and control about the messages being spread (Parker et al., 2020). Simultaneously, the interaction modalities were kept to a minimum (i.e., scrolling gestures and selection) to avoid frustration of users (Ardito et al., 2015). With data stemming from a variety of different sources, we also tried to prevent issues seen with awareness tools that yield information from only one data source such as incomplete information (Ye, Ye, & Liu, 2018). Both displays (each $\geq$ 46 inches in size) were mounted on a rack with rolling wheels in a landscape configuration. The total height of each installation was circa 1.80 meters, while both displays provided a 1080p resolution (i.e., 1,920 $\times$ 1,080 pixels) and infrared touch sensors (up to 2 and 32 touches respectively). Two compact desktop computers operated the displays and ran the software application.

At the time of this research, two Ambient Surfaces (labeled *System 1* and *System 2*) were deployed in the ASD department of a German company where roughly 70 to 80 people were employed. The department's organization could be characterized as a professional, large-scale ASD context (Dingsøyr, Fægri, & Itkonen, 2014) as to, during the entire study, the total number of agile teams varied between four to eight teams. The staff was equipped with a variety of tools that assisted them in developing the company's custom software product. For the context of this research, the following tools deserve a special mention: *Atlassian Jira* (e.g., used to store user stories), *Atlassian Confluence* (e.g., utilized to share architectural decisions), *Jenkins* (used for continuous integration purposes), *GoCD* (leveraged to automate the build and deployment infrastructure), *Tetris* (a custom tool to visualize test summaries), and *Avatar* (also a custom tool to highlight graphical test metrics). Both Ambient Surfaces were situated in a common area (see Figure 3) which was selected because it seemed to provide the chance of opportunistic interactions (Ardito et al., 2015; Parker et al., 2020). Card boards, beverages, tables, restrooms, and chairs were in the vicinity of both systems. In Figure 3, we highlight one area, labeled *Lobby*, which we will repeatedly refer to below. The area *Lobby* highlights the space in which people went in and out from the restrooms and fetched beverages.

In total, eight information views were deployed between 2014 and 2019 that built on the aforementioned tools. At the time of the present research, six of the eight views were available. *System 1* contained the views labeled *Team Charts* (based on Atlassian Jira), *Confluence*, and *Bug Survey* (based on Atlassian Jira), while *System 2* included the views named *Jenkins*, *Test Suites* (i.e., based on Tetris), and *GoCD*. The software application displayed either two of these views at a time or one view on its own, depending on the data being presented. Overall, the views leveraged different means to present data, whereas some views were entirely custom-made. For some of the views, the available APIs provided the foundation to gather information (e.g., the *GoCD* view), while for other views existing web interfaces were incorporated as is on the Ambient Surfaces (e.g., the *Confluence* view).

4. Methods

In this section, we present the procedures, software tools, and hardware utilized for the data collection and analysis process.

4.1. Data collection

We decided to use Microsoft Kinect cameras in our research because they were, back in 2017, reported to be state-of-the-art depth sensors (Elhart et al., 2017). Specifically, we utilized two Kinect v2 sensors (analogously to *System 1* and *System 2* labeled *Camera 1* and *Camera 2*), whose depth sensors have a resolution of 512 × 424 pixels, operate at 30 Hz, and can track up to 6 people at a time. Both cameras were mounted on both Ambient Surfaces roughly 1,80 meters above ground (see Figure 2), were tilted down at circa 13 degrees, and were approximately 0.85 meters apart (edge to edge).

We created *AmbiLogger* (see Figure 4) to automatically collect skeletal data. It was developed in WPF and combined custom source code with source code from the different available examples in the *Kinect for Windows SDK 2.0*[3]. The tool's resulting data is, similar to the data in Mäkelä et al. (2018), anonymous and contains information of coordinates in a 3D space. Furthermore, AmbiLogger was configured to gather data from only one specific skeleton joint—the *Spine Shoulder* joint (see Figure 4). On the one hand, we wanted to limit the expected large amounts of data and, on the other hand, we wished to concentrate on the general location of people in front of the Ambient Surfaces. We, fundamentally, experienced this body joint to be less affected by body rotations. Compared to arm and hand joints, this joint was less often lost during tracking. AmbiLogger also allowed disabling the preview window to reduce resource allocations and could track up to 6 people simultaneously.

Between 6 a.m. and 8 p.m., AmbiLogger was continually executed on two separate

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

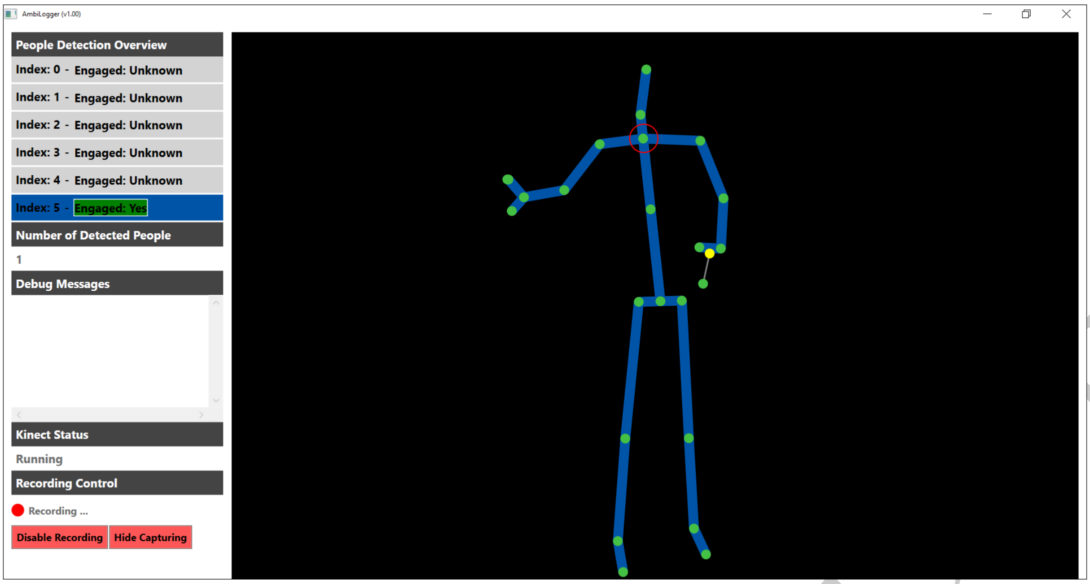


Figure 4.: The user interface of AmbiLogger while recording a scene with one person. Left: The control panel of the tool. Right: The preview window indicating the skeleton silhouette of a person. The red circle at the top indicates the *Spine Shoulder* joint, whose coordinates were the basis for analysis.

computers. Every time a person entered the cameras' field of view (circa 70 degrees horizontally and roughly 60 degrees vertically), an event to start a record was automatically triggered. A record ended, when the camera did not detect any more people in the field of view. Every record included zero to many frames, depending on the amount of data received. Listing 1 illustrates an exemplary frame extracted from one of the record text files. In total, data had been collected between weeks 14 and 31 in 2017. Over the course of 18 weeks, AmbiLogger archived a total of 97,618 individual records, including 34,682,188 million frames. *Camera 1* contributed 50,636 records (excluding five empty files) with 19,185,149 million frames resulting in material of circa 176 hours. *Camera 2* added 46,951 records (excluding 26 empty files) with 15,497,039 million frames (material of roughly 143 hours).

As the data collection process was conducted as part of a PhD study, the ethics committee of the awarding university was initially consulted to elaborate on issues to be considered. We explained what data was intended to be collected and how anonymity of users was thought to be ensured (e.g., the tracking of the *wearing glasses* property, included in Listing 1, was disabled). As a consequence, an ethics consent form to be signed was prepared and a handout describing the details of the data collection process was drafted (see handout below the right display in Figure 2). We then contacted the software department's management and discussed ethical concerns with them. As a result, the management internally consulted the workers' council, while also the staff was encouraged to raise any potential questions and concerns. Ultimately, the ethics consent form was signed by the management and we were allowed to deploy the two Kinect cameras, whereas we were limited to skeleton data and data that indicated, for instance, whether a person looked at the Kinect sensors.

```
2017-04-04 13:53:43.945 +02:00 #    // Timestamp
72057594037929268 #                 // Body tracking id
1019936248 #                        // Record id
5 #                                 // Body id (1-6)
1 #                                 // Number of people in frame
Unknown #                           // Property happy
No #                                // Property engaged
Unknown #                           // Property wearing glasses (disabled)
Unknown #                           // Property left eye closed
Unknown #                           // Property right eye closed
Unknown #                           // Property mouth open
Unknown #                           // Property mouth moved
Yes #                               // Property looking away
Unknown #                           // Property hand left state
NotTracked #                        // Property hand right state
0,3046448 #                         // x (meters)
0,2297372 #                         // y (meters)
4,262015 #                          // z (meters)
4,27906041643319                    // Distance to Kinect sensor (meters)
```

Listing 1: An exemplary frame gathered by the AmbiLogger tool and extracted from an archived record text file.

4.2. Data analysis

We touch on three topics in this section: first, the means chosen to analyze spatial and temporal audience behavior are introduced; second, the data set utilized is presented; and third, the topic of data filtering is elaborated on.

4.2.1. Analysis approaches

It can be complicated to capture aspects of spatial and temporal audience behavior (Elhart et al., 2017). We therefore consulted the literature and built our analysis on existing work. In the end, we followed three central themes: first, we aimed at creating means to explore the more general facets of the spatial audience behavior in front of the Ambient Surfaces; second, we targeted identifying people's entry and exit directions with respect to the cameras' field of views; and third, we pursued quantifying temporal aspects of the audience behavior.

General facets of spatial audience behavior We developed *AmbiMapper*, a console application to selectively choose skeletal data and to create custom heat map visualizations for the purpose of data exploration. These visualizations have a 1080p resolution and consist of 20,736 individual tiles (each 10 × 10 pixels in size), a scaling in meters, nineteen different color increments to differentiate values, and four highlighted zones (labeled *Z1*, *Z2*, *Z3*, and *Z4*). These zones aided investigating spatial behavior in relation to the different distances (d) users had with respect to both Ambient Surfaces. We used metrics presented by Michelis and Müller (2011) to define these zones. The authors describe that direct interaction typically to occur in a small area around a display of about 1 meter, while they specify a *passerby* as everyone that is in a radius of 4 meters. Between both these scales, we evenly divided the four zones: (1) zone *Z1*: $d \leq 1$ meter; (2) zone *Z2*: 1 meter $< d \leq 2$ meters: (3) zone *Z3*: 2 meters $< d \leq 3$ meters; (4) zone *Z4*: 3 meters $< d \leq 4$ meters.

We used different metrics in the heat map visualizations which we will henceforth refer to as modes *M1* to *M4*. Any of these modes has a central measurement upon which a heat map visualization was created and that was calculated on a per-tile basis. Depending on these metrics, the background color of each individual tile was configured. Fundamentally, the four modes *M1*–*M4* were analyzed in relation to the four zones *Z1*–*Z4*, while they address different questions:

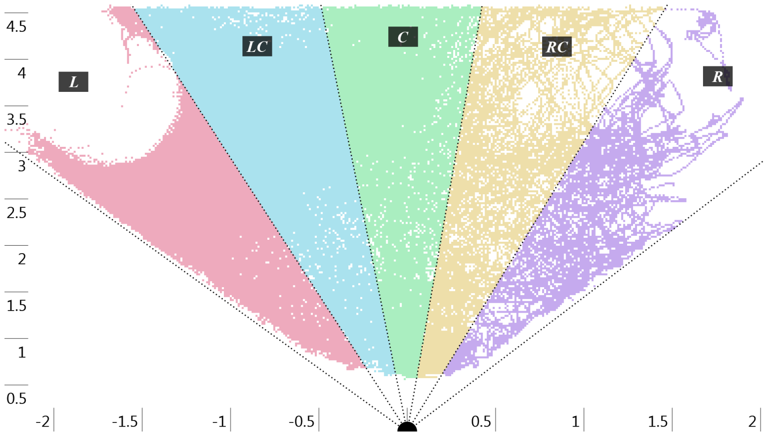


Figure 5.: A visualization of exemplary data from *Camera 1* split into the five entry and exit direction areas *L–R* (scaling in meters).

- *M1* concerns the question of in which areas were most of the frames collected (measure: the number of frames).
- *M2* targets the question of in which areas people spent most of their time (measure: the number of frames in relation the number of records).
- *M3* addresses the question of in which areas people were most notably walking past (measure: the number of tracking ids detected).
- *M4* aims at the question of in which areas people showed strong levels of engagement with the Ambient Surfaces. To this end, we leveraged one specific .NET framework property—the property *Engaged*. This property indicates whether it appears that a person is engaging with the Kinect sensor. Values for this property are *Yes*, *No*, *Maybe*, and *Unknown* (measure: the number of *Yes* values).

Spatial audience behavior: Entry and exit directions To investigate how individuals and groups of individuals walked through the scene in front of both Ambient Surfaces, we divided the horizontal field of view of both cameras in five evenly-sized areas (see Figure 5). These areas are labeled as follows: *Left* (*L*), *Left to center* (*LC*), *Center* (*C*), *Right to center* (*RC*), and *Right* (*R*). For each record, we considered both the first and the last frame and calculated the corresponding areas in relation to the location of the Kinect sensors (i.e., the coordinate system's origin). To this end, we used the *arctan* function given that a frame's x and z coordinates were known in advance to map a calculated angle to the five mentioned areas.

Temporal facets of audience behavior We, similarly to Mäkelä et al. (2018) and Williamson and Williamson (2017), leveraged the Audience Funnel framework (Michelis & Müller, 2011), whereas we concentrated on the model's phases denoted as *Passing by*, *Viewing and Reacting*, and *Direct Interaction*. The reason was twofold: first, we wanted to limit the overall complexity of analyses. For instance, the fifth phase (i.e., *Multiple Interactions*) and the sixth phase (i.e., *Follow up Actions*) would have required additional time-intensive steps such as interviews; second, one phase (i.e., *Subtle Interaction*) was not applicable to our research. As Wouters et al. (2016) explain, applying existing models to other domains can unveil differences with respect to the existence or absence of certain parts. Specifically, as most employees were familiar

```
2017-04-04 13:58:38.107 +02:00 #     // Timestamp
72057594037929880 #                  // Body tracking id
2052127216 #                         // Record id
0,9871798 #                          // -x (meters, inverted)
1,593787 #                           // z (meters)
No                                   // Value of engaged property
```

Listing 2: An exemplary frame extracted from a revised record text file. The x-axis is inverted because the Kinect sensor interprets the scene from the user's point of view.

with both the presence of the Ambient Surfaces and the information conveyed since February 2014, we did not expect staff members to proactively cause reactions by the display (e.g., because they expected unknown interaction capabilities).

The analysis built on two specific metrics. On the one hand, on the mentioned .NET framework's tracking id, which is an identifier that is automatically assigned to a person when a Kinect sensor detects their presence. Every record can include one to many of these tracking ids. On the other hand, on the mentioned *Engaged* property. In combining both tracking ids and the *Engaged* property, we were able to deduce conclusions considering the three mentioned phases as follows:

(1) *Passing by*: Everyone who was in the vicinity of the Ambient Surfaces without paying any attention to one of the Kinect sensors (i.e., *Engaged* property values other than *Yes*).
(2) *Viewing and Reacting*: People who in some way seemed to show some level of engagement. To elaborate on this phase, we grouped the different levels of engagement in ten categories, while a category indicates a tracking id's amount of *Yes* values in relation to all other values of the *Engaged* property in percent.
(3) *Direct Interaction*: Every person who was in the close vicinity to both Ambient Surfaces (i.e., in zone *Z1*).

4.2.2. Selecting data for analysis

Initially, we tried to answer the question of what data to focus on. During this initial stage and like Mäkelä et al. (2018), we experienced our working machines sometimes running out of memory due to the sheer amount of data. We played with several parameters of the heat map visualizations and repeatedly revised resolutions, the scaling, and colors to name just a few of these parameters. As a result, we reduced the overall data required to be analyzed. Among others, we removed values from the y-axis, because we were not interested in, for instance, the size of a person. Instead we geared toward determining a person's location in a 2D plane (i.e., values of both the x-axis and z-axis). AmbiMapper crawled all archives stored by AmbiLogger and respectively created revised record files for *Camera 1* and *Camera 2* (see Listing 2).

In total, we applied six rules in the selection process. First, we removed all data recorded on weekends (16 records). Second, we did not consider data collected prior to 7:00 a.m. and after 6:00 p.m. to focus on the working hours (8,514 records) and, simultaneously, to avoid temperature influences—it is recommended to run the Kinect v2 sensor at least 25 minutes prior to capturing data (Wasenmüller & Stricker, 2017). Third, we removed all records that included tracking ids with a total of less than 15 frames to ensure a certain minimum of activity (5,208 records). Fourth, records were not considered if they incorporated tracking ids that showed 100 % of engagement. Due to the *Engaged* property's sensitivity, it seemed implausible to us that a person was looking at the Kinect sensors for the entire duration of a record (1,147 records). Fifth,

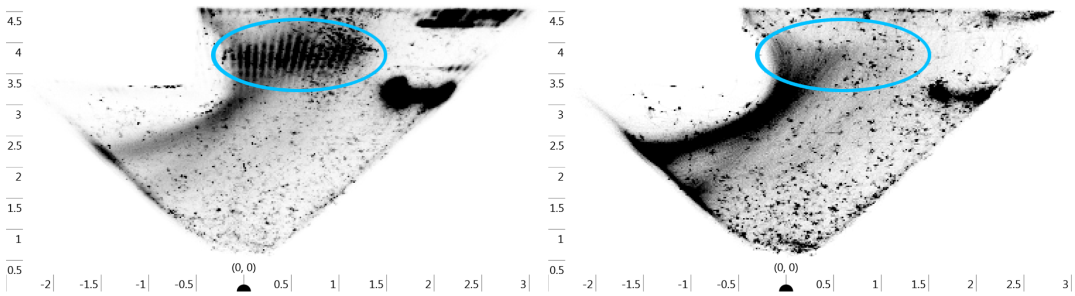


Figure 6.: Heat maps of data (scaling in meters) collected with *Camera 2* indicating areas with many frames (black) and areas with no frames (white). The amount of tiles in the heat maps was doubled to better contrast the erroneous data (see highlighted areas). Left: Data without any rules applied; Right: Data with the six rules plus a moving average filter applied.

we dealt with erroneous data. The Kinect v2 sensor can sometimes produce erroneous data in the form of erratic 3D coordinates (Mangal & Tiwari, 2020), jitter (Niu, Wang, Wang, & Ran, 2020), and noise in the form of flying pixels or multipath interference (Wasenmüller & Stricker, 2017). We found that *Camera 2* was particularly prone to erroneous data. Similar to Niu et al. (2020), we also experienced that occlusion or a subject slightly moving out of the camera's field of view seemed to cause these issues. Erroneous data was most notably observable in front of area *Lobby*. The resulting heat map visualizations showed vertically aligned patterns comparable to sand ripples (see Figure 6). To us, this observation seemed to best fit the description of flying pixels (i.e., erroneous depth estimates), which are a common problem of Time-of-Flight cameras (Wasenmüller & Stricker, 2017). For *Camera 1*, however, we did not visually observe this issue to a comparable extent. We, ultimately, decided to discard records, where huge jumps ($>$ 0,5 meters) on the x-axis or z-axis were measured (4,614 records). In doing so, we could remove the observed patterns. Finally, we excluded all records from analysis which showed no level of engagement (54,964 records). We did so mainly because we wanted to focus on user behavior that could be somewhat related to the Ambient Surfaces and not to random events such as spontaneous discussions in the vicinity of both systems. In total, roughly 24 % of records respectively remained (*Camera 1*: 11,989 records; *Camera 2*: 11,135 records). Table 1 summarizes the records that were used for the subsequent data filtering process.

4.2.3. Data filtering

Applying filters is suggested in the literature (Niu et al., 2020). In the end, we decided for a moving average filter with an interval of 20 values. We did so, because this interval seemed to provide a reasonable trade-off with respect to filtering the data and not losing nuances in the data for investigations such as walking paths analyses. Specifically, the filter was applied to the values of both the x-axis and z-axis. The result of the filtering process were eight text files that respectively included records according to their number of people detected (i.e., six files for 1–6 people, one file for more than 6 people, and one file for all the data). We note, however, that there is additional research required with respect to more sophisticated filter techniques. We chose the moving average filter primarily because it is easily implemented, while also being capable of improving data quality by up to 21 % (Niu et al., 2020).

Table 1.: Records summary for *Camera 1* and *Camera 2*.

	Camera 1	***Camera 2***
Records in total	11,989 (1,960,886 frames)	11,135 (1,807,892 frames)
People detected (within 4 meters)	16,187	14,384
***Engaged* property *Yes* values (median)**	9.84 %	13.04 %
Records with 1 person	9,425	9,027
Records with 2 persons	1,553	1,354
Records with 3 persons	525	416
Records with 4 persons	257	169
Records with 5 persons	112	97
Records with 6 persons	50	36
Records with 6+ persons	67	36
Duration (min)	0.42 seconds (15 frames)	0.41 seconds (15 frames)
Duration (max)	0.34 hours	0.12 hours
Duration (overall)	17.86 hours	16.48 hours
Duration (median)	1.95 seconds	2.48 seconds

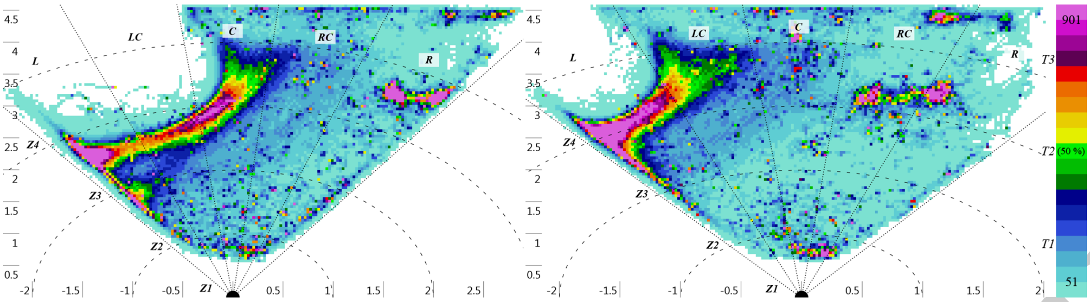
5. Results

The results presentation is organized analogously to the earlier indicated three central themes of analysis. Initially, results regarding the more general facets of spatial audience behavior are presented. Then, an overview of entry and exit directions is provided. Lastly, the temporal facets of the audience behavior are illustrated.

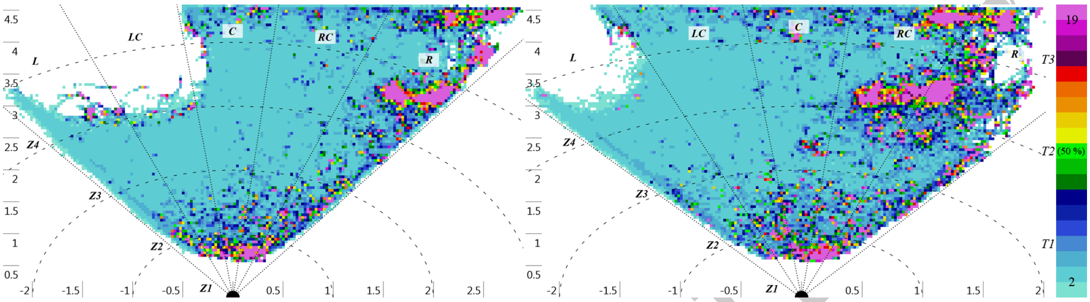
5.1. General facets of spatial audience behavior

Figure 7 shows the heat maps according to modes *M1–M4*. It is noted that the data of both Kinect sensors contained a different scaling on the x-axis, while the heat maps were mapped to a 1,080p resolution hence zone boundaries and the field of view vary. However, *Camera 1* detected 1,665,013 frames in zones *Z1–Z4* (*Camera 2*: 1,599,574 frames). Of the total number of tiles included in each visualization (i.e., 20,736 tiles), the data of *Camera 1* recorded values for 9,296 tiles within the radius of zones *Z1–Z4*, while 7,858 tiles got no corresponding data in this area (*Camera 2*: 7,288 tiles with a value; 9,356 tiles without one). Beyond zones *Z1–Z4*, *Camera 1* stored values for 2,499 tiles and no values were available for 1,083 tiles (*Camera 2*: 2,177 tiles with a value; 1,915 tiles without a value). In sum, the data of *Camera 1* resulted in 11,795 tiles (circa 57 %) having values (Camera 2: 9,465 tiles, roughly 46 %). It can be observed that a concrete wall near to *Camera 2* resulted in an increased number of tiles without any values (see zone *Z4* in Figure 7a). The aforesaid 9,296 tiles linked to *Camera 1* divide across zones *Z1–Z4* as follows: 343 tiles in *Z1*; 1,879 tiles in *Z2*; 3,199 tiles in *Z3*; and 3,875 tiles in *Z4*. Analogously, the 7,288 tiles relating to *Camera 2* break down as follows: 292 tiles in *Z1*; 1,630 tiles in *Z2*; 2,178 tiles in *Z3*; and 2,648 tiles in *Z4*. This total of 16,044 tiles defines the field of view of both cameras in zones *Z1–Z4*.

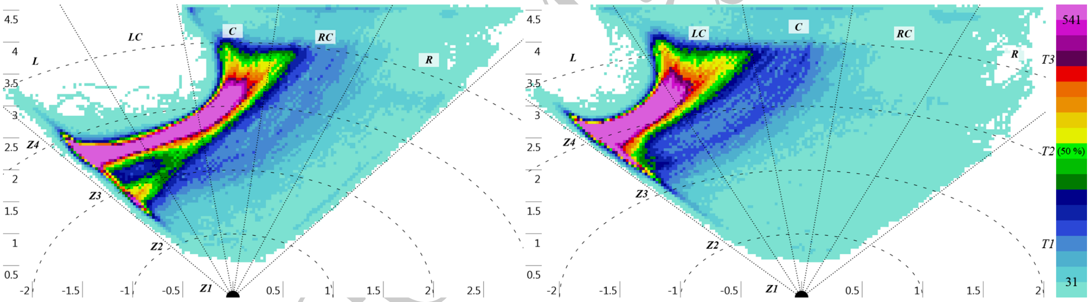
Table 2 presents percentage values that are a product of relating all tiles in zones *Z1–Z4* to a predefined threshold. It allows to deduce conclusions regarding the magnitude



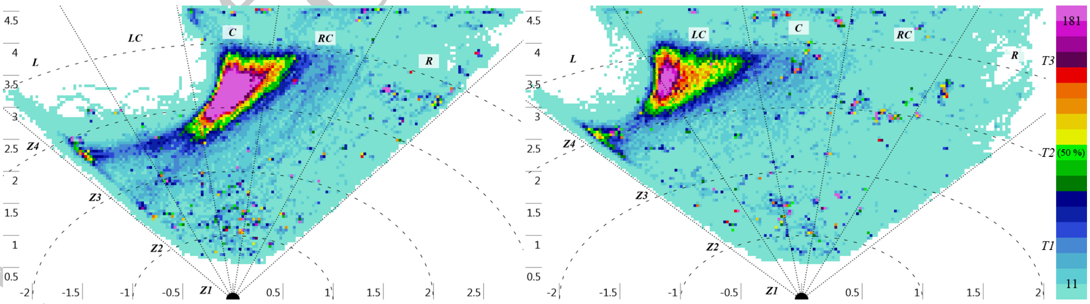
(a) Mode *M1*: Increment of 50 frames; 100 % = 901 frames or more.



(b) Mode *M2*: Increment of 1 ratio of frames to records; 100 % = 19 times more (or higher) frames than records.



(c) Mode *M3*: Increment of 30 tracking ids; 100 % = 541 tracking ids or more.



(d) Mode *M4*: Increment of 10 *Yes* values (*Engaged* property); 100 % = 181 *Yes* values or more.

Figure 7.: Heat map visualizations of both *Camera 1* (right) and *Camera 2* (left) according to modes *M1*–*M4* (scaling in meters). Zones *Z1*–*Z4* as well as the entry and exit direction areas *L*–*R* are indicated in each of the visualizations.

Table 2.: Percentage values of relating the total number of tiles in zones *Z1–Z4* to tiles meeting the thresholds *T1–T3* according to modes *M1–M4*.

		Camera 1				**Camera 2**			
		Z1	*Z2*	*Z3*	*Z4*	*Z1*	*Z2*	*Z3*	*Z4*
T1	*M1*	40 %	13 %	27 %	49 %	45 %	35 %	49 %	43 %
	M2	83 %	43 %	23 %	28 %	83 %	33 %	9 %	22 %
	M3	0 %	1 %	32 %	43 %	0 %	20 %	56 %	39 %
	M4	4 %	6 %	5 %	29 %	8 %	11 %	25 %	32 %
T2	*M1*	15 %	2 %	6 %	15 %	12 %	7 %	17 %	12 %
	M2	44 %	12 %	6 %	13 %	45 %	8 %	2 %	11 %
	M3	0 %	0 %	6 %	18 %	0 %	4 %	22 %	13 %
	M4	0 %	2 %	1 %	11 %	1 %	2 %	5 %	15 %
T3	*M1*	7 %	1 %	3 %	6 %	4 %	2 %	6 %	4 %
	M2	29 %	6 %	3 %	8 %	28 %	3 %	1 %	7 %
	M3	0 %	0 %	3 %	8 %	0 %	1 %	13 %	4 %
	M4	0 %	1 %	0 %	3 %	1 %	1 %	2 %	8 %

of tile values in light of one of the four predefined modes. Specifically, Table 2 provides three thresholds: first, *T1* at the fourth color increment of a heat map visualization (i.e., values $\geq \frac{4}{19}$); second, *T2* at the ninth increment (i.e., values $\geq \frac{9}{19}$); and, finally, *T3* at the fifteenth increment (i.e., values $\geq \frac{15}{19}$). In the case of *Camera 1*, for example, zone *Z1* in mode *M1* indicates roughly 40 % of its tiles (136 out of 343 tiles) met *T1* with values of at least 151 frames each (50 frames per increment). This value correspondingly lowers to 15 % for *T2* and 7 % for *T3*.

Based on Figures 7a–7d and Table 2 the following conclusions are drawn. Evidently, the closer distance of *Camera 2* to the main walking path of employees left to both Kinect sensors is reflected in the results. Figure 7a, for instance, demonstrates this as to zone *Z3* on both cameras shows notable visual and numerical differences. Whereas roughly 27 % of tiles (threshold *T1*) in the case of *Camera 1* contain at least 151 frames, this number increases to circa 49 % for *Camera 2*. Furthermore, Figure 7a pinpoints areas where the most frames were collected (i.e., magenta clusters). While there are spots in the near vicinity of the Kinect sensors, foremost the main pathway as well as area *Lobby* and the card boards are highlighted. Areas with a lower number of frames are indicated, for instance, on the right in zone *Z3* of *Camera 1*. Figure 7b markedly contrasts Figure 7a. It is evident that people spent more time in front of the Ambient Surfaces and in the area around the card boards. Compared to all other zones, zone *Z1* shows in both cases the highest relative percentage values in mode *M2* (thresholds *T1–T3*) which is reflected by the accumulation of magenta-colored tiles in this area. Aside from the location in front of the card boards, zones *Z3–Z4* play a negligible role in the visualizations of mode *M2*. Figure 7c underlines that the most people were detected in the main walking path left of both Kinect sensors. This area was, apparently, heavily used to move through the building (e.g., arriving at work or having lunch in the canteen). There are no other areas where a comparably high number of people was detected, especially near to the Ambient Surfaces (i.e., zones *Z1* and *Z2*). Interestingly, the measurement of people's levels of engagement reflects this finding since the highest levels of engagement were detected in the same area, specifically in area *Lobby* (see Figure 7d). Apparently, people, although in a notable distance to both systems (roughly 3–4 meters), regularly looked at the Ambient

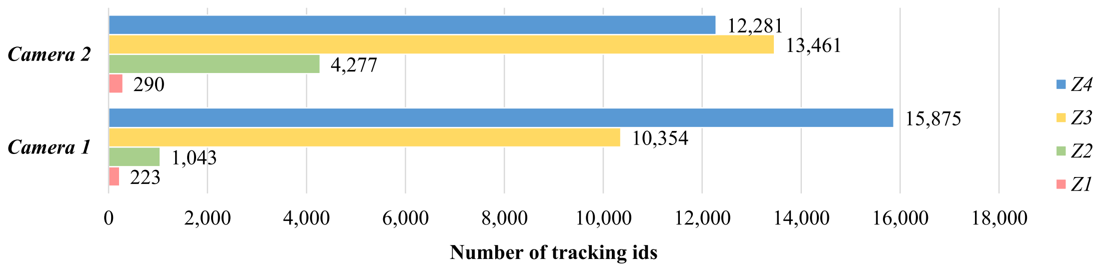


Figure 8.: The number of tracking ids detected for zones *Z1–Z4* with both *Camera 1* and *Camera 2*. Note: In the underlying calculations, a tracking id could be included in multiple classes (i.e., zones *Z1–Z4*).

Surfaces while passing by. In all other instances, the data is scattered across both field of views. Compared to *Camera 1*, the heat map visualizations of *Camera 2* include more tiles showing at least a minimum of data (see the darker teal color tones in Figure 7d) and, simultaneously, more tiles with higher values according to thresholds *T1–T3* (see mode *M4* in Table 2).

Finally, *Camera 1* detected more unique tracking ids (16,188) than *Camera 2* (14,383) in zones *Z1–Z4*. Only a little number of tracking ids (768) was recorded outside the four zones. Following Figure 8, *Camera 1* recorded the most tracking ids in Zone *Z4* (15,875), while *Camera 2* did the same in zone *Z3* (13,461). Zone *Z2*, in particular, shows markedly strong differences for both cameras. The amount of detected tracking ids by *Camera 2* is more than four times the number recorded by *Camera 1*. It is evident that the closer distance of *Camera 2* to the main walking path led to notably higher numbers of passersby detected in zones *Z2* and *Z3*.

5.2. *Spatial audience behavior: Entry and exit directions*

Figure 9 presents the entry and exit directions grouped by the number of tracking ids detected, while the different areas *L*, *LC*, *C*, *RC*, *R* can be visually interpreted in Figures 7. Considering all records, most of the records of *Camera 1* (8,022 records) originated in area *LC*, while the majority of records of *Camera 2* (5,297 records) began in area *C*. Principally, there is an observable shift regarding the numbers of record entry directions. Whereas the entry direction values for *Camera 1* are always the highest in areas *LC* and *C*, they are the highest in areas *C*, *RC*, and *R* for *Camera 2*. Both cameras, however, have in common that most records ended in area *L*—i.e., most people exited the scene left of both Ambient Surfaces. Except in one case (i.e., *Camera 2*, *6+ Persons*), the corresponding column has the highest values compared to all other exit direction areas. In the case of *Camera 2*, the area *RC* has always the highest values of the remaining exit directions. With respect to *Camera 1*, the second highest values for the exit directions can be observed in areas *LC* and *C*.

5.3. *Temporal facets of audience behavior*

Below, findings are presented regarding the Audience Funnel framework's phases of *Passing by*, *Viewing and Reacting*, and *Direct Interaction*.

	Entry					Exit				
Camera 1	*L*	*LC*	*C*	*RC*	*R*	*L*	*LC*	*C*	*RC*	*R*
All records	399	8,022	3,041	464	63	11,043	284	457	120	85
1 Person	271	7,218	1,798	105	33	8,942	86	264	69	64
2 Persons	59	497	808	174	15	1,319	106	98	21	9
3 Persons	28	158	234	98	7	423	48	42	9	3
4 Persons	16	81	95	62	3	205	23	23	6	0
5 Persons	12	28	45	26	1	86	9	12	4	1
6 Persons	5	14	21	8	2	32	6	7	3	2
6+ Persons	9	17	28	10	3	35	6	12	7	7

	Entry					Exit				
Camera 2	*L*	*LC*	*C*	*RC*	*R*	*L*	*LC*	*C*	*RC*	*R*
All records	407	830	5,297	4,019	582	9,401	140	534	711	349
1 Person	300	626	4,694	3,170	237	7,957	75	337	415	243
2 Persons	63	137	426	548	180	977	48	122	155	52
3 Persons	36	37	99	173	71	263	13	43	71	26
4 Persons	9	8	52	55	45	108	3	22	28	8
5 Persons	10	6	21	36	24	65	1	7	16	8
6 Persons	1	3	8	12	12	21	1	3	8	3
6+ Persons	2	1	7	12	14	11	1	3	12	9

Figure 9.: Entry and exit directions of records for both Kinect sensors were grouped by the number of people detected. The two highest values of each group and direction are highlighted.

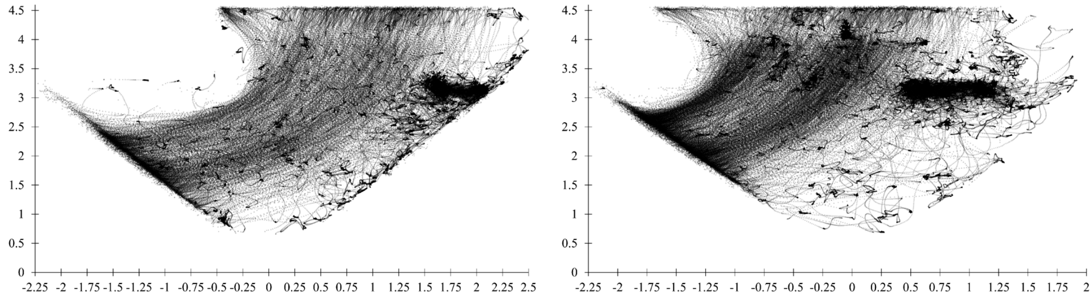
5.3.1. Phase: Passing by

Considering that all records were excluded where no level of engagement was measurable, still almost 17 % of all tracking ids showed no engagement at all. In total, 5,167 out of 30,571 tracking ids could be categorized as passersby. Figure 10a illustrates the corresponding pathways with only little activity directly in front of both Ambient Surfaces, while there are two places that indicate high numbers of passersby: first, the area top-left to both systems (i.e., the main walking path of employees); and, second, the area in front of the card boards. It is evident that the area around the card boards had been frequently used for activities such as meetings and discussions.

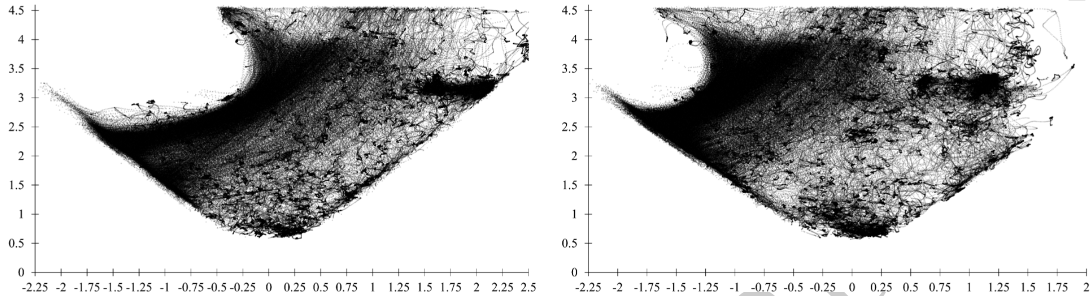
5.3.2. Phase: Viewing and Reacting

The 25,404 remaining tracking ids of both cameras (circa 83 %) show at least a minimum level of engagement. Most tracking ids (10,966) indicated a level of engagement of between 0 % and 10 %. Compared to the total number of passersby, more than twice as many people evidently paid at least a little attention to the Ambient Surfaces. For all other cases, Figure 11 indicates that the higher the measured level of engagement, the lower is the corresponding number of detected tracking ids. Interestingly, the according line chart curvatures illustrated suggest that this observation seems to follow behavior, to a lesser or greater extent, seen in exponential decay functions. In sum, the relative changes in the levels of engagement are slightly higher on *Camera 1* than on *Camera 2*. It is also apparent that both cameras detected notably different numbers of tracking ids for the level of engagement between 0 % and 10 %. We attribute this result to the fact that *Camera 1* was able to better capture the area in front of the cards boards where meetings were regularly held. Furthermore, independently of the level of engagement, *Camera 1* initially captured the most people in area *LC*. While for *Camera 1* no other correlation was found between the different levels of engagement and the way people were recorded to enter the scene, the data of *Camera 2* indicated a shift from the area *RC* to the area *LC*. Whereas the lowest levels of engagement could be linked to tracking ids originating in area *RC*, higher levels could be correlated to area *C*. However, the two highest levels of engagement could be related to people that were initially observed in area *LC*.

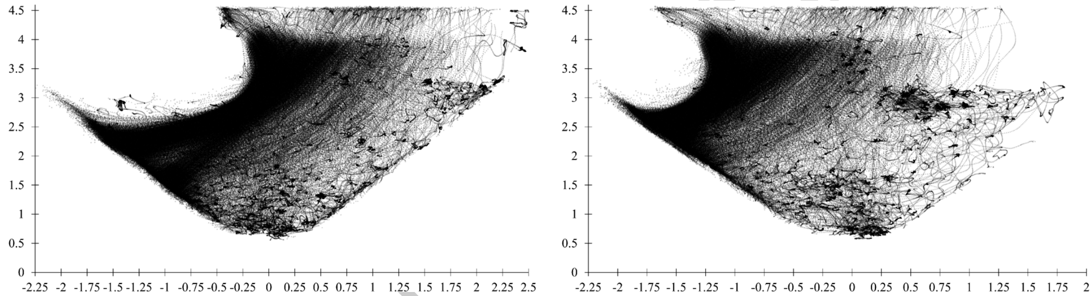
Finally, Figure 10b and 10c visualize the pathways of the 25,404 tracking ids. While Figure 10b shows the data with a 1–9 % level of engagement, Figure 10c depicts the sum of the remaining levels of engagement (i.e., 10–99 %). In contrast to Figure 10a,



(a) Level of engagement: 0 %.

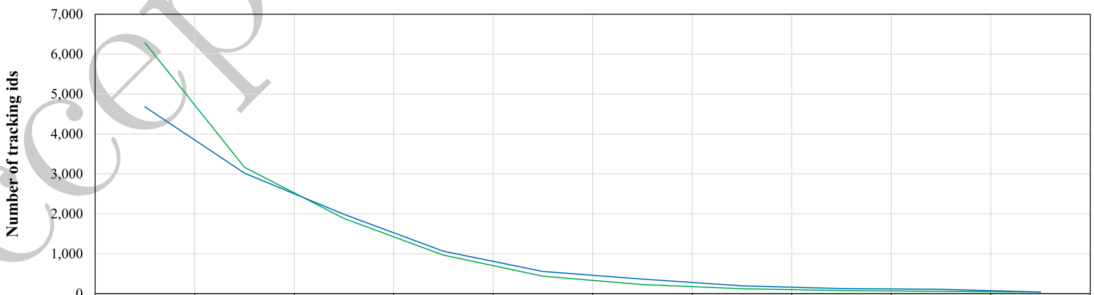


(b) Level of engagement: 1–9 %.



(c) Combined levels of engagement: 10–99 %.

Figure 10.: Scatter plots of *Camera 1* (right) and *Camera 2* (left) showing the walking paths of people according to their levels of engagement (scaling in meters; black = data points; white = no data).



	0 %<e<10 %	10 %≤e<20 %	20 %≤e<30 %	30 %≤e<40 %	40 %≤e<50 %	50 %≤e<60 %	60 %≤e<70 %	70 %≤e<80 %	80 %≤e<90 %	90 %≤e<100 %
Camera 1	6,286	3,165	1,883	967	435	229	123	76	57	39
Camera 2	4,680	3,020	1,989	1,062	555	366	196	126	109	41

Figure 11.: The number of detected tracking ids in relation to the ten different levels of engagement (e).

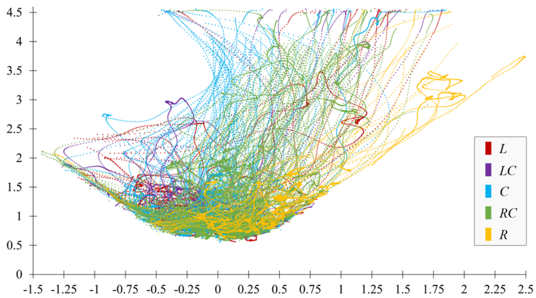


Figure 12.: Scatter plots of *Camera 1* (right) and *Camera 2* (left) illustrating the walking paths of people that engaged in direct interaction (scaling in meters). The different paths are highlighted according to their area of origin (i.e., *L*–*R*).

these two illustrations indicate notable amounts of gatherings in front of both Ambient Surfaces. While the main walking path and the area in front of the card boards still show high numbers of detected people, there are simultaneously multiple spots where the data accumulated. In Figure 10b these spots are scattered across the entire field of view of both cameras, whereas in Figure 10c these areas are more in closer distance to the Ambient Surfaces. Figure 10c, furthermore, unveils that the area in front of the card boards is less frequently used considering the increased level of engagement. In all visualizations (i.e., Figures 10a–10c), however, the main walking path left to both Ambient Surfaces shows frequent usage. Consequently, any extent of viewing and reacting behavior can be linked to this particular area, including area *Lobby*.

5.3.3. Phase: Direct Interaction

In total, 223 unique tracking ids were identified in zone *Z1* of *Camera 1*, whereas 290 tracking ids were correspondingly detected with *Camera 2* in the same area. This accumulates to roughly 0.01 % (*Camera 1*) and to circa 0.02 % (*Camera 2*) of all people detected in zones *Z1*–*Z4*. Figure 12 illustrates the walking paths of the 513 tracking ids, which, at times, look like somewhat straight lines, while there are also instances where a person seemed to move back and forth or up and down. In the near vicinity of both systems, the walking paths accumulate and are hardly separable from one another. Interestingly, the main walking path left to both Ambient Surfaces is hardly recognizable in these illustrations—particularly in the case of *Camera 1*. In addition, Figure 12 depicts from which direction people entered the scene. While *Camera 1* initially detected most people in area *C* (*L*: 26, *LC*: 58, *C*: 82, *RC*: 43, *R*: 14), *Camera 2* did the same in area *RC* (*L*: 37, *LC*: 29, *C*: 78, *RC*: 104, *R*: 42).

6. Discussion

A careful study of the literature led us to discuss the originality of our work in light of four individual contributions. First, it is returned to the leveraged approach to separate passersby from real users and how it extends existing knowledge. Second, attention is drawn to findings expanding on existing spatial and temporal models suggesting the necessity to incorporate incidents of unprovoked, direct interaction in future models. Third, it is concentrated on the novel characteristics of our skeletal data set. Finally, it is focused on how our research adds to the exploration of skeletal data. Analogously to these contributions, we describe research implications, underline their practical and scientific usefulness, and elaborate recommendations for future work. The discussion is concluded with a presentation of research limitations.

6.1. Separating passersby from real users

Our work strongly builds on the .NET framework's *Engaged* property to allow for deducing different levels of engagement. In a nutshell, the property assists in determining whether a person is looking at a display or not. It helps to isolate and gather data about, for example, people that choose not to interact (Williamson & Williamson, 2014). To the best of our knowledge, related research, so far, has not been focusing on the task of separating passersby from real users. While already a decade ago Michelis and Müller (2011) discussed this issue, also more recent research notes limitations in this regard (Mäkelä et al., 2018). The implication is that audience behavior can prospectively be analyzed in greater detail compared to how it was done in the past. Future research may start to more systematically identify areas in front of display installations in which people show lower or higher levels of engagement. We think that our results can add manifold quantitative nuances to, for example, passive engagement zones seen in observation-based models such as these from Memarovic et al. (2012) or Boucher et al. (2021).

As there is a demand for improved evaluations methods in long-term ambient display deployments (Mäkelä et al., 2018), the opportunity to classify users more granular may contribute a useful tool for scientists on this avenue. Said separation may also assist with the challenge of modeling people's behavior with public and semi-public displays (Ardito et al., 2015). Future work may collect and compare levels of engagements from public and semi-public spaces to unveil similarities and differences between these settings. Additionally, future models may incorporate the different levels of engagement and linking them to, potentially even statistically significant, outcomes. We do think that the presented approach can also serve as a useful tool for practitioners such as user interface designers and staff members in companies. Based on the varying levels of engagement, interface designers may investigate whether changes in the user interface entail different behavior, while employees may test different display locations to increase the visibility of content. As Elhart et al. (2017) note, comprehensive reports of audience behavior can assist display owners to optimize their installations.

6.2. Direct interaction without apparent incentives

Existing spatial and temporal models assume passive, unrelated behavior at the early phases of the interaction process, followed by phases describing increased attention leading to direct interaction (e.g., Boucher et al., 2021; Germany et al., 2019). Considering, however, results from this research, we see reason to hypothesize that this chain of subsequent phases is, at times, bypassed entirely. In this context, attention is drawn to Figure 10a and 12. While Figure 10a demonstrates the curvature of the main walking path, Figure 12 allows to inspect the pathways of people engaging in direct interaction. What makes Figure 12 unique is that it highlights pathways that are, to a lesser or greater extent, directly orientated toward the Ambient Surfaces (e.g., the purple walking paths for *Camera 1*). Without any notable reason or incentive to do so, people, in such instances, seemingly were coming from area *Lobby* to engage in direct interaction. Ultimately, we attribute this behavior to the fact that, as we previously learned (Schwarzer et al., 2021), at least some of the employees appropriated the Ambient Surfaces as to they regularly checked for new information on the screens. In said cases, people might have fetched a beverage from area *Lobby* and, on their way back to the office, intended to inspect, for instance, a project's build status. Apparently, there was no need to entice people in these instances to interact, what is usually a major challenge in related, public display settings (Williamson & Williamson, 2014).

Overall, we do think that this discovery is, foremost, useful from a scientific point of view. While we acknowledge that most existing models build on rather public environments (i.e., usually more first-time users and less repeating users), we nonetheless do see implications for future research. What other scientists can draw from this finding is that future audience behavior models should consider some degree of appropriation as, we believe, existing models lack to incorporate such behavior. Considering the increasing number of both public and semi-public display installations (Ardito et al., 2015), we do see the relevance for such research in the future. Simultaneously, we do conclude that these examples contribute some nuances toward the goal of developing a more comprehensive understanding of the attention people give to ambient displays (Parker et al., 2018). Future research may focus on said instances of direct interaction to, for instance, systematically train machine learning algorithms that are subsequently applied to data of similar contexts. The goal would be to automatically gather more cases to underpin and describe their occurrence quantitatively.

6.3. The size and nature of the skeletal data set

Following Mäkelä et al. (2018), data collection approaches based on skeletal data are ecologically valid methods for evaluations. The data gathered allows rich insights regarding interaction and non-interaction that would be otherwise hard to obtain (Williamson & Williamson, 2014), whereas real-world experiences are integral to authentic experience (Williamson & Williamson, 2017). We did collect people's experience in the wild and gathered skeletal data over the course of several months. Hence, we do believe that the skeletal data underlying the results shows high ecological validity. Additionally, and similar to Williamson and Williamson (2014), we do also think that the data collected is without bias (i.e., incorporating information from both interacting and non-interacting users alike) and has high spatial and temporal accuracy. We build this argument on the fact that the Ambient Surfaces were operational for more than 3 years prior to the deployment of both Kinect sensors. Furthermore, we do not think that the presence of the cameras led to any notable unnatural behavior as, for instance, discussed by Shi and Alt (2016). We do so primarily because employees and the management were part in the Kinect sensors' deployment process (e.g., discussions of ethical concerns) and the cameras went operational several weeks after their initial installation. We did also not leverage attention-grabbing features to artificially attract users that could have likely raised concerns regarding the novelty effect.

The data set analyzed incorporating information of more than 30,000 passersby, is, in comparison, one of just a few sets that large. As far as we know, the data sets from Elhart et al. (2017) and Mäkelä et al. (2018) are one exception to this rule (circa 40,000 and 100,000 passersby respectively). We do, furthermore, believe that the present study is the first one building on skeletal data stemming from a professional context. Our data set arguably allows rich insights on how ambient displays are appropriated in everyday work life and our findings can, therefore, build the foundation for revelatory research to come. Patterns in the data set are a product of authentic, social behavior and are not skewed by, for instance, a high number of first-time users who could have very likely affected the data due to effects related to the issue of curiosity (e.g., the novelty effect). As we concluded recently (Schwarzer et al., 2021), more research is necessary with respect to longitudinal deployments of ambient displays in authentic environments such as in the ASD context. We do think that the data set collected is a promising stepping stone on this avenue. The quantitative results drawn from the data set extend previous research by, for example, substantiating existing spatial models with respect to their largely observation-based distance measurements of the various interaction zones (see Figure 8). Researchers and practitioners may derive from that suggestions on how to redesign workspaces or whether to relocate display installations.

6.4. Exploration of skeletal data

A recent study from Mangal and Tiwari (2020) concludes that skeleton tracking with Kinect v2 sensors needs additional exploration. Similarly, Mäkelä et al. (2018) conclude that skeletal data has not been fully utilized. Among others, a lot of manual work is still required to go through such data. At the same time, depth data has often been used in combination with other methods as supporting evidence (e.g., observations and interviews). With our study, we hope, however, demonstrating useful insights on how depth data, particularly Kinect v2 skeletal data, can serve as the primary data source. By respectively illustrating our data collection, preprocessing, and analysis procedures, our work adds to the development of fully automated tools to capture audience behavior in real-world deployments. Simultaneously, with the exploration of our skeletal data set, we hope to contribute to a better understanding of how ambient displays are really used in the wild (Parker et al., 2018). Considering that patterns in data will be the foundation for future innovations, services, and interactions (Brown, Bødker, & Höök, 2017), our research can arguably distill insightful nuances on this path. Like Elhart et al. (2017), we do believe that our analysis approach can be both useful for other researchers to evaluate their own display solutions and beneficial in terms of strengthening the quantitative foundations for our field.

Results drawn from the data set are, to a lesser or greater extent, similar to and different from those captured in related studies. We now discuss these aspects in more detail. As shown in Figure 7d, the highest levels of engagement were found circa 3–4 meters away from both Ambient Surfaces (i.e., in area *Lobby*). We therefore, on the one hand, concur with Germany et al. (2019) that attention leading to awareness and, finally, screen interaction may not necessarily depend so much on the question of distance but more on the question of a message's clarity conveyed by a display. On the other hand, however, the mere presence of the Ambient Surfaces might have as well created incidents of interaction (Wouters et al., 2016) which, on their own, could be one of the main triggers to start an interaction in the first place (Ardito et al., 2015). Evidently, we (Schwarzer et al., 2021) and others (Michelis & Müller, 2011) observed people typically looking briefly at display installations while passing by. Our data suggests this effect occurring most markedly in area *Lobby* (see Figure 3)—after people, for instance, fetched a beverage. Placing the Ambient Surfaces in the near vicinity of the main pathway of employees, including area *Lobby*, apparently created, time and time again, situations to interact actively and passively. We therefore see parallels to related research in public settings. Here, a closer proximity of a display to the main walking path of pedestrians correlates to higher user engagement—and vice versa (Parker et al., 2018). Our results underline this relationship as to the levels of engagement are indeed slightly higher in the case of *Camera 2* (see Table 2). Similarly, *Camera 2* indicates higher levels of engagement, for example, in areas *C* and *RC* (see Figure 7d). Yet, the larger distance of *Camera 1* to area *Lobby* led to no measurable and observable shifts in the levels of engagement to the right. It seems that staff members, at many occasions, looked at both Ambient Surfaces somewhat equally when a situation happened to occur. We know from past research (Schwarzer et al., 2021) that interaction is largely linked to spontaneous incidents which, in our view, underlines the relevance of a closer distance between the main walking path and a display's location. Had both systems been placed distant to any major pathway, instances of interaction would have been created less often.

Attention is now drawn to the Audience Funnel framework's user classification. In comparison to the studies from Michelis and Müller (2011) and Mäkelä et al. (2018), the following can be observed. First, as we removed all records including no engagement at all from analysis (54,964 records in total), we drastically lowered the number of passersby. While in both mentioned studies, most people where categorized as passersby, we only identified roughly 17 % of people spending no attention toward the Ambient Surfaces. All other tracked staff members showed at least a minimum level of viewing and reacting behavior (i.e., circa 83 %). Second, we identified both smaller and similar numbers of people engaging in direct interaction (*Camera 1*: 1.4 % of people; *Camera 2*: 2.02 % of people). While Michelis and Müller (2011) observed almost a third of people directly interacting, Mäkelä et al. (2018) identified, interestingly, a comparable small number of direct interaction (1.4 %). For the large differences between our numbers and those from Michelis and Müller (2011), we see two main reasons. On the one hand, there was only a very little number of first-time users in out setting, if at all. There was, in fact, a high number of returning users who were familiar with the contents displayed. On the other hand, the user interface of the Ambient Surfaces did not require touch interactions to unveil information in every case. Hence, people could often interpret information without the need to enter zone *Z1*. Third, similar to Mäkelä et al. (2018), Table 1 suggests that most times a person was detected, they were tracked as lone passersby (circa 80 % of records). However, we saw a higher relative amount of people entering zone *Z3*. While Mäkelä et al. (2018) tracked roughly 22 % of passersby to enter the space less than 2.80 meters away, our results respectively indicate 64 % (*Camera 1*) and 95 % (*Camera 2*) of employees were tracked at least once in Zone *Z3*. We attribute these differences primarily to the very close proximity of both Ambient Surfaces to the main walking path. Finally, in contrast to findings from Mäkelä et al. (2018), people in our study most notably approached the Ambient Surfaces from the front, meaning areas *LC*, *C*, and *RC* (see Figure 9). Yet, we concur with Mäkelä et al. (2018) as to direct interaction is more likely to occur when people were entering the scene from the top.

6.5. Research limitations

Our work is not without limitations. Similar to Elhart et al. (2017) and Mäkelä et al. (2018), we note that the Kinect sensor's restricted horizontal and vertical field of view allowed only to capture a limited corridor, not the entire scene, in front of both Ambient Surfaces. Thus, we were able to solely investigate this narrower corridor. The field of view was also restricted by the earlier mentioned concrete wall that led to the prominent curvature of the main pathway of people. Hence, the investigated audience behavior underlies limitations regarding spatial conditions. Furthermore, situations such as passersby walking behind other people or people standing in front of others may have resulted in wrongly grouping records according to their total number of detected people. For such incidents, Mäkelä et al. (2018) discuss the possibility of relocating cameras. In our case, a setup consisting of multiple Kinect sensors placed at different locations (e.g., in each corner) could have mitigated this issue by cross-comparing the sensor data. It is, additionally, not possible to determine returning users based on skeletal data alone (Mäkelä et al., 2018), while there is also the chance that one to many staff members particularly affected the results. As we, again, learned in previous research (Schwarzer et al., 2021), people reported to have, to a lesser or greater extent, appropriated the Ambient Surfaces in their daily routines. Also, the fact that the Kinect sensor better recognizes people coming from the front, may have led to a bias in the calculations of entry and exit directions.

We also found that the *Engaged* property has its limitations. For example, it is possible to trick the underlying algorithms by facing the sensor directly, while, simultaneously, gazing in another direction. In such cases, the algorithm wrongly assumes that a person is directly looking at the sensor (i.e., the *Engaged* property value equates to *Yes*). We do not believe that any employee was aware of this technical issue. For instance, there was no visual feedback implemented in the user interface that would have indicated changes in the property's value. Nonetheless, we found this information worth noting because we experienced this effect throughout our experimentation. The *Engaged* property sometimes also wrongly estimates that a person is directly facing the sensor at all times while walking past. In instances, the sum of a tracking id's *Yes* values added up to a level of engagement of 100 %, while a person was clearly walking away from the sensor. We double-checked such examples by replaying them in the Kinect studio software. As a result of this observation, we ultimately removed all tracking ids showing such behavior prior to analysis. The rather small distance between both Kinect sensors (roughly 0.85 meters apart), we do, however, not consider an issue because we experienced the *Engaged* property to be very sensitive in detecting people's principle viewing directions.

It is also worthwhile mentioning that our analysis entirely built on the *Spine Shoulder* joint. While we observed this joint not being as prone to tracking issues such the arm joints (e.g., during body rotation), other joints may have unveiled richer findings. Like Mäkelä et al. (2018) we did not track joints from the lower body because, again, we were interested in a person's general location in front of both Ambient Surfaces. Finally, we like to draw attention to our rather strict data selection procedure. Based on six specific rules, we limited the available data for the present study. We admit that a less rigorous selection process would have potentially revealed additional results, while we, at the same time, still see difficulties in assuring high quality in a chosen subset of the data (e.g., defining thresholds for erroneous data). Thus, the conclusions drawn in this article could have pointed one way or the other depending on the underlying data set. For example, the total number of passersby would have been notably higher when no data would have been removed.

7. Conclusion

In this research, we investigated the spatial and temporal audience behavior around a custom, semi-public ambient display installation. Specifically, we observed said behavior in a professional, large-scale ASD environment where we deployed two of our Ambient Surfaces solutions for roughly 5 years. Throughout 18 weeks in 2017, behavioral information of more than 30,000 passersby was collected using our AmbiLogger tool and two Microsoft Kinect v2 cameras. Ultimately, the present research adds to existing knowledge in four specific ways. First, the study envisions novel means to separate passersby from real users, including their levels of engagement. Second, findings suggest exiting models of audience behavior lacking to incorporate instances of unprovoked, direct interaction. Third, building on a large, revelatory skeletal data set distilled from a professional context, this research contributes manifold nuances to a field predominantly informed by studies in public environments. Finally, by elaborating the procedures to collect, preprocess, and analyze skeletal data, our study adds to the exploration of such data.

Overall, we see great potential for future research on audience behavior around ambient displays—with both a positivist and a more pragmatic lens. We concur with Williamson and Williamson (2014) as to we similarly find our work to only scratch the surface when it comes to fruitful directions for research in this domain. To advance on the present study, we are currently working on correlating skeletal data with touch interaction data to learn more about the context of direct interaction. We are also elaborating on the question of whether there is a link between people's pathways and the visualization of specific contents. Furthermore, while we were largely focusing on absolute positions of people around the Ambient Surfaces so far, we are now shifting our emphasis to vector-based analyses to compare curvatures of pathways in light of, for instance, similarities in the movement.

Acknowledgments

We would like to thank the company and the ASD department for their long-standing participation in this research.

Funding

This research is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—project number 451069094.

References

Alt, F., Buschek, D., Heuss, D., & Müller, J. (2021, March). Orbuculum - predicting when users intend to leave large public displays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, *5*(1), 1–16. Retrieved from `https://doi.org/10.1145/3448075`

Ardito, C., Buono, P., Costabile, M. F., & Desolda, G. (2015). Interaction with large displays: A survey. *ACM Computing Surveys*, *47*(3), 1–38. Retrieved from `https://doi.org/10.1145/2682623`

Boucher, M., Blumenstein, K., de Jesus Oliveira, V. A., & Seidl, M. (2021). Byod - bringing your own device into single-surface interaction models. In R. Wimmer (Ed.), *Proceedings of the fifteenth international conference on tangible, embedded, and embodied interaction* (pp. 97:1–97:13). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3430524.3446073`

Brignull, H., & Rogers, Y. (2003). Enticing people to interact with large public displays in public spaces. In *Human-computer interaction interact '03: Ifip tc13 international conference on human-computer interaction* (pp. 17–24). Amsterdam, Netherlands: IOS Press.

Brown, B., Bødker, S., & Höök, K. (2017). Does hci scale? scale hacking and the relevance of hci. *Interactions*, *24*(5), 28–33. Retrieved from https://doi.org/10.1145/3125387

Dalton, N., Dalton, R., & Hoelscher, C. (2015). People watcher: An app to record and analyzing spatial behavior of ubiquitous interaction technologies. In *Proceedings of the 4th international symposium on pervasive displays* (pp. 1–6). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2757710.2757714

Davies, N., Clinch, S., & Alt, F. (2014). *Pervasive displays: understanding the future of digital signage.* Morgan and Claypool Publishers.

Dingsøyr, T., Fægri, T. E., & Itkonen, J. (2014). What is large in large-scale? a taxonomy of scale for agile software development. In A. Jedlitschka, P. Kuvaja, M. Kuhrmann, T. Männistö, J. Münch, & M. Raatikainen (Eds.), *Product-focused software process improvement* (pp. 273–276). Cham, Switzerland: Springer. Retrieved from https://doi.org/10.1007/978-3-319-13835-0_20

Elhart, I., Mikusz, M., Mora, C. G., Langheinrich, M., & Davies, N. (2017). Audience monitor: An open source tool for tracking audience mobility in front of pervasive displays. In *Proceedings of the 6th acm international symposium on pervasive displays* (pp. 1–8). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3078810.3078823

Germany, J. O., Speranza, P., & Anthony, D. (2019). Eliciting public display awareness and engagement: An experimental study and semantic strategy. *International Journal of Human–Computer Interaction*, *35*(20), 1975–1985. Retrieved from https://doi.org/10.1080/10447318.2019.1597572

Greenberg, S., Marquardt, N., Ballendat, T., Diaz-Marino, R., & Wang, M. (2011, January). Proxemic interactions: The new ubicomp? *Interactions*, *18*(1), 42–50. Retrieved from https://doi.org/10.1145/1897239.1897250

Hall, E. T. (1966). *The hidden dimension.* Doubleday: Garden City, NY, USA.

Huang, E. M., Mynatt, E. D., Russell, D. M., & Sue, A. E. (2006). Secrets to success and fatal flaws: The design of large-display groupware. *IEEE Computer Graphics and Applications*, *26*(1), 37–45. Retrieved from https://doi.org/10.1109/MCG.2006.21

Koch, M., von Luck, K., Schwarzer, J., & Draheim, S. (2018). The novelty effect in large display deployments – experiences and lessons-learned for evaluating prototypes. In *Proceedings of 16th european conference on computer-supported cooperative work - exploratory papers* (Vol. 2). European Society for Socially Embedded Technologies (EUSSET). Retrieved from https://doi.org/10.18420/ecscw2018_3

Mäkelä, V., Heimonen, T., & Turunen, M. (2018). Semi-automated, large-scale evaluation of public displays. *International Journal of Human–Computer Interaction*, *34*(6), 491–505. Retrieved from https://doi.org/10.1080/10447318.2017.1367905

Mangal, N. K., & Tiwari, A. K. (2020). Kinect v2 tracked body joint smoothing for kinematic analysis in musculoskeletal disorders. In *2020 42nd annual international conference of the ieee engineering in medicine biology society (embc)* (pp. 5769–5772). IEEE. Retrieved from https://doi.org/10.1109/EMBC44109.2020.9175492

Memarovic, N., Langheinrich, M., Alt, F., Elhart, I., Hosio, S., & Rubegni, E. (2012). Using public displays to stimulate passive engagement, active engagement, and discovery in public spaces. In *Proceedings of the 4th media architecture biennale conference: Participation* (pp. 55–64). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2421076.2421086

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

Niu, J., Wang, X., Wang, D., & Ran, L. (2020). A novel method of human joint prediction in an occlusion scene by using low-cost motion capture technique. *Sensors*, *20*(4), 1119:1–1119:12. Retrieved from https://doi.org/10.3390/s20041119

Parker, C., Tomitsch, M., Davies, N., Valkanova, N., & Kay, J. (2020). Foundations for designing public interactive displays that provide value to users. In *Proceedings of the 2020 chi conference on human factors in computing systems* (pp. 1–12). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3313831.3376532`

Parker, C., Tomitsch, M., & Kay, J. (2018). Does the public still look at public displays? a field observation of public displays in the wild. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, *2*(2), 73:1–73:24. Retrieved from `https://doi.org/10.1145/3214276`

Prante, T., Röcker, C., Streitz, N., Stenzel, R., Magerkurth, C., van Alphen, D., & Plewe, D. A. (2003). Hello.wall – beyond ambient displays. In *Video track and adjunct proceedings of the fifth international conference on ubiquitous computing (ubicomp'03)* (pp. 277–278).

Schwarzer, J., Draheim, S., von Luck, K., Wang, Q., Casaseca, P., & Grecos, C. (2016). Ambient surfaces: Interactive displays in the informative workspace of co-located scrum teams. In *Proceedings of the 9th nordic conference on human-computer interaction.* New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/2971485.2971493`

Schwarzer, J., Draheim, S., von Luck, K., Wang, Q., & Grecos, C. (2021). Spontaneous utilization: A classic grounded theory of utilizing ambient displays in professional, large-scale agile software development environments. *International Journal of Human–Computer Interaction.* Retrieved from `https://doi.org/10.1080/10447318.2021.2002051`

Shi, J., & Alt, F. (2016). The anonymous audience analyzer: Visualizing audience behavior in public space. In *Proceedings of the 2016 chi conference extended abstracts on human factors in computing systems* (pp. 3766–3769). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/2851581.2890256`

Stephanidis, C., Salvendy, G., of the Group Margherita Antona, M., Chen, J. Y. C., Dong, J., Duffy, V. G., … Zhou, J. (2019). Seven hci grand challenges. *International Journal of Human–Computer Interaction*, *35*(14), 1229–1269. Retrieved from `https://doi.org/10.1080/10447318.2019.1619259`

Streitz, N., Röcker, C., Prante, T., Stenzel, R., & van Alphen, D. (2003). Situated interaction with ambient information: Facilitating awareness and communication in ubiquitous work environments. In *Proceedings of hci international 2003, 10th international conference on human–computer interaction: [held jointly with] symposium on human interface (japan) 2003, 5th international conference on engineering psychology and cognitive ergonomics, 2nd international conference on universal access in human–computer interaction* (pp. 133–137).

Vogel, D., & Balakrishnan, R. (2004). Interactive public ambient displays: Transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of the 17th annual acm symposium on user interface software and technology* (pp. 137–146). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/1029632.1029656`

Wasenmüller, O., & Stricker, D. (2017). Comparison of kinect v1 and v2 depth images in terms of accuracy and precision. In C.-S. Chen, J. Lu, & K.-K. Ma (Eds.), *Computer vision – accv 2016 workshops* (pp. 34–45). Cham: Springer International Publishing. Retrieved from `https://doi.org/10.1007/978-3-319-54427-4_3`

Williamson, J. R., & Williamson, J. (2014). Analysing pedestrian traffic around public displays. In S. Gehring (Ed.), *Proceedings of the international symposium on pervasive displays* (pp. 13–18). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/2611009.2611022`

Williamson, J. R., & Williamson, J. (2017). Understanding public evaluation: Quantifying experimenter intervention. In *Proceedings of the 2017 chi conference on human factors in computing systems* (pp. 3414–3425). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3025453.3025598`

Wouters, N., Downs, J., Harrop, M., Cox, T., Oliveira, E., Webber, S., … Vande Moere, A. (2016). Uncovering the honeypot effect: How audiences engage with public interactive systems. In *Proceedings of the 2016 acm conference on designing interactive systems* (pp. 5–16). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://`

doi.org/10.1145/2901790.2901796
Ye, E., Ye, X., & Liu, C. (2018). Teamwatch: Visualizing development activities using a 3-d city metaphor to improve conflict detection and team awareness. *PLOS ONE*, *13*(3), 1–27. Retrieved from https://doi.org/10.1371/journal.pone.0193562

Biographies

Jan Schwarzer is a postdoc researcher at CSTI. His primary research interests include ambient displays, machine learning, methodology development, field deployment research, and agile software development. He received his PhD in computer science from UWS. Contact him at jan.schwarzer@haw-hamburg.de.

Susanne Draheim is managing director of the research and transfer institute Smart Systems and the affiliated research laboratories CSTI and Living Place Hamburg. She has a background in Sociology, Educational Sciences and Cultural Sciences and holds a PhD in Sociology from TU Dresden. Contact her at susanne.draheim@haw-hamburg.de.

Kai von Luck is a full professor at HAW Hamburg for computer science and artificial intelligence. He acts as academic director of the research and transfer institute Smart Systems and the affiliated research laboratories CSTI and Living Place Hamburg. Contact him at kai.vonluck@haw-hamburg.de.