Schwarzer, J.; von Luck, K.; Draheim, S.; Koch, M. (2019): Towards Methodological Guidance for Longitudinal Ambient Display In Situ Research. In: Proceedings of the 17th European Conference on Computer-Supported Cooperative Work: The International Venue on Practice-centred Computing an the Design of Cooperation Technologies - Exploratory Papers, Reports of the European Society for Socially Embedded Technologies (ISSN 2510-2591), DOI: 10.18420/ecscw2019_ep07

Towards Methodological Guidance for Longitudinal Ambient Display In Situ Research

Jan Schwarzer[†], Kai von Luck[†], Susanne Draheim[†], Michael Koch[‡]
[†]Hamburg University of Applied Sciences
{jan.schwarzer, kai.vonluck, susanne.draheim}@haw-hamburg.de
[‡]Universität der Bundeswehr München
michael.koch@unibw.de

Abstract. Field deployment research represents a promising way for understanding how technology is utilised in the wild. It gained relevance in both HCI and CSCW, and allows, for instance, to investigate how technology is socially embedded in real world contexts. However, such enterprises are considered complex in nature due to continuously changing conditions such as practices surrounding technology. In situ research has yet to gain momentum, leaving researchers with little theoretical guidance. In response, the present paper proposes the application of classic grounded theory in longitudinal field deployment studies for ambient displays. We argue that the methodology is a valuable choice in coping with the challenges surrounding in situ evaluations and simultaneously ensures methodological rigour. This paper contributes a practical systematisation of the methodology's two core concepts, namely constant comparison and theoretical sampling. It sheds light on their exemplary application in investigating quantitative interaction data in the early stages of our ongoing research. With that, we hope to encourage future research and provide a first stepping stone towards developing methodological guidance for evaluations of ambient displays in the wild.

Introduction

Socially embedded technology arguably challenged and changed practices like no other technological artefact had done before (Wulf et al., 2015). Research in this context embarks on affecting the fundamental ways of how research is conceptualised, as people and practices are more than just their relationships with technology—the concept of the 'user' becomes problematic (Bjørn and Boulus-Rødje, 2015). Here, practices and technology are being understood as intertwined and as continuously changing entities (Bjørn and Boulus-Rødje, 2015). Despite its challenging nature, research in the wild sheds light on real usage and highlights how technology interacts with environmental conditions such as technology already in use (Siek et al., 2014). Contrary to other research approaches, field deployment research is capable of investigating longitudinal effects and enabling researchers to cope with issues such as the novelty effect (Alt et al., 2012). Generally, long-term deployment studies are considered promising as they allow one to scrutinise how a technology is adapted in a particular context (Alt et al., 2012; Preim et al., 2018; Siek et al., 2014). In fact, operational feasibility of novel technology can only be determined in the field (Nunamaker Jr. et al., 2015). Although, field deployment research is considered a messy enterprise (Alt et al., 2012; Siek et al., 2014), it simultaneously affects most notably both science and society (Nunamaker Jr. et al., 2015). In situ research has gained momentum throughout HCI and particular relevance in the CSCW and Ubicomp disciplines (Siek et al., 2014) as their data can be assumed of high value (Alt et al., 2012). Recent contributions from a variety of domains such as information systems (Nunamaker Jr. et al., 2015), HCI (Börner et al., 2013; Hazlewood et al., 2011; Matthews et al., 2007; Messeter and Molenaar, 2012; Siek et al., 2014), information visualisation (Preim et al., 2018), and CSCW (Bjørn and Boulus-Rødje, 2015) stress the relevance of in situ research. Some authors place a particular emphasis on calls for long-term in the wild evaluations (Börner et al., 2013; Hazlewood et al., 2011; Preim et al., 2018).

In ambient display research, a strong technology-driven focus is observable, failing to consider how actual operation relates to people's everyday lives (Matthews et al., 2007). Social aspects, per se, have received little attention and broadening the scope of investigation is recommended (Messeter and Molenaar, 2012). Evidently, real-world evaluations indicate a lack of methodological guidance as traditional evaluation approaches do not apply. For instance, Hazlewood et al. (2011) conclude that ambient displays, in all their forms, require methodological development for in the wild evaluations.

In response, this paper introduces a holistic methodological approach applied in our ongoing longitudinal in situ evaluation of our custom 'Ambient Surfaces' solution. The paper sheds light on the utilisation of classic grounded theory (GT) and the practical systematisation of its two core processes, namely constant comparison and theoretical sampling (Boeije, 2002). To this end, their procedure is exemplary illustrated by discussing quantitative analyses of interaction data during

the early stages of our research. By following classic GT methodology, we argue that we can both adhere to methodological rigour and maintain flexibility in light of the aforementioned challenges.

The paper is organised as follows: firstly, an overview of the current state in evaluating socially embedded technology in the wild is presented, primarily focusing on the domain of ambient displays. Secondly, the methodological foundations of our research are highlighted, including the research setting and purpose, a consideration of GT's fit to conduct longitudinal in situ research, and an introduction to GT methodology. Thirdly and foremost, the paper presents the practical systematisation of the two core processes. Finally, before concluding the paper with some recommendations for future research, it discusses our approach, presents contributions to existing knowledge, and illustrates some limitations.

Evaluation of Ambient Displays in the Wild

Discussions in the HCI literature on how to conduct evaluations of user interfaces range back to early work such as Bannon and O'Malley (1984). However, these discussions often either focus on laboratory settings or short-term in situ deployments. Siek et al. (2014) provide some guidance on how to realise field deployments in HCI, including how to design data collection instruments. However, they stay rather superficial on this issue.

Some information on the utilised methods for evaluating long-term deployments of ambient displays can be found in reports about single research projects such as Peltonen et al. (2008), Rogers et al. (2010), Alt et al. (2012), Ojala et al. (2012), Memarovic et al. (2016), and Shelton and Nesbitt (2017). However, while addressing the topic of gathering and analysing data, the authors usually do not provide any background motivation as to why they conducted their evaluations in the way it was demonstrated.

Börner et al. (2013) report that reviewed studies used a variety of methodologies. However, it seems that the literature review rather lists a set of different data collection methods (e.g. interviews and observations) and design approaches (e.g. user-centred design and exploratory design). Following Crotty (1998), we found that surveys and heuristic inquiries were the only identified methodological choices.

Input regarding potential research goals can be found in the work from Nunamaker Jr. et al. (2015). The authors address the broader field of information systems and structure research into the three phases: 'proof-of-concept' research, 'proof-of-value' research, and 'proof-of-use' research. For each phase, the authors envision both field studies and laboratory studies as valid approaches and list potential evaluation goals. They argue that "The research is not complete ... until proof-of-use research demonstrates that a self-sustaining and growing community of practice has emerged around the solution" (Nunamaker Jr. et al., 2015, p. 43). However, their contribution fails to provide further guidance on how to conduct the respective evaluations.

In conclusion, there is currently no methodological advice on how to conduct research in longitudinal ambient display deployments.

Methodological Foundations for Longitudinal Ambient Display In Situ Research

This paper aims to provide a first stepping stone towards developing methodological foundations for long-term ambient display research in the wild. This development process builds on an ongoing study, where we utilise our custom Ambient Surfaces solution. The study seeks to understand, how the solution is appropriated in an authentic environment. We propose building this knowledge-seeking process on classic GT as *a* way of conducting such enterprises. To this end, this paper illustrates a practical systematisation of GT's two core processes—constant comparison and theoretical sampling. While the next section thoroughly elaborates this practical systematisation, the following emphasises the overall research setting and purpose, envisions classic GT as a means to cope with the challenges in field deployment research, and briefly introduces GT methodology.

The Study: Research Setting and Purpose

For field deployment research, the complexity and scope are embodied in choices made with respect to the *target population*, *scale*, and *duration* (Siek et al., 2014). Accordingly, this paper subsequently focuses on these topics. More details regarding the research setting and the custom solution can be found in previous work (Schwarzer et al., 2016).

The field deployment commenced in February 2014 with one large and interactive screen, while a second one followed in August 2015. Data collection is still ongoing in 2019. Our Ambient Surfaces solution is deployed in a German company that can be characterised as a large-scale agile software development environment with eight agile teams at present (Dingsøyr et al., 2014). Foremost, the practice of Scrum (Schwaber and Beedle, 2001) is adhered to and accompanied by selected practices from Extreme Programming (Beck, 2000). The Ambient Surfaces were located in two distinct locations and show information from different tools utilised in the department. In the first setting, roughly 70 to 80 people had access to the screens—this included Scrum Masters, Product Owners, management personnel, and foremost software developers (i.e. almost two-thirds). Around 90% of these staff members were between 31 and 50 years old and approximately 75% of them had been working in the company for at least three years. The number of potential users increased substantially to over 400 from the entire company's campus in the second setting (including further management personnel and consultants). This is due to the fact that both systems were relocated in 2017 to a newly constructed building which also includes a canteen. Generally, a large number of passers-by is characteristic for this new setup as it was for the old one. For in-

Figure 1. Both Ambient Surfaces in their current setup as of 2019, situated in a hallway near the canteen. In this instance, the left system mostly shows contents from the 'Confluence View' (e.g. architecture articles) and the right screen illustrates information from the 'Test Suites View' (e.g. list of test suites).

stance, people typically have to walk past the systems when having lunch in the canteen. Figure 1 shows the setup in its current configuration as of 2019.

Fundamentally, the study sets out to contribute missing longitudinal findings of ambient displays in real world contexts (Schwarzer et al., 2016). Our ongoing multiple-year enterprise embarks on generating a substantive theory—i.e. a theory that sheds light on a particular empirical area in the real world (Glaser, 1978)—which conceptually explains how the Ambient Surfaces solution is appropriated in this particular setting. Foremost, we are interested in what the literature refers to as 'naturalistic usage' (Siek et al., 2014) rather than, for instance, usage originating in instances of a novelty effect (Koch et al., 2018). We position our research in the domain of proof-of-use research as we are largely dealing with issues surrounding operational feasibility (Nunamaker Jr. et al., 2015).

Fit of Classic Grounded Theory

To rigorously strive towards our research goal, we pondered over an appropriate research methodology. Due to the issues below, we finally decided to utilise classic GT as the methodological foundation for our research:

1. Fundamentally, GT methodology sheds light on *social* phenomena, independently of a particular research discipline (Glaser and Strauss, 1967). It therefore assists in coping with the issue that socially embedded technology cannot be investigated without its social components (Bjørn and Boulus-Rødje, 2015).

2. Furthermore, GT enables one to deal with the dynamic nature of field deployments as it is considered messy and may require changes in the data collection procedure (Siek et al., 2014). GT does not ask for preconceiving of any sort of data but asks to let the data emerge and to openly chose the most appropriate data collection method (Glaser, 2008).

3. Classic GT utilises both quantitative and qualitative methods beyond boundaries of specific research paradigms such as positivism and constructivism as it is considered a general methodology (Glaser, 1998). In fact, Glaser (1998, 2008) considers all kinds of data as valuable in the process of generating theory (e.g. documents, magazines, and interviews). In comparison to other GT variants such as Straussian GT (Corbin and Strauss, 2015), classic GT suits the requirements of field deployment research arguably better as, typically, a mixture of different methods (e.g. interviews, observations, and log files) is utilised in such endeavours (Alt et al., 2012). This methodological openness primarily led to the decision to commence our research with classic GT.

4. As long-term in situ research is such an unexplored territory (Börner et al., 2013; Hazlewood et al., 2011; Preim et al., 2018), scarce theoretical guidance arguably exists to pose any initial research questions or hypotheses. In fact, ambient display research lacks existing general theories (Alt et al., 2012). GT follows the notion of starting any research open-minded without any preconceived problem statements (Glaser, 2008). It asked the open question of "What's going on[here?]" (Glaser and Strauss, 1967, p. 97), which is—in one form or the other (e.g. Glaser and Strauss, 1967; Glaser, 1978, 1992)—the opening question in every GT study (Charmaz, 2006).

5. Proof-of-use research generally faces the issue of externalising and codifying a researcher's tacit knowledge (Nunamaker Jr. et al., 2015). GT provides a means to report a researcher's own experiences and thereby increases the traceability and credibility of a study (Boeije, 2002). Fundamentally, GT strives towards situating study participants' actions and interpretations in the relevant circumstances and thus making them explicit (Morse et al., 2009).

Introduction to Grounded Theory Methodology

In the 1960s, GT was an inductive response to predominant hypothetico-deductive research approaches (Glaser and Strauss, 1967; Morse et al., 2009; Stol et al., 2016). Over the last decades, however, GT evolved from its origins into two major streams: since the 1990s referred to as 'Glaserian GT' or 'classic GT' (Glaser, 1978, 1992, 2006; Stern, 1995) and 'Straussian GT' (Corbin and Strauss, 2008, 2015; Strauss, 1987; Strauss and Corbin, 1990, 1998). Following this development, further variants emerged, which are summarised under the term 'second generation' (Morse et al., 2009; Muller, 2014). According to Morse et al. (2009), differences in GT variants arise from epistemological stances, methodological

strategies, assumptions about what constitutes theory, and lastly conceptional directions.

Principally, GT represents a way of thinking about and conceptualising based on data (Morse et al., 2009). It is aimed at proposing grounded hypotheses, not facts (Glaser, 1978, 2008). Descriptions are put forth to foster an understanding of the rationale behind hypotheses (Glaser, 1978). Generally, it is geared towards discovering *a* not *the* theory (Heath and Cowley, 2004). Following Glaser (1978, p. 4), "... a theory should be able to explain what happened, predict what will happen and interpret what is happening in an area of substantive or formal inquiry." The process of generating theory is a continuous one of modification. Glaser (1978) refers to this attribute of a theory as 'modifiability'. GT is acknowledged to be a methodology which is a highly individually conducted endeavour (Morse et al., 2009). It fundamentally turns a human weakness into a strength as it allows the researcher to theorise about data during analyses whilst explaining data to oneself or colleagues (Muller and Kogan, 2012). As Muller and Kogan (2012) further note, GT seeks to formalise this cognitive process into a quality process to generate new insights and theories.

Practical Systematisation of the Analysis Process

Guided by other GT studies that structured their research in different phases (e.g. Boeije, 2002; Walsh, 2015), we organised our research in *four* research phases. In the first phase, only quantitative interaction data was considered. Subsequently, observations, a group interview, and a survey enriched the theory generating process in the second phase. While during the third phase statistical tests were additionally conducted, the last stage is ongoing and incorporates a respondent validation to conclude the research.

Below, it is concentrated on the practical systematisation of GT's two core processes during the first stage, denoted as 'Phase 1'. Therefore, we initially describe the theoretical underpinnings of the analysis process and subsequently highlight findings obtained by adhering to this rationale.

How the Analysis was Conducted

Fundamentally, our work builds on a constant comparison step-by-step approach introduced by Boeije (2002). She suggests *four* distinct criteria to be elaborated throughout each step: firstly, the analysis activities (i.e. a description); secondly, the aim of comparisons; thirdly, important questions asked and, lastly, the findings (see next section). In the following, it is now continued with the application of the first three of these four criteria in Phase 1, starting with the aim of comparisons, the questions, and finally a description of the comparison process.

Aim of Comparisons

The primary aim of comparisons in Phase 1 was to reveal latent patterns in interaction data. To this end, the identified patterns were used to state initial hypotheses about the actual utilisation, to pose new questions, and consequently to guide future data collection activities. It was intended to generate descriptive figures which provided a first theoretical glance at the issue of utilisation.

Important Questions

The following two questions were of particular relevance in Phase 1:

1. *During what daytimes is the Ambient Surface most prominently being utilised?* This question aimed at shedding light on the first emerging latent patterns that were prevalent in the material.

2. *How long can a novelty effect be notably observed in the collected data?* With this question, it was intended to further investigate the anticipated novelty effect at the beginning of the study.

Description of Comparison Activities

In particular, the guidelines regarding the use of quantitative data in GT had been consulted in Phase 1 (Glaser, 2008). Accordingly, the concept of *crude indices* had been applied, mainly for two reasons. Firstly, they reportedly "... suffice to indicate the concepts of the theory ..." (Glaser, 2008, p. 41). Secondly, the material at hand felt to be unsuitable to proceed any further with respect to the elaboration analysis introduced in quantitative GT as item discovery was not the issue at hand.

Below, the items of the crude index (i.e. the comparison candidates), the memoing process, and the sampling strategy are introduced.

Declaring Comparison Candidates The relevant interaction data log file included different variables containing information relating to *touch events*, which are triggered in the software framework when a person interacts with the display's surface (e.g. variable *Timestamp of Event*). Analogously to Glaser (2008), every variable in this file represented an *item* for a possible crude index. The general aim is to create a crude index, incorporating at least two items to indicate the concepts of a theory by leveraging cross-tabulations. However, the selection of items is a highly individual choice. Glaser (2008) refers to this process as a type of piloting study, given that the researcher literally plays with a set of different items.

The variable *Timestamp of Event* seemed most promising, primarily due to its nature to describe usage over time. A crude index *Utilisation of the Ambient Surface* consisting of two items was created. Firstly, the item *Number of Interactions*, which summarises data from *Timestamp of Event*. Secondly, the item *Daytimes* was utilised. This variable cuts *Number of Interactions* in temporal segments (i.e.

27 half-hourly segments from 07:00 to 20:00). The chosen time frame felt reasonable as it happened to account for the majority of interactions.

Memos and Diagrams Fundamentally, Glaser (1978) considers memos as the core stage of each GT research endeavour. Yet, he fails to clearly explain how memos and diagrams were utilised during the elaboration analysis in quantitative GT (Glaser, 2008). However, other books on GT provide sufficient detail on this matter (e.g. Charmaz, 2006; Corbin and Strauss, 2015; Glaser, 1978, 1998).

One distinct feature of memos and diagrams in this work is that they are primarily digital and, in some cases, printed pieces of data artefacts. Corbin and Strauss (2015), on the contrary, utilised long hand-written memos to reflect on their research process. While they found that qualitative data entails complex and cumulative thinking, this work initially considered only quantitative data. However, Corbin and Strauss (2015) acknowledge that memo-writing is a highly individual process and conclude that the important part is that memos are created, especially in longitudinal research.

A Microsoft Excel file was created to store memos, incorporating the aforesaid crude index with its items in cross-tabulations. This file also left room for additional notes and comments. For example, software bugs, change requests, and relevant correspondences were also documented in this file. A memo wall in the authors' office space was utilised to collaboratively reflect on data and simultaneously to inspect and sort memos at a glance (see Figure 2).

Figure 2. Parts of the utilised memo wall displaying a diverse set of touch events statistics, user interface screenshots, and charts in the authors' office space.

Theoretical Sampling The following suggestions were considered in approaching the sampling procedure: firstly, Muller (2014) encourages researchers to

choose methods which allow them best to perceive and know and, secondly, Stol et al. (2016) recommend selecting a primary data source as the basis for further data collection activities.

In the end, it was decided to select the Ambient Surfaces' custom interaction logging mechanism as the primary data source throughout the research. The motivation was fourfold:

- Firstly, interaction data allows one to shed light on a variety of usage patterns (e.g. content utilisation). Sensor data to track user activities is commonly utilised in ambient display research (Börner et al., 2013). Generally, logging is considered helpful in long-term enterprises (Alt et al., 2012).
- Secondly, phenomena such as the novelty effect and display blindness pose certain relevance (Koch et al., 2018). It was initially anticipated that a novelty effect would be present to some extent. Interaction data arguably allows the identification of uncommon patterns in the material.
- Thirdly, this method helped in keeping initial resources in check. For instance, some studies report that they extended their research due to the prevalence of an initial novelty effect (e.g. Gallacher et al., 2015; Hazlewood et al., 2011). Additionally, behavioural sciences show that it can take up to several months until a new behaviour takes hold (Prochaska and Diclemente, 1982). Data collection techniques such as observations would had arguably accounted for more time-intensive workloads (e.g. travel time), which is also mentioned by Corbin and Strauss (2015).
- Lastly, as it is crucial in ambient display research to collect data unobtrusively (Börner et al., 2013), the logging mechanism arguably allows for the collection of data without distracting potential users.

Initially, it was anticipated to sample data for at least a couple of months, primarily due to the novelty effect. Analyses were scheduled to be conducted weekly. Generally, data saturation is a crucial part of GT research (Corbin and Strauss, 2015). At its core, it develops—or saturates—the core categories of an emerging theory (Corbin and Strauss, 2015; Glaser, 2008). In Phase 1, data saturation referred to an incremental mitigation process that resulted from the decreasing impact of newly integrated interaction data on manifested latent patterns.

Findings of the Analysis

Below, findings obtained by continuously comparing data in Phase 1 are discussed. However, as this paper primarily concerns the methodological foundations of our research, this section exemplary presents some results stemming from this comparison process and indicates conclusions (i.e. hypotheses and emerging questions) drawn on its basis.

Analyses Stages

Overall, analyses in 2014 happened to be organised in four stages, while each of these stages concerned interaction data from periods of up to several weeks. This organisation is a result of particular events and observations. These four stages were:

- *Weeks 8 to 9 (Stage 1)*: As the Ambient Surface was operational on a Friday right before midday, it was decided to initiate analyses with the data from the first two weeks (i.e. six working days).
- *Weeks 10 to 11 (Stage 2)*: It was then decided to summarise the analysis activities from weeks 10 and 11, primarily as in Week 11 an update was deployed. Simultaneously, Week 11 marked the end of the reported novelty effect's overall existing time in some studies (e.g. Gallacher et al., 2015; Hazlewood et al., 2011).
- *Weeks 12 to 19 (Stage 3)*: Here, the analyses stood in light of the now arguably less prevalent effects of novelty towards the data. It was focused on a longer time period for two main reasons: firstly, while comparing the weeks in question, it was found that weeks 8 to 11 indicated the highest number of interactions; secondly, a five-week time period followed, in which the Ambient Surface was largely not operational due to software-related issues.
- *Weeks 20 to 52 (Stage 4)*: Of particular interest in this stage was whether the Ambient Surface could tackle the threats stemming from display blindness beyond a novelty effect and whether the previously obtained patterns endured over time. Consequently, it was decided to focus on analysing interaction data until the end of the year. The reason for this decision was threefold: firstly, to allow a profound comparison, it felt reasonable to collect a notable amount of data; secondly, as the Ambient Surface was lastly revised in Week 16 and the project partner was still occasionally reviewing its contents, it was hypothesised that more time had to pass to grasp on issues relating to display blindness; finally, the almost entirely non-operational state of the screen in weeks 20 to 24 posed a particular threat to display blindness as this issue could have resulted in disuse of the system afterwards.

Throughout all stages, memos in the form of what is depicted in Figure 3 were intensively being utilised. While cross-tabulations in Stage 1 provoked the idea that events such as arriving at work might play an important role relating to actual usage, Stage 2 revealed that changes to the system resulted in a measurable difference in the total number of interactions. With respect to display blindness, it was observable during Stage 3 that the Ambient Surface was still frequently being utilised. While the total number of interactions notably decreased, the system was seemingly incorporating some positive contributing factors. Otherwise, it was believed that the usage would have dropped more substantially or would have halted entirely. In Stage 4, it seemed that data became more representative compared to data stemming from preceding weeks. It was concluded that threats resulting from display blindness were likely to be prevalent to some extent, but it was also recog-

Daytimes	# of	# of (in %)
07:00	0	0.00%
07:30	0	0.00%
08:00	0	0.00%
08:30	0	0.00%
09:00	0	0.00%
09:30	0	0.00%
10:00	0	0.00%
10:30	0	0.00%
11:00	0	0.00%
11:30	0	0.00%
12:00	167	8.96%
12:30	161	8.64%
13:00	259	13.90%
13:30	302	16.21%
14:00	317	17.02%
14:30	27	1.45%
15:00	61	3.27%
15:30	5	0.27%
16:00	24	1.29%
16:30	208	11.16%
17:00	176	9.45%
17:30	36	1.93%
18:00	120	6.44%
18:30	0	0.00%
19:00	0	0.00%
19:30	0	0.00%
20:00	0	0.00%
Total:	**1,863**	**100.00%**

Daytimes	# of	# of (in %)
07:00	0	0.00%
07:30	47	1.22%
08:00	139	3.62%
08:30	155	4.03%
09:00	194	5.05%
09:30	83	2.16%
10:00	117	3.04%
10:30	293	7.62%
11:00	113	2.94%
11:30	146	3.80%
12:00	324	8.43%
12:30	184	4.79%
13:00	330	8.58%
13:30	213	5.54%
14:00	234	6.09%
14:30	159	4.14%
15:00	195	5.07%
15:30	86	2.24%
16:00	215	5.59%
16:30	128	3.33%
17:00	109	2.83%
17:30	125	3.25%
18:00	122	3.17%
18:30	97	2.52%
19:00	37	0.96%
19:30	0	0.00%
20:00	0	0.00%
Total:	**3,845**	**100.00%**

Daytimes	# of	# of (in %)
07:00	0	0.00%
07:30	47	0.82%
08:00	139	2.44%
08:30	155	2.72%
09:00	194	3.40%
09:30	83	1.45%
10:00	117	2.05%
10:30	293	5.13%
11:00	113	1.98%
11:30	146	2.56%
12:00	491	8.60%
12:30	345	6.04%
13:00	589	10.32%
13:30	515	9.02%
14:00	551	9.65%
14:30	186	3.26%
15:00	256	4.48%
15:30	91	1.59%
16:00	239	4.19%
16:30	336	5.89%
17:00	285	4.99%
17:30	161	2.82%
18:00	242	4.24%
18:30	97	1.70%
19:00	37	0.65%
19:30	0	0.00%
20:00	0	0.00%
Total:	**5,708**	**100.00%**

Figure 3. Three memos including the crude index *Utilisation of the Ambient Surface*, collating interaction data from Week 8 (left), Week 9 (centre), and weeks 8 and 9 in combination (right). While darker blue segments indicate stronger usage, white segments indicate the opposite.

nised that employees continued utilising the system. Therefore, the obtained latent patterns arguably became sufficiently saturated to seek additional material (e.g. observational data) to be included in the overall analysis and, hence, to strengthen theoretical sensitivity.

2014: A Retrospective

Before Phase 1 was concluded, two further investigations were retrospectively conducted: firstly, the all-year usage in terms of interactions per calendar week was elaborated and, secondly, data representativeness was analysed.

Comparing all of the interaction data from 2014, it was found that the mean number of interactions per week first fell below the all-year mean number of interactions in Week 18. We demonstrated this eleventh week into the field study elsewhere as the threshold which indicated that the initial novelty effect finally subsided (Koch et al., 2018). In investigating data representativeness, we intended to identify the individual interaction incidents that potentially affected conclusions drawn from analyses. In the end, it was not evident that the results during weeks 18 to 52 were substantially affected by any single incident. The Ambient Surface was used regularly and on most days at around 12:30 (see Figure 4).

Consequently, it was assumed that the findings showed 'naturalistic usage' (Siek et al., 2014) to a certain degree beyond both a novelty effect and threats relating to display blindness.

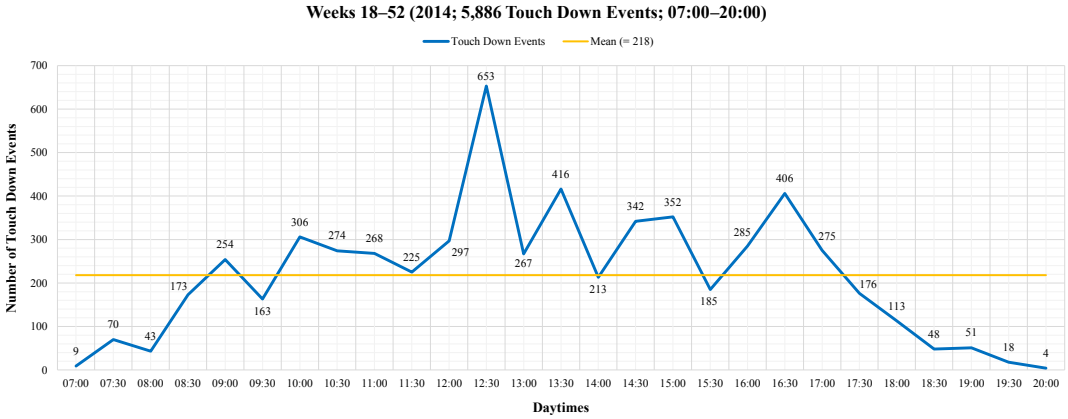

Figure 4. The total number of interactions (i.e. touch down events) with respect to corresponding daytimes for weeks 18 to 52 in 2014.

Conclusions

It is concluded with an illustration of hypotheses and questions which emerged throughout Phase 1. Both represented the basis for subsequent research phases and fostered theoretical sensitivity. Overall, the rather descriptive findings indicated that appropriation might be somewhat related to *informal* occasions (e.g. lunch breaks). In summary, the following hypotheses were posed:

- *The Ambient Surface is notably utilised in the early morning when people are arriving at work*
- *The Ambient Surface is mostly utilised when people are going to or are returning from lunch*
- *The Ambient Surface is notably utilised in the late afternoon when people are leaving work*
- *There are daytimes in which the use of the Ambient Surface notably decreases*
- *A novelty effect results in the usage of the Ambient Surface which distracts from latent patterns due to the magnitude of interactions and the time of their occurrence*
- *Changes to an existing system contribute towards or extend a prevalent novelty effect*
- *The Ambient Surface provides positive contributing factors beyond a prevalent novelty effect*

In addition to these hypotheses, Phase 1 revealed questions that largely surrounded the limitations of quantitative interaction data. These questions included:

- *What positive contributing factors is the Ambient Surface promoting?*
- *What are the reasons for the varying reductions in interactions throughout the day?*
- *How is the Ambient Surface passively utilised by staff members?*

- *How do the relationships of variables investigated in the analysis change when incrementally compared to additional interaction data?*
- *What can be learnt with respect to display blindness by utilising further data collection techniques?*
- *How do the conclusions regarding data saturation change when compared to additional interaction data from subsequent years?*

Discussion

The paper presents a practical systematisation of GT's two core concepts—constant comparison and theoretical sampling—in the early stages of our ongoing research. Generally, we do not argue that our approach is the modus operandi to conduct longitudinal ambient display in situ studies. For example, Siek et al. (2014) recommend a more sequential organisation of the research process, where analyses are carried out at the end of the field research. In fact, this post-deployment analysis approach is also chosen by some GT studies in the context of HCI and CSCW (Muller, 2014). Above, further utilised methodologies were also introduced (e.g. heuristic inquiries)—admittedly, other circumstances may seek a different methodological choice (e.g. limited time resources). In this respect, we seek to illustrate *a* way to methodologically guide in situ ambient display research.

Given that we are fundamentally interested in evaluating 'naturalistic usage' (Siek et al., 2014), we have committed to the challenge that research prototypes typically do not withstand daily use in authentic environments (Nunamaker Jr. et al., 2015; Siek et al., 2014). It is with the utmost certainty that the Ambient Surfaces would not be in operation as of 2019 had we not committed to this proactive engagement. Consequently, we faced several challenges throughout the entire study. As the Ambient Surfaces did and do change regularly (see Figure 5), so did and does the environment, including staff members, meeting schedules, holiday and illness seasons, as well as trainings—to name but a few. Following GT helped us to stay sensitive towards the data. For example, sometimes there were no interactions during an entire week. We immediately started to ask questions and postulate possible explanations. Here, the primary data source served as a sensitive indicator to rapidly start wondering about the data. We found this quantitative data source to be very helpful, especially when the research commenced but also later during the study (e.g. to isolate a novelty effect).

We find it difficult to convey the tacit knowledge regarding issues that are not directly presented in charts, diagrams, and statistical tests. However, as we are constantly comparing data and theorising about the implications, GT ensures that we explicitly track the progress while, for instance, asking questions and conducting analyses. Overall, by only considering such occasionally intertwined issues and by bringing them to the fore, we argue that field deployment reports reach their full potential and the reader is able to enjoy a text to its full extent.

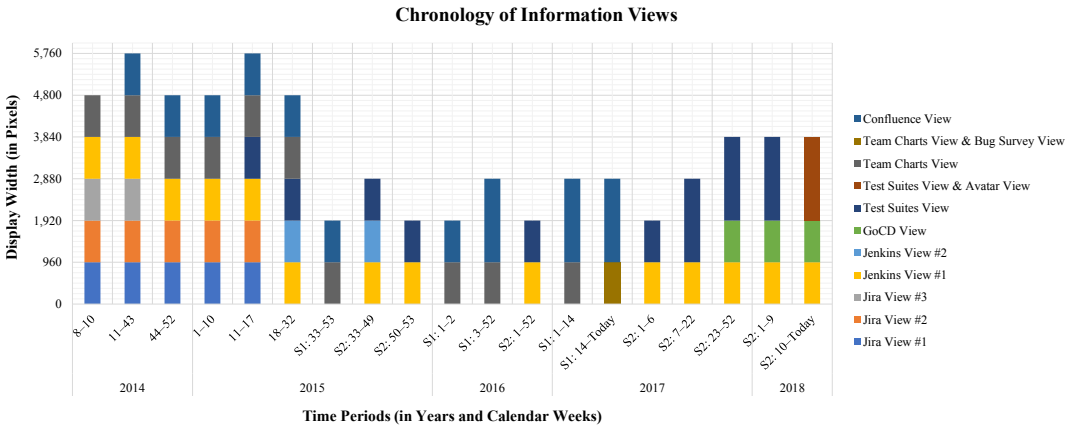

Figure 5. A bar chart diagram depicting the evolution of the Ambient Surfaces (i.e. 'S1' and 'S2' as of Week 33 in 2015) over time by relating added, removed, and resized (i.e. increases in width) information views to actual display widths (i.e. in pixels) and time periods.

Contributions

The following elaborates the different contributions this paper adds to existing knowledge, categorised in terms of field deployments, ambient displays, GT methodology, and the novelty effect.

Long-term Field Deployment Research

- Firstly, by shedding light on our ongoing long-term field deployment study, this paper responds to recent calls for more longitudinal in situ evaluations (Börner et al., 2013; Hazlewood et al., 2011; Preim et al., 2018).

Methodological Advances in Ambient Display In Situ Research

- Secondly, by envisioning the application of classic GT in ambient display field deployment research, the paper introduces a way to cope with the lack of methodological development in this domain (Hazlewood et al., 2011).

Grounded Theory Methodology

- Thirdly, it contributes to the general rare application of GT methodology in the domains of software engineering (Stray et al., 2016) as well as HCI and CSCW (Muller and Kogan, 2012).
- Fourthly, it demonstrates a practical systematisation of both core processes in GT. The literature remains silent as to how to proceed during the constant comparison process and fails to indicate what constitutes fertile comparison candidates to develop a theoretical model (Boeije, 2002). Stol et al. (2016) also note that the process of theoretical sampling remains unclear, particularly considering the implications of data magnitude in this process.

- Fifthly, it envisions the utilisation of quantitative data in GT research. There have been no recent attempts in this regard (Glaser, 2008).
- Sixthly, as the systematisation approach is used with both quantitative and qualitative data, this paper puts forward a means to use a mixed-methods approach in GT. Rarely are both data sources combined in GT research (Walsh, 2015).

Novelty Effect Research

- Seventhly and lastly, in addition to a discussion of selected examples of the novelty effect in a previous publication (Koch et al., 2018), this work illustrates how we coped with this effect on a methodological level.

Limitations

Overall, our study is conducted in one particular environment—other settings will likely reveal notable differences in terms of comparisons and the sampling strategy. Furthermore, as there is little practical guidance on conducting GT with quantitative data, it is possible that we misinterpreted certain parts that Glaser (2008) foresees for such research. Additionally, the issue of incorporating literature in GT research was disregarded (Giles et al., 2013), primarily due to the fact that we would have had to frame the paper more holistically. In a similar vein, the paper only briefly draws on the complexity of our research and does not convey any information on how the descriptive findings from Phase 1 transcended into conceptually representative categories. Also, social aspects such as how usage relates to practices (e.g. team meetings) are not further considered. Again, Phase 1 was intended as being the first stepping stone towards the goal of generating a substantive theory.

Conclusion and Recommendations for Future Work

In response to the lacking methodological advances in ambient display in situ research, this paper illustrates the application of classic GT in our ongoing longitudinal study. It seeks to put forward practical guidance for the researcher to go along with the methodology's two core concepts, namely constant comparison and theoretical sampling. To this end, it demonstrates a practical systematisation of both processes to compare and sample quantitative interaction data. The concept of crude indices was leveraged to unveil latent patterns and to indicate directions for future data collection as well as analysis activities on the basis of emergent hypotheses and questions.

In terms of future research, we propose the following directions. Firstly, a promising avenue would be to apply our approach to further quantitative analyses in related studies. Secondly, we encourage other researchers to conduct studies with GT as it has yet to gain momentum in HCI and CSCW research (Muller and

Kogan, 2012). Thirdly, irrespective of the selected research methodology, this paper puts forward the issue that in situ research is highly relevant (Börner et al., 2013; Hazlewood et al., 2011; Nunamaker Jr. et al., 2015; Preim et al., 2018; Siek et al., 2014). Consequently, any research with any chosen methodology conducted in this manner, would contribute valuable knowledge to the community. Fourthly and lastly, we concur with Bjørn and Boulus-Rødje (2015) and invite other researchers to rethink their research approaches when planning to conduct research in dynamic and heterogeneous environments. With the selection of a primary data source (e.g. to cope with the novelty effect), we hope to indicate some first stepping stones in this regard.

References

Alt, F., S. Schneegaß, A. Schmidt, J. Müller, and N. Memarovic (2012): 'How to Evaluate Public Displays'. In: *Proceedings of the 2012 International Symposium on Pervasive Displays*. New York, NY, USA, pp. 17:1–17:6, ACM.

Bannon, L. and C. O'Malley (1984): 'Problems in Evaluation of Human-Computer Interfaces: A Case Study'. In: B. Shackel (ed.): *Proceedings of IFIP INTERACT'84: Human-Computer Interaction*. Amsterdam, pp. 709–713, North-Holland.

Beck, K. (2000): *Extreme Programming Explained: Embrace Change*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.

Bjørn, P. and N. Boulus-Rødje (2015): 'Studying Technologies in Practice: "Bounding Practices" When Investigating Socially Embedded Technologies'. In: V. Wulf, K. Schmidt, and D. Randall (eds.): *Designing Socially Embedded Technologies in the Real-World*. London: Springer, pp. 341–356.

Boeije, H. (2002): 'A Purposeful Approach to the Constant Comparative Method in the Analysis of Qualitative Interviews'. *Quality and Quantity*, vol. 36, no. 4, pp. 391–409.

Börner, D., M. Kalz, and M. Specht (2013): 'Closer to You: Reviewing the Application, Design, and Evaluation of Ambient Displays'. *Int. J. Ambient Comput. Intell.*, vol. 5, no. 3, pp. 16–31.

Charmaz, K. (2006): *Constructing Grounded Theory*. Thousand Oaks, CA, USA: Sage.

Corbin, J. M. and A. L. Strauss (2008): *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Thousand Oaks, CA, USA: Sage, 3rd edition.

Corbin, J. M. and A. L. Strauss (2015): *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Thousand Oaks, CA, USA: Sage, 4th edition.

Crotty, M. J. (1998): *The Foundations of Social Research: Meaning and Perspective in the Research Process*. London: Sage Publications Ltd.

Dingsøyr, T., T. E. Fægri, and J. Itkonen (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, and M. Raatikainen (eds.): *Product-Focused Software Process Improvement*. Cham, pp. 273–276, Springer International Publishing.

Gallacher, S., J. O'Connor, J. Bird, Y. Rogers, L. Capra, D. Harrison, and P. Marshall (2015): 'Mood Squeezer: Lightening Up the Workplace Through Playful and Lightweight Interactions'. In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. New York, NY, USA, pp. 891–902, ACM.

Giles, T., L. King, and S. de Lacey (2013): 'The Timing of the Literature Review in Grounded Theory Research An Open Mind Versus an Empty Head'. *Advances in Nursing Science*, vol. 36, no. 2, pp. E29–40.

Glaser, B. G. (1978): *Theoretical Sensitivity*. Mill Valley, CA, USA: Sociology Press.

Glaser, B. G. (1992): *Emergence versus forcing: Basics of grounded theory analysis*. Mill Valley, CA, USA: Sociology Press.

Glaser, B. G. (1998): *Doing Grounded Theory: Issues and Discussions*. Mill Valley, CA, USA: Sociology Press.

Glaser, B. G. (2006): *Doing formal grounded theory: A proposal*. Mill Valley, CA, USA: Sociology Press.

Glaser, B. G. (2008): *Doing Quantitative Grounded Theory*. Mill Valley, CA, USA: Sociology Press.

Glaser, B. G. and A. L. Strauss (1967): *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York, NY, USA: Aldine de Gruyter.

Hazlewood, W. R., E. Stolterman, and K. Connelly (2011): 'Issues in Evaluating Ambient Displays in the Wild: Two Case Studies'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA, pp. 877–886, ACM.

Heath, H. and S. Cowley (2004): 'Developing a grounded theory approach: a comparison of Glaser and Strauss'. *International Journal of Nursing Studies*, vol. 41, no. 2, pp. 141–150.

Koch, M., K. von Luck, J. Schwarzer, and S. Draheim (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 of *Reports of the European Society for Socially Embedded Technologies*. European Society for Socially Embedded Technologies (EUSSET).

Matthews, T., T. Rattenbury, and S. Carter (2007): 'Defining, Designing, and Evaluating Peripheral Displays: An Analysis Using Activity Theory'. *Human–Computer Interaction*, vol. 22, no. 1-2, pp. 221–261.

Memarovic, N., I. Elhart, and E. Rubegni (2016): 'Developing a Networked Public Display System'. *IEEE Pervasive Computing*, vol. 15, no. 3, pp. 32–39.

Messeter, J. and D. Molenaar (2012): 'Evaluating Ambient Displays in the Wild: Highlighting Social Aspects of Use in Public Settings'. In: *Proceedings of the Designing Interactive Systems Conference*. New York, NY, USA, pp. 478–481, ACM.

Morse, J. M., P. Noerager Stern, J. Corbin, B. Bowers, A. E. Clarke, and K. Charmaz (2009): *Developing Grounded Theory: The Second Generation*. London: Routledge.

Muller, M. (2014): 'Curiosity, Creativity, and Surprise as Analytic Tools: Grounded Theory Method'. In: J. S. Olson and W. A. Kellogg (eds.): *Ways of Knowing in HCI*. New York, NY: Springer, pp. 25–48.

Muller, M. J. and S. Kogan (2012): 'Grounded Theory Method in Human-Computer Interaction and Computer-Supported Cooperative Work'. In: J. A. Jacko (ed.): *The Human Computer Interaction Handbook*. Boca Raton, FL, USA: CRC Press, 3rd edition, pp. 1003–1024.

Nunamaker Jr., J. F., R. O. Briggs, D. C. Derrick, and G. Schwabe (2015): 'The Last Research Mile: Achieving Both Rigor and Relevance in Information Systems Research'. *Journal of Management Information Systems*, vol. 32, no. 3, pp. 10–47.

Ojala, T., V. Kostakos, H. Kukka, T. Heikkinen, T. Linden, M. Jurmu, S. Hosio, F. Kruger, and D. Zanni (2012): 'Multipurpose Interactive Public Displays in the Wild: Three Years Later'. *Computer*, vol. 45, no. 5, pp. 42–49.

Peltonen, P., E. Kurvinen, A. Salovaara, G. Jacucci, T. Ilmonen, J. Evans, A. Oulasvirta, and P. Saarikko (2008): 'It's Mine, Don'T Touch!: Interactions at a Large Multi-touch Display in a City Centre'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA, pp. 1285–1294, ACM.

Preim, B., T. Ropinski, and P. Isenberg (2018): 'A Critical Analysis of the Evaluation Practice in Medical Visualization'. In: A. Puig Puig, T. Schultz, A. Vilanova, I. Hotz, B. Kozlikova, and P.-P. Vázquez (eds.): *Eurographics Workshop on Visual Computing for Biology and Medicine*. Granada, Spain, pp. 45–56, The Eurographics Association.

Prochaska, J. and C. Diclemente (1982): 'Transtheoretical therapy: Toward a more integrative model of change'. *Psychotherapy: Theory, Research & Practice*, vol. 19, no. 3, pp. 276–288.

Rogers, Y., W. R. Hazlewood, P. Marshall, N. Dalton, and S. Hertrich (2010): 'Ambient Influence: Can Twinkly Lights Lure and Abstract Representations Trigger Behavioral Change?'. In: *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*. New York, NY, USA, pp. 261–270, ACM.

Schwaber, K. and M. Beedle (2001): *Agile Software Development with Scrum*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1st edition.

Schwarzer, J., S. Draheim, K. von Luck, Q. Wang, P. Casaseca, and C. Grecos (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, pp. 69:1–69:4, ACM.

Shelton, B. and K. Nesbitt (2017): 'Evaluating WaveWatch: An Ambient Display of Web Traffic'. In: *Proceedings of the Australasian Computer Science Week Multiconference*. New York, NY, USA, pp. 9:1–9:9, ACM.

Siek, K. A., G. R. Hayes, M. W. Newman, and J. C. Tang (2014): 'Field Deployments: Knowing from Using in Context'. In: J. S. Olson and W. A. Kellogg (eds.): *Ways of Knowing in HCI*. New York, NY: Springer, pp. 119–142.

Stern, P. N. (1995): 'Eroding grounded theory'. In: J. M. Morse (ed.): *Critical issues in qualitative research methods*. Thousand Oaks, CA, USA: Sage, pp. 212–223.

Stol, K. J., P. Ralph, and B. Fitzgerald (2016): 'Grounded Theory in Software Engineering Research: A Critical Review and Guidelines'. In: *2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)*. New York, NY, USA, pp. 120–131, ACM.

Strauss, A. L. (1987): *Qualitative Analysis for Social Scientists*. New York, NY, USA: Cambridge University Press.

Strauss, A. L. and J. M. Corbin (1990): *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Newbury Park, CA, USA: Sage.

Strauss, A. L. and J. M. Corbin (1998): *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Newbury Park, CA, USA: Sage, 2nd edition.

Stray, V., D. I. K. Sjøberg, and T. Dybå (2016): 'The daily stand-up meeting: A grounded theory study'. *The Journal of Systems and Software*, vol. 114, no. April 2016, pp. 101–124.

Walsh, I. (2015): 'Using quantitative data in mixed-design grounded theory studies: an enhanced path to formal grounded theory in information systems'. *European Journal of Information Systems*, vol. 24, no. 5, pp. 531–557.

Wulf, V., C. Müller, V. Pipek, D. Randall, M. Rohde, and G. Stevens (2015): 'Practice-Based Computing: Empirically Grounded Conceptualizations Derived from Design Case Studies'. In: V. Wulf, K. Schmidt, and D. Randall (eds.): *Designing Socially Embedded Technologies in the Real-World*. London: Springer, pp. 111–150.