CARL VON OSSIETZKY universität OLDENBURG

Fakultät II – Informatik, Wirtschafts- und Rechtswissenschaften
Department für Informatik

Sensor-supported Unsupervised Observation Techniques for Field Studies

Dissertation zur Erlangung des Grades eines
Doktors der Ingenieurwissenschaften

vorgelegt von

Benjamin Poppinga

Gutachter:

Prof. Dr. Susanne Boll
Prof. Charlotte Magnusson, PhD

Tag der Disputation: 17.07.2014

Publication Details

Printed Edition

This thesis originally appeared as: *Sensor-supported Unsupervised Observation Techniques for Field Studies*, Oldenburg Computer Science Series, Vol. 32, X, 151 S., OlWIR (Oldenburger Verlag für Wirtschaft, Informatik und Recht), Edewecht, Germany, 2015, 49.80 EUR, ISBN-13 978-3-95599-019-0.

Author Contact Information

Benjamin Poppinga
bp@benjaminpoppinga.de
http://www.benjaminpoppinga.de/

Revision History

22nd January 2017	First digital edition, equals the printed edition.

Zusammenfassung

Wir leben in einem mobilen Zeitalter, in dem wir täglich und mehrfach mit alltagsdurchdringenden und interaktiven Schnittstellen und Anwendungen konfrontiert werden. Um für diese Umgebung geeignete Nutzungsschnittstellen zu designen und zu entwickeln, ist ein tiefgehendes Verständnis der Nutzer und der mobilen Situationen essentiell.

Gewöhnliche Laborversuche oder Beobachtungsmethoden für Laborstudien eignen sich kaum, um mobile Anwendungen, deren Nutzer oder das mobile Nutzererlebnis zu untersuchen, da sie der hohen Dynamik und der Flexibilität in der Benutzung nicht gerecht werden können. Deshalb werden mobile Anwendungen zunehmend in der natürlichen Umgebung in Feldstudien untersucht. Die dabei angewendeten Beobachtungsmethoden müssen es schaffen, dem Beobachter situationsnahe Einblicke zu gewähren, dürfen den Nutzer dabei nicht stören und müssen effizient sein, um viele Nutzer in kurzer Zeit beobachten zu können. Leider involvieren die meisten Beobachtungsmethoden einen menschlichen Beobachter und störende Beobachtungswerkzeuge, wie z.B. Kameras, was es schwierig macht, die genannten Anforderungen zu erfüllen.

Deshalb werden in Feldstudien zunehmend Beobachtungsmethoden eingesetzt und untersucht, die keinen menschlichen Beobachter mehr erfordern. Bekannte Beispiele für sogenannte unbegleitete Beobachtungsmethoden sind Logging, die Experience Sampling Method (ESM) oder Tagebuch-Studien. In der Regel wird der menschliche Beobachter dabei durch Sensoren oder durch vom Nutzer unmittelbar bereitgestellte Einblicke, z.B. Tagebucheinträge, ersetzt. Obwohl diese Techniken sich bereits wesentlich besser für die Untersuchung von mobilen Interaktionen eignen, sind sie dennoch nicht in der Lage, alle genannten Anforderungen gleichermaßen zu erfüllen.

In dieser Arbeit wird untersucht, wie sich unbegleitete Beobachtungsmethoden durch den Einsatz von Sensorik und modernen Datenverarbeitungstechniken auf einem Mobiltelefon weiter verbessern lassen. Es wird gezeigt, dass sich Informationsgehalt und Situationsbezug von Logging durch die Kombination mit statischen Informationen über die Umgebung und durch den Einsatz von statistischen Analysen verbessern lassen. Außerdem wird demonstriert, dass sich ESM durch maschinelles Lernen insoweit modifizieren lässt, dass Aufforderungen zur Eingabe von Daten an den Nutzer in geeigneten Momenten erscheinen und deshalb wahrscheinlicher beantwortet werden. Letztlich wird an zwei Beispielen illustriert, wie sich durch geschicktes Aufbereiten und Präsentieren von getätigten Beobachtungen ein im Anschluss an eine Studie stattfindendes Interview anreichern und verbessern lässt.

Diese Arbeit leistet zwei Beiträge. Zum einen stellt sie einen neuartigen, sensorbasierten Beobachtungsansatz vor, dessen Mehrwert und Vorteile in verschiedenen Studien und Szenarien demonstriert werden. Zum anderen wird ein zugehöriges, ganzheitliches Framework vorgestellt, das als Referenzimplementierung dient und die Anwendung meiner Methode durch interessierte Entwickler erleichtert. Beide Beiträge ermöglichen es Forschern und Wissenschaftlern, unbegleitete Studien durchzuführen, die effizient und unaufdringlich sind und einen hohen Situationsbezug gewährleisten.

Abstract

We live in a mobile era and ubiquitous applications and interactive services surround us everywhere. To design and develop targeted and successful applications for these environments a detailed understanding of the users and their mobile context is needed.

Traditional lab studies and lab-based observation methods are mostly unsuited to investigate mobile applications, their users and the users' experiences in such mobile and highly dynamic settings. Consequently, more and more research is done in field studies, where the users, applications and services are investigated in their natural environments. Common field observation methods, such as shadowing, are facing three essential challenges, i.e., situatedness, obtrusiveness, and scalability, and are in most cases unable to meet all of them. While mobile technologies continue to disappear, field observation techniques often still require the presence of human observers and observation tools, like video cameras.

In the last years, research has identified unsupervised observation techniques as promising approach for field studies. Popular examples for these techniques are logging, the Experience Sampling Method (ESM), and diary studies. In such techniques the human observer is replaced by sensors or self-reporting techniques, which are typically hosted on a mobile phone. However, although they inherently overcome many limitations of traditional field study observation techniques, they still fail to meet the three key challenges fully.

In this thesis, I investigate to what extent the sensing and computation capabilities of smart phones can improve unsupervised observation techniques. I demonstrate that the information gain of logging can be increased, if statistical analysis is applied and the sensor data is combined with static, environmental information. Further, I show that situation-aware self-reporting techniques, such as ESM, can trigger inquiries in opportune moments and, therefore, are perceived less obtrusive and are more likely to be answered. Finally, I illustrate that properly prepared and presented in situ information can serve as a valuable key resource for post hoc studies, like follow-up interviews.

I contribute a novel, sensor-based observation approach, which has shown its benefits and advantages in a set of structured field studies of different complexities and dynamics. Further, I contribute the Virtual Observer, a holistic observation framework and reference implementation of my observation approach, which facilitates use and application of my method in other contexts and domains. Both, the novel observation approach and its reference implementation, will enable researchers to conduct unobtrusive, scalable field studies that feature convincing situational details.

Contents

1 Introduction

"Focus on the user and all else will follow" is the first of ten philosophy principles of one of today's most influential and successful IT companies: Google[1]. This statement indicates, representative for many other companies, that a positive user experience is considered a key factor for customer satisfaction and market success. To design and develop a targeted and pleasant user experience, a detailed understanding of the users and their context is needed. Only such in-depth understanding truly allows to identify the users' underlying desires and problems, and to come up with a likely accepted and successful solution.

Although a company's overall philosophy is probably omnipresent, its practical implications are often vague and unclear. Design processes help to pursue defined goals and to implement related actions, e.g., by establishing an in-depth user understanding before any design or realisation activities start. The Human-Centred Design (HCD) process [ISO10] is a standardised design process which is widely applied in industry and academia. It is iterative and consists of four phases, of which the first two are to understand the context and to specify user requirements. The third phase is the actual design and prototyping step, which is followed by an evaluation of the design solution. Ideally, the user is involved in each of these four phases.

User observation is a key aspect of the HCD process, either to understand their needs or to evaluate their experiences with recent design solutions. In general, it can be differentiated between two types of observation approaches: surveys, i.e., directly asking the user for oral or written feedback, and ethnographic studies, i.e., the observation of the user in the context of interest. For both types of observation approaches there exist a variety of different observation techniques. Popular examples of surveys are *questionnaires* or *interviews*; in ethnography, *shadowing* is a frequently used technique. Typically, both observation approaches are combined to gather a holistic understanding.

In earlier days, when design and development were targeted at stationary desktop platforms and workstations, user studies were conducted in lab environments, and the aforementioned observation techniques were sufficient. In fact, researchers could easily video record users and observe their interactions, their successes and failures. Further, users could fill out questionnaires or answer interviews about their expectations and whether they are met. Lab studies and traditional observation techniques still are the best practise if fundamental research questions about, e.g., perception or cognition, should be studied on a controlled level.

[1] `http://www.google.com/about/company/philosophy/`, last visited January 5, 2015.

Figure 1.1: Traditional observation techniques have significant limitations when applied in field environments. For example, shadowing requires the permanent presence of a human observer, which can cause study participants and passersby to change their natural behaviour.

However, as foreseen by Marc Weiser in the early 90s, the era of ubiquitous computing has begun [Wei91]. Instead of using traditional desktop computers, which we used in mostly static contexts in designated offices, we now use smart phones, tablet computers, or other pervasive interfaces to interact, work, and communicate in our everyday life. Instead of a single device for all tasks, we now have several interconnected devices of various form factors—all mobile and often seamlessly integrated in our environment. These devices communicate and collaborate to give us an immersive, longitudinal experience and assist us in fulfilling tasks of all kinds.

Observing people using ubiquitous technologies and interaction concepts "in the wild" is a major challenge [AM00, Kru09]. One reason for this is that the interaction is more and more distributed across several devices. Given that all devices are interconnected, the focus of the interaction can easily shift from one device to another. Further, there often is no concrete start or ending point for an interaction. The users intentions have a more prolonged character, requiring an interactive systems to provide longitudinal support of varying degree and intensity. Eventually, interactions might reach the border between explicit and implicit interaction, making it context-dependent how a system is exactly operated.

In such highly dynamic, uncontrolled field environments, traditional observation methods, like shadowing (see Figure 1.1), come with significant limitations. Some techniques fail to get situated insights, which are crucial to fully understand the users and their situations. Other techniques are perceived as too obtrusive, likely causing study participants to change their natural behaviour. In addition, the scalability of many methods in the field is limited, limiting the efficiency with which users can be observed. Literature has referred to these challenges under the terms *situatedness*, *unobtrusiveness*, and *scalability* [FCC+07, MHK05].

Because many of the limitations of traditional techniques can be credited to the presence of a human observer, *unsupervised* observation techniques were invented. These techniques don't require a human observer and use other sources of information, e.g., sensors or user input, to provide researchers with insights. The most established and prominent unsupervised techniques are logging, the Experience Sampling Method (ESM) and diary studies [FCC+07, MHK05]. Logging uses quantitative data streams, such as the data provided by sensors, to observe the user. Self-reporting methods, like ESM, are used to gather subjective feedback by confronting the user with inquiries and record their response. In diary studies, it is left to the user to record relevant or interesting experiences, and communicate these to the observer in a later post-hoc interview.

Although the overall assessment of nowadays unsupervised observation techniques is promising [FCC+07, MHK05], they are still far from being optimal in every regard. Overall, unsupervised methods in their current state come with less situatedness than traditional supervised techniques. That means that unsupervised techniques have limitations in covering the full extend of situational details. However, because of the absence of a human observer, they are overall perceived as less obtrusive and perform better in terms of scalability. Altogether, literature agrees that unsupervised methods are well-suited for the observation of ubiquitous systems [AM00, MHK05, FCC+07], but there still is a lot of potential in research about better methodologies [AM00, Kru09, KG03, KCS+12].

1.1 Challenges

The literature identified three key challenges, which need to be addressed to eventually have better observation techniques for field studies in the era of ubiquitous computing: situatedness, obtrusiveness, and scalability [MHK05, FCC+07, KG03, KCS+12]. We give a detailed description of each challenge in the following.

The key outcome of a user observation are the gathered insights. To get the maximum out of made insights, it is important that an analyst, e.g., a researcher, is well **situated** and gets the best possible impressions of a user's situation and related experiences. These situated insights allow a researcher to eventually understand the users' underlying problems and find appropriate solutions. Unsupervised techniques have a limited ability to provide analysts with what was coined as situatedness, because in contrast to shadowing they don't allow to get and work with first-hand experiences. Instead, the provided observations come from the external perspective of, e.g., sensors or self-reports of users, making analysis and interpretation a challenge. Consequently, one of the challenges to be addressed in this thesis is to improve the *situatedness* of unsupervised observation techniques.

The observations of users in the field should be as **unobtrusive** as possible, to neither make the users themselves nor passersby changing their natural behaviour. That users change their behaviour under observation is known under the term "Hawthorne effect" [May33, RD39] and was widely discussed in literature, e.g., [Mac07]. It allows researchers to potentially draw invalid conclusions, limiting the overall validity of the observation and the study. In addition, obtrusive observation techniques, such as reoccurring self-report inquiries, can be perceived as annoying, demanding, and frustrating. Again, this is something which in the end could affect the overall insights and their validity. Consequently, to reduce the *obtrusiveness* of unsupervised observation techniques is another challenge that needs to be addressed in this thesis.

Another quality criteria for observation techniques is **scalability**. That means how efficient a researcher could observe how many users and gather relevant information. Unsupervised observation techniques have the advantage that no human observer needs to be present during the actual field observation. Overall, this already positively affects the scalability of these methods. Nevertheless, recorded data from unsupervised techniques still needs to be analysed and studied, which takes time and, thus, makes room for further scalability improvements. Consequently, the third challenge to be addressed in this thesis is *scalability*.

1.2 Approach

Like any other interactive system, ubiquitous systems are also typically based on the input-process-output paradigm. Wearable sensors or other input devices create an actual input for a system. A memory-supported processing unit uses this input to compute and trigger an output behaviour of the system. Dedicated output devices, such as displays, speakers, or vibrotactile interfaces, then represent and communicate this information to the user.

For many of today's ubiquitous and pervasive applications at least one element of the paradigm, i.e., input, process, output, is taken over by a smart phone. Smart phones can record user input through any of the available sensors or the touch screen. In addition, nowadays phones are equipped with very capable

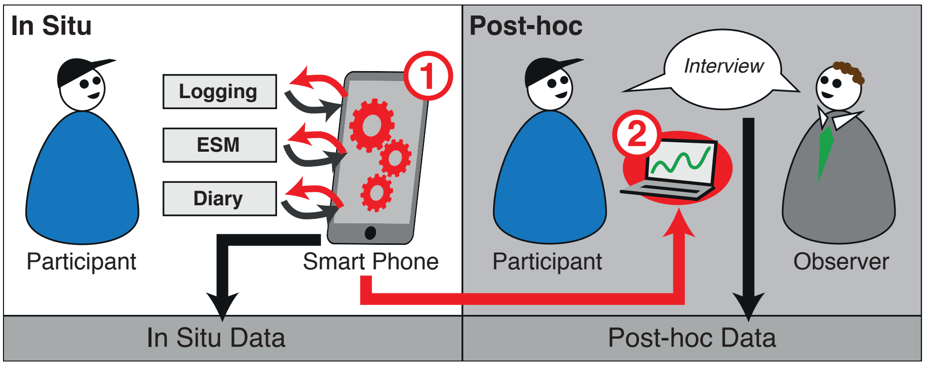


Figure 1.2: The approach followed in this thesis is two-fold. First, we make unsupervised in situ techniques smart by using a mobile phone's sensing and processing capabilities (Step 1). In addition, we do a hybrid observation by re-using recorded in situ data in post-hoc settings to elicit further insights (Step 2).

multi-core processing units and a huge amount of memory, which qualifies to host the actual logic of a ubiquitous application. Further, modern phones come with several multi-modal output capabilities, like speakers or a vibration actuator, allowing for a broad diversity of feedback.

With all these capabilities and its prominence in the fields of ubiquitous and mobile computing, many unsupervised observation techniques make use of these capabilities for unsupervised user observation. In fact, nowadays realisations of today's most common unsupervised observation techniques are highly dependent on mobile phones. Logging uses many of the mobile phone's sensors to collect data [HR00, HPSH00], the Experience Sampling Method (ESM) uses the phone's notification capabilities [FCC^{+}07], and diary studies benefit from the multi-modal recordings, such as videos, that a phone can make [CM05].

Our approach is two-fold and further stresses the capabilities of a mobile phone to address the identified challenges. The first part of our approach, labelled *smart observation*, is to use the input and processing capabilities of the mobile phones to further refine existing unsupervised observation concepts (see Figure 1.2, Step 1). Through sensing of user input many information is present, which we plan to process in different ways, e.g., through data mining. The ultimate goal is that this additional information can be used to optimise unsupervised techniques to become more scalable, less obtrusive, and more situated.

The second part of the approach, named *hybrid observation*, is to re-use collected in situ data in an intelligent way to further improve the overall user observation and quality of insights. The idea is that the recorded in situ information can support later reflection processes of a study participant (see Figure 1.2, Step 2).

Thereby, very situated details about the user experience could be revealed. Further, we expect positive implications on a technique's scalability and obtrusiveness.

We illustrate our approach with three unsupervised observation techniques: logging, ESM, and diaries. We decided for these techniques because they are very popular in the field of HCI and have been used for a broad variety of different use cases and scenarios. Further, the vast majority of other unsupervised observation techniques use one or more of these three techniques as a basis and conceptual foundation. Consequently, our insights will likely enhance and improve these and future unsupervised observation techniques as well.

All three, unsupversived observation techniques are investigated in real and representative study contexts. We illustrate how unsupervised observation techniques can be applied in the domain of navigation and orientation as well as in the field of longitudinal mood monitoring. Both are fundamentally well understood topics, which at the same time receive a significant and ongoing attention in acadamia and industry. This combination allows us to discuss and ground our study observations on established theories, but also address timely research topics and illustrate the applicability and relevance of the unsupervised observation techniques in these settings.

We show that each of the three unsupervised observation techniques benefits from both, the *smart observation* and the *hybrid observation*, or both at the same time. Consequently, we study each of the three techniques with each of the two sub-approaches. Overall, this allows us to get a detailed and holistic understanding of our approach on the individual techniques. In the end, this understanding allows us to give targeted recommendations, which observation technique and which sub-approach are most appropriate for certain scenarios and settings. Further, we can conclude which modified, unsupervised technique is best to overcome which of the three identified challenges, i.e., situatedness, unobtrusiveness, scalability.

1.3 Research Questions and Contributions

An initial literature analysis of the aforementioned most prominent unsupervised observation techniques with our two-fold approach in mind leads to five key research questions, which are presented in the following.

RQ1: To What Extent Environmental and Physiological Information Can Be Used to Enhance the Situatedness of Logging

Static environmental data, such as the locations of crossings and buildings, is known to be an essential source for spatio-temporal analysis of pedestrian behaviour [RS99, MG11]. Further, physiological data is frequently used in HCI to study aspects like stress or arousal in the field [SFR+13, Var07]. In existing research, both data types were separate subjects to supervised analysis and interpretation. It is unclear to what extent these information can be combined, integrated, and processed in an unsupervised approach and how this will change the situatedness of the logging technique.

In this thesis, we combined environmental and physiological data with other logging information and studied which advantages and beneficial insights this causes. We contribute that environmental information is a valuable means to structure and separate logging data. We further observed that physiological information is unsuited to increase situatedness, mostly because of its limited dynamics in outdoor field study settings. Altogether, we conclude that additional data can be beneficial for the overall situatedness of the logging observation technique.

RQ2: In What Way a Rule-Based Analysis of Logging Data Can Reveal Valuable Insights

Rule-based systems allow the analysis and interpretation of data in a useful way, which in contrast to many machine learning approaches can be followed and understood by a human being. The rule-based analysis and processing of data has shown its advantages in various applications, e.g., to layout magazines [KKO12] or to improve activity recognition systems [MBIC13]. Yet, it is unclear to what extent it can be applied to data, which was obtained from dynamic field settings, and if a rule-based analysis allows to draw conclusions on user behaviour and interaction level.

We did a rule-based analysis of logging data with the aim to draw conclusions about users' scanning interaction behaviour. Scanning describes an interaction technique, where the user explores an environment by pointing a mobile device in various directions. We present a set of rules, which we derived based on insights from earlier work, and which is able to automatically identify this behaviour in logging data. A detailed analysis shows that the made observations are, overall, similar to findings of previous studies. Consequently, we argue that rule-based analysis is a valuable means to increase situatedness and scalability of the logging technique.

RQ3: If Machine Learning Could be Used to Issue Self-Reporting Inquiries at Opportune Moments

Self-reporting techniques often actively ask the user for input, e.g., through pop-ups or notifications. It is known that these inquiries often appear in inopportune moments, causing stress to the study participants [MGK08]. As potential solution to this problem, machine learning technologies were proposed, which allow to automatically analyse contexts and adapt an application's behaviour accordingly. However, it is questionable if machine learning is capable to predict opportune moments to issue self-reporting inquiries in the field.

We conducted a large-scale study through the Google Play store, where we collected over 6500 of answered and ignored self-reporting notifications. In addition, we recorded contextual information, i.e. sensor data, in the situation where the notifications were issued. This data was used to create a machine learning classifier that predicts opportune moments with an accuracy of almost 78 %. A detailed investigation shows that notifications should be presented at the right time, i.e., towards later hours of a day, or if the device is already at hand. We envision that the applied model can lead to a substantial decrease of notifications appearing in inopportune moments, which therefore reduces the number of annoying and unpleasant interruptions at no costs. Consequently, this approach allows to decrease the obtrusiveness of self-reporting techniques.

RQ4: How Storytelling and In Situ Reflection on Diary Entries Affect a Post-Hoc Interview

Diary studies give study participants a means to freely capture their qualitative everyday experiences in situ. Typically, the captured material is reviewed and discussed in a post-hoc setting after the actual study. Diary studies come with two problems: participants forget to create a diary entry, and individual entries have lost their contextuality and recentness when it comes to a post-hoc interview. Authoring tools were proposed to be a potential solution to these problems, because they simplify to frequently record new material and typically increase overall expressiveness. In HCI, multimedia recording [CM05] and Storytelling [Bon11] were studied and discussed. However, it is unclear to what extent these approaches affect 'in the wild' diary studies and if these modifications will change related post-hoc interviews.

We developed the Storyteller, a mobile storytelling application which simplifies media capture and story creation. The application was deployed in a field study to investigate how the storytelling concept compares to a regular diary study approach. We found that participants liked the quick and easy mechanism to capture new diary entries, although not significantly more entries were created.

Further, participants liked to interlink diary elements with each other to create a holistic story. They agreed that they appreciate to tell the story with the help of the Storyteller and felt well supported. We conclude that diary studies with storytelling are a valuable approach to improve the overall situatedness of diary studies and their post-hoc interviews.

RQ5: How a Hybrid Observation Approach Performs in Longer-Term Non-Everyday Field Studies

Hybrid observation techniques use unobtrusive in situ observation techniques, like logging, that do not require an experimenter for the actual observation. The gathered data is then used by an experimenter in a post-hoc study to jog the participants memories, revive the experience and, consequently, obtain additional feedback [ELF93]. Hybrid observation techniques have mostly been applied in short-term, day-to-day activities. It is unclear how hybrid techniques perform in longer-term, non-everyday field studies. It is particularly questionable which of the in situ observation techniques is least obtrusive in a hybrid context, how the hybrid observation technique will affect the relation and communication between experimenters and participants, and which in situ insights foster the most relevant discussions in post-hoc studies.

We conclude that the hybrid observation approach enables field study participants to remember details, which they were not aware of any more. The presented in situ data supports many of the made participant statements, although they also enabled experimenters to identify weaknesses in explanations. We got mixed impressions regarding the obtrusiveness of various in situ observation techniques. Because of the recorded photographs, the SenseCam was assessed most obtrusive, followed by ESM, and logging. We observed that post-hoc studies tend to become more lively, if in situ material is present. This simplifies the creation of empathy and supports discussions. We found that the map was particularly used to refer to places, and visualised SenseCam images were used to further explain situated experiences.

1.4 Thesis Outline

The next chapter, Chapter 2, provides an extended introduction into field study techniques. Each of the following four chapters addresses a single of the selected unsupervised observation techniques, i.e., logging, ESM, diary, and investigates one to two research questions (see Figure 1.3). Chapter 3 is about logging and presents findings on Research Questions RQ1 and RQ2. Chapter 4 contains a study on ESM, a self-reporting technique with active inquiries, to provide an answer to RQ3. Chapter 5 focuses on the diary study technique, which is extended

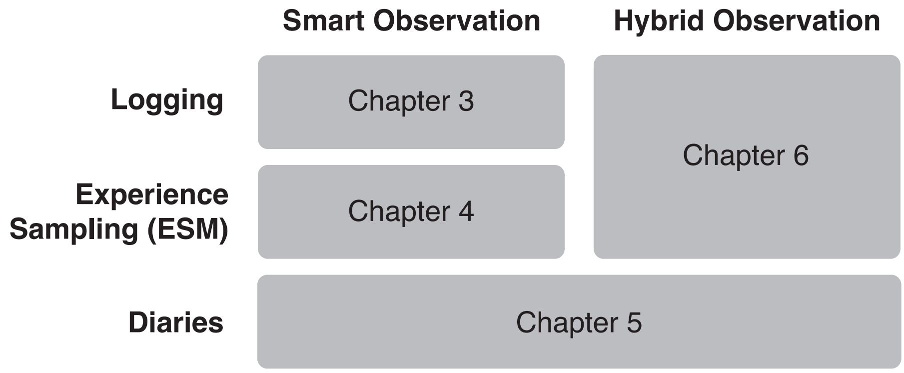


Figure 1.3: A visualisation of the thesis structure. We apply both of our two sub-approaches to each of the three most popular, unsupervised observation techniques.

with a Storytelling approach, and addresses RQ4. In Chapter 6, we propose a combination of in situ observation techniques, followed by a data-supported post-hoc interview, to clarify on RQ5. The last chapter of this thesis, Chapter 7, concludes the work and summarises the key findings.

Chapter 2 gives an overview of observation techniques in HCI field studies. Survey-based and ethnography-based study methodologies are presented, their advantages and disadvantages are discussed, and it is illustrated how they were applied in earlier days and how they are applied today. Further, the most prominent modifications to these techniques are introduced and common frameworks to realise these methods are highlighted. To allow a proper assessment of the thesis contributions, the following chapters take up and specifically extend these fundamental insights wherever necessary.

Chapter 3 focuses on the logging observation technique and reports about two conducted field studies, each investigating a separate method approach to improve logging. The first study was conducted with 11 participants and addresses RQ1 (see Section 3.2). Our presented findings indicate that environmental information is helpful to structure logging data, and physiological data has insufficient quality to be of use in the investigated pedestrian navigation scenario. The second study investigates RQ2 and shows that rule-based analysis of logging data allows an experimenter to get a detailed understanding of the user behaviour (see Section 3.3). We conclude this chapter with a summary on the made improvements to the logging observation technique, i.e., our modifications lead to an increased situatedness and scalability.

In Chapter 4 we focus on the Experience Sampling Method (ESM) as most prominent representative for self-reporting techniques that use active inquiries [FCC+07]. We present a large-scale study with 79 participants over a period of six months which addresses RQ3. Our results show that the machine learning-based prediction of opportune moments is possible with an accuracy of almost 78 %. Consequently, inquiries of self-reporting methods could appear in opportune moments, which will likely lead to less stress and effort. We summarise that this study method positively affects the obtrusiveness of self-reporting study techniques, like ESM.

In Chapter 5 we present a field study where we address RQ4 and investigate how storytelling and an in situ reflection on diary elements advance a traditional diary study approach. The idea is that users thereby reflect on diary entries at an early stage and combine their experiences to a holistic story. In a study with 10 participants we found that participants liked the quick and easy mechanism to capture new diary entries. Further, we found that participants appreciated the overall idea to combine entries to a coherent story. We summarise that the diary study has a lot of potential for modifications to further increase the situatedness of the technique.

Chapter 6 extends the idea of the Day Reconstruction Method and combines several in situ techniques with an assisted post-hoc interview. This concept is also known under the term 'hybrid observation'. A conducted field study addresses RQ5, i.e., how this hybrid approach performs in longer-term field studies. We present that data of many unsupervised in situ techniques can be a valuable resource in post-hoc interviews and that a hybrid study design positively affects the overall communication between experimenters and participants. We summarise that hybrid observation increases the situatedness and, together with appropriate post-hoc analysis tools, also can increase the scalability of this approach.

We conclude this thesis in Chapter 7, where we highlight the key contributions, illustrate the expected impact, point out important limitations, and give ideas for future work in field of unsupervised observation techniques.

1.5 Publications

Excerpts of this work have been published in international journals, conferences, or workshops [PPHB12c, PHB14b, POHB13, PPHB14]. These excerpts are clearly marked within the thesis.

Further, I published several papers on related topics, which also eventually contributed to the overall idea, objective, and outcome of this thesis [HPB10, HPP+11, HRB12, MP11, PPB09a, PPB10, PVK+10, PPH+11, PPHB11a, PPB11, PPHB12b, PPHB12a, PPB09b, PB10b, PB10a, PMH+10, PPHB11c, PMPRG11, PTHB11, PPHB11b, PHF+12, PCB+12, PHB14a].

During the work on this dissertation I co-supervised various Diploma, Master, and Bachelor theses. Of these, many form the basis for the presented research and the resulting publications. Most notably, I want to refer to [Koe10], [Str12], [Nie12], [Oeh12], and [Zer12].

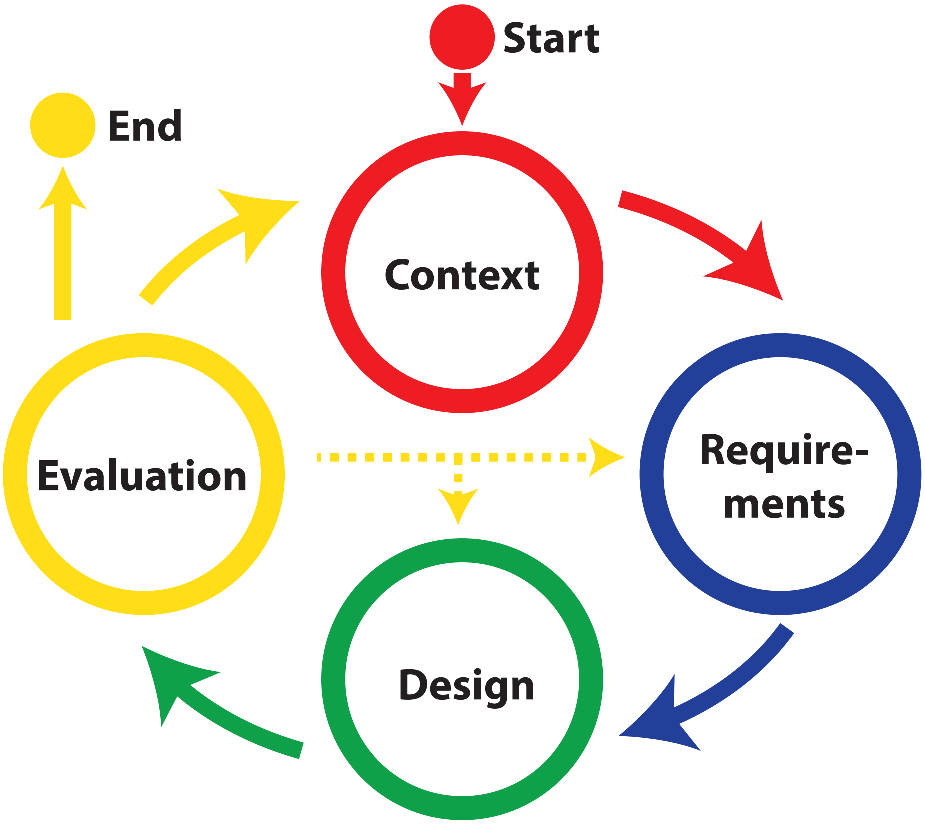


Figure 2.1: The Human-Centred Design (HCD) process is iterative and consists of four phases [ISO10]. User observation is typically done in the first, i.e., context analysis, and last phase, i.e., evaluation.

2 Observation Techniques in HCI Field Studies

In the field of Human Computer Interaction (HCI) it is common to involve the user in many if not all aspects of the design and development process. This is done to reduce misunderstandings and misinterpretations of the requirements and to have the outcome matching the users' actual needs as closely as possible. For commercial products, this will eventually result into high customer satisfaction, which is known to be a major factor for the overall success of a product.

The human-centred design (HCD) process for interactive systems (see Figure 2.1) is among the most frequently used processes for the design of HCI research prototypes [ISO10]. The HCD process is iterative and basically consists of four successive phases: understand and specify the context of use, specify the user and organisational requirements, produce design solutions, and evaluate designs against requirements. While the user is involved in all phases, only the first and last phase deal with the observation of the user, either to understand the user or to evaluate how a design solution performs in practise.

Literature shows several ways to differentiate between various, user-involving observation and study setups. Most prominently, it can be differentiated between lab studies and field studies. Lab studies are typically used to observe phenomena which don't necessarily need external influences or context information, e.g., usability studies. In contrast, field studies are used if context is important for the whole user experience, e.g., for the evaluation of a navigation device. In the last few years, researchers have reached consensus that field studies are inevitable for authentic and successful research in mobile HCI and ubiquitous computing [NOP+06, RCT+07, KCS+12, Kru09].

A huge variety of different observation techniques exists to do the actual user observation. Vermeeren et al. recently identified 96 techniques of which 64 can be applied in field studies [VLR+10]. Most of these techniques are based on fundamental observation techniques, which have been created in ethnography, social sciences or psychology, but were modified to address more specific HCI scenarios. Some are designed to be applied with fully functional products, while some are optimised for conceptual design ideas. Further, some techniques are optimised for long-term usage, some for observing single task experiences.

There is no single definition of what fundamental observation techniques are. However, there is a set of techniques, which is repeatedly referred to, and whose methods serve as a basis for many other user observation techniques. Among these popular techniques for mobile HCI field studies are: shadowing, interviews, questionnaires, diaries, the Experience Sampling Method (ESM) and logging [MHK05, FCC+07, HRKS05, KG03, KCS+12].

This chapter gives a brief introduction into each of these six observation techniques. We explain the basic rationale behind each technique and indicate advantages and disadvantages. For unsupervised techniques, such as diary studies, we further give insights into existing and prominent technique modifications, and present key frameworks and tools. We conclude this chapter in Section 2.7 with a comparative assessment of these observation methods.

2.1 Shadowing

Shadowing is also known under the term *observation* and means that one or more observers closely follow and observe a study participant in any situation for a certain period of time (see Figure 1.1). Thereby, the observer pays close attention to what the user does, says or experiences. Further, the observer will also try to understand the contextual information, e.g., which people are interacted with, what was the rationale behind a certain action. McDonald identified that there are three motivations to do shadowing: to experientially learn about the context of a subject, to record behaviour, or to understand roles and perspectives [McD05]. Because of the huge amount of information, typically notes are taken or the subject is being video recorded. If notes are taken it is recommended to use a framework to record the data as otherwise important information might be lost. Robson proposed to record *space, actors, activities, objects, acts, events, time, goals, and feelings* for every observation [Rob11].

Shadowing can be differentiated by the degree of participation of the observer. In most studies the observer is neither particularly involved nor particularly invisible. However, a variation in the degree of involvement can be necessary and influences the obtrusiveness and situatedness of the observation. An active observer who seamlessly integrates into the daily life of the observed subject is known as *participant observer*. Given the thereby created closeness and familiarity an observer might observe much more intimate details than a regular observer would do. The other extreme is known as *passive observer*, who does not participate in anything and tries to be as unobtrusive and invisible as possible. Both kinds of observation have their advantages and can provide unique insights. It depends on the study setting which one should be chosen.

The biggest advantage of shadowing is its situatedness. The observer gets many contextual information, which allow a broad understanding of the overall situation. Even unexpected information can be recorded, which is impossible for many other techniques, like structured interviews. However, all observations need to be interpreted by the observer which can lead to subjective or biased insights. Further, shadowing is quite obtrusive. If the user knows about the observation, which is typically the case, this might lead to a unintended behaviour change, which is known as the Hawthorne effect [Lan58]. As a consequence, this can results in invalid observations and wrong conclusions. Shadowing requires the physical presence of a human observer and is therefore hardly scalable and costly.

2.2 Interview

Interviews are another observation technique to gather qualitative data, which requires one or more interviewers and one or more interviewees. Typically, the interviewer confronts the interviewee with different facts, statements or questions and is then awaiting the verbal response to the earlier elicitation. There exist several types of interviews, which can be categorised in *structured*, *semi-structured*, and *unstructured* interviews [SRP07].

In structured interviews the procedure of the whole interview is fixed. That means that the interviewer asks certain facts in a specific order. In most cases structured interviews are close-ended, which means that the interviewee is only allowed to respond from a choice of answers. However, also open-ended answers are possible in structured interviews. Independent of any reply, the interviewer goes on to the next question. Thus, typically no dialogue evolves which is great if the whole interview should be conducted efficiently. The downside is that open responses can not be considered, leading to potential loss of valuable information.

Semi-structured interviews are less strict regarding the questions and responses than structured interviews. Thus, the interviewee is allowed to respond with open-ended answers and further go into the details. The interviewer could also dynamically pick up open-ended answers and elaborate on them further, asking for more information or in-depth background. In practise, semi-structured interviews often come with a combination of structured questions and more open questions. Thereby, interviewers make sure to get well-defined replies on their questions, but still leave some room for further discussion.

Compared to structured or semi-structured interviews, unstructured interviews come with no guiding questions or statements. In contrast, the whole interview is started with an overall lead topic which should be discussed. As a consequence it is very hard to assess the actual outcome or results of the interview in advance. Once the discussion becomes fuzzy and less focused, an interviewer might intervene and guide the discussion back to the original statement. In practise, unstructured interviews are often applied as part of focus groups, which is also referred to as group interview, where several people discuss the topic of relevance from various perspectives.

Depending on which interview category, i.e., structured, semi-structured, unstructured, is selected the situatedness varies. For structured interviews it is typically lower than for semi-structured or unstructured interviews, as the interviewee is not allowed to elaborate on any questions. Similar to shadowing, the gathered insights are subject to the interviewer's interpretation. This might lead to invalid or biased conclusions. Interviews are obtrusive and don't scale very well, because both, interviewers and interviewees, are completely engaged in this task.

2.3 Questionnaire

A questionnaire is typically a paper sheet that comes with a set of questions or statements, which should be filled by a participant with a pen. For each question or statement a variety of response formats can be used, e.g., free text entry, check boxes, rating scales. The static way how the data is collected allows an easy analysis and interpretation. The experimenter is only involved when the questionnaire is issued and returned.

There are custom questionnaires, which come with a set of individually defined questions, and standard questionnaires, which come with a predefined set of questions and have often been validated. Custom questionnaires allow a flexible and very individual combination of questions, statements and response formats. However, this flexibility frequently leads to misunderstanding when answering the questionnaire. Thus, it is recommended to do a pilot test and check whether there is a common understanding of the used questions and response options. Custom questionnaires are not validated. Thus, it remains unclear what actually can be derived from the collected responses, e.g., a question for ‘workload’ can be interpreted in various ways, leading to inconsistent results. The use of non-standard questionnaires makes it hard to relate own findings with those presented in the literature.

To overcome these limitations there are standard questionnaires which typically have been empirically validated. That means that there is significantly less risk that these questionnaires do not provide insights into a subjective measure, like ‘workload’. Further, standard questionnaires allow to compare own findings with literature. Two prominent examples for standard questionnaires are the NASA Task Load Index (TLX) [HS88] and the System Usability Scale (SUS) [Bro96], which both result into a single numeric value which allows a unified assessment and comparison.

Questionnaires have the advantage that they scale quite well. The experimenter just needs to set up the questionnaire, distribute it, and eventually collect and analyse the results. Nowadays, many questionnaires are distributed digitally, thereby reaching much more people than the non-digital counterpart. However, questionnaires are often not answered in the situations of interest. That means that the user is often reflecting on earlier experiences, which limits the situatedness of the methodology. To overcome this limitations methods like diary studies (see Section 2.4) or the Experience Sampling Method (ESM, see Section 2.5) were invented. Answering a questionnaire is normally quite obtrusive as a user has to read the questions, decide on the answer and write the response.

2.4 Diary Study

A diary study is longer-term study methodology where users are asked to autonomously note important aspects regarding given research questions or other aspects of interest. The traditional realisation of a diary uses pen and paper and, thus, was mostly limited to textual descriptions (see Figure 2.2). What should be recorded can be pre-defined or of open nature. Pre-defined diaries ask the participant to capture material or a statement regarding a certain aspect, e.g., which food was consumed for lunch. Open diaries do not have specific topics, and participants are allowed to record whatever they think is of importance. For example, a participant, who is testing a wearable bracelet, might summarise

Figure 2.2: Originally, diaries were kept with pen and paper. Today, various tools and applications support the creation of a diary.

his/her impressions regarding wearing comfort, battery issues, and what else is considered important.

Diary realisations are not proactive and, thus, do not actively remind a user to create a new entry. Instead the interest of researchers are communicated or defined as part of the study task and procedure description and it is left up to the participant to create a diary entry. For example, Hinze et al. asked their study participants to create a diary entry every time they use a mobile search engine [HCN10]. Sohn et al. asked the participants to create a new diary entry, if they feel a mobile information need at all [SLGH08].

Diary entries are recorded for two different purposes: to collect feedback or to reuse diary entries in later stages to elicit further insights. In feedback studies a user is asked to elaborate on certain aspects of the experience, e.g., a user can be asked to answer questions or report on perceived feelings. The outcome of a feedback diary study are the notes itself, which are made available to the experimenter. In contrast, in elicitation diary studies the recorded material is reused in post-hoc interviews to inspire additional comments of the participants.

Diary studies scale about as good as questionnaires. The content, where the user is asked to elaborate on, needs to be defined and the analysis of the generated content must be done. These are the only two tasks require the experimenter's attention. With its in-situ character diary studies have a reasonable situatedness.

They intent to capture the user impressions close to the relevant situation, and at least on a daily basis. However, it has been shown that people frequently forget to create a diary entry and that diary entries are of poor quality and depth [BWK07], which reduces the overall situatedness.

The creation of each diary entry requires participants to recall what happened earlier. This introduces some cognitive bias, as participants might prefer to remember negative aspects over positive impressions. In elicitation diary studies, diary entries are typically reflected on several days after they have been initially created. This makes it hard for the user to remember the context and motivation behind an entry and the potential interconnection to earlier or later diary entries. In fact, the entries lose their currentness and contextuality over time, which was definitely present during creation. From a researchers' perspective these problems eventually result into missing or distorted information, which are threats for the overall study outcomes and validity. Diary studies further have the downside that they put a continuous burden on the participants, as these have to autonomously take care to create new entries. Consequently, diary studies also have an obtrusive nature.

Diary studies have a very long history in psychology and have been used for a diversity of contexts (see Bolger et al. [BDR03] for a comprehensive overview). However, traditional feedback diary studies were proposed to the CHI community in 1993 [Rie93], and elicitation diaries in 2005 [CM05]. Thus, these methodologies have not been widely applied in HCI and there is still much methodological research going on. Some of the most relevant modifications to the original technique are presented in the following.

2.4.1 Existing Modifications of This Technique

Originally, diaries only contained textual descriptions of an event, although plain text is known to have a limited expressive power. Consequently, researchers quickly started to include other media into their diary studies. For example, a camera can be used to capture information about places, and the resulting photos can enrich a textual description in a diary [VKO10].

Carter and Mankoff further extended the traditional feedback diary method by allowing the participants to capture media and use this media in a post-hoc setting to jog their memories [CM05]. They conducted and analysed three different diary studies, whereby they discuss how each media type can support participants' recall process. They found that photos are very supportive to recognise the *who* and *where* of a situation. They further found that the timing and sequencing of events is very important for activity reconstruction. Based on their insights they designed a modified diary study pipeline, implemented in the 'Reporter' tool.

Brandt et al. [BWK07] addressed the problem that study participants are unwilling or unable to invest time in thorough, reflective entries when these are initially created. They proposed to record small snippets in-situ, which serve as prompts for participants when completing full diary entries at a convenient time. They found that their idea basically works and it seems that almost no accuracy is lost. However, participants preferred to use text as only media. The participants were reluctant to use audio, because that felt awkward to them. Further, they did not took photos as inspiration, because the camera quality was not sufficient for their purposes.

In a typical diary study the participants do not have any time restrictions when a diary entry should be created. In contrast, the Day Reconstruction Method (DRM) by Kahneman et al. requires the users to record their experiences exactly once a day [KKS+04]. This single report should reconstruct the previous day, split into a set of several arbitrary episodes. Then, the participants are asked to answer certain study-specific questions each about the identified episodes. Arguably, the experimenter will end up with very detailed insights into the participants' days, assessments of individual activities, and related study-relevant experiences. Kahneman et al. argue that this method has a lower respondent burden than comparable methods, such as diary studies, but results into a similar complete coverage of the day. In contrast to other self-reporting techniques, DRM has a lower susceptibility to retrospective reporting biases.

Another significant variation of traditional diary studies are cultural probes [GDP99, GBPW04, Rob08]. Cultural probes were invented to remotely observe users in their everyday situations and while doing their everyday activities. Each cultural probe is basically a set of self-reporting and self-observation tools, such as a camera or a notepad (see Figure 2.3). When initially handed, users are asked to use the materials to capture their everyday experiences. Thereby, users can either decide on their own, which situations are relevant, or follow potentially given advices and instructions, e.g., to take a photo of the home. All recorded materials will eventually be send or handed to the experimenters. Gaver et al. summarised that cultural probes provided them with a rich and varied set of different materials and insights, which supported them in understanding the subjects' local cultures.

2.4.2 Frameworks and Tools

The purpose and reasons to take a diary vary between different studies. Further, participants often have personal preferences how they would like to keep a diary. Both results into a huge diversity of different tools that are used for diary studies. In fact, it was reported that researchers and participants used traditional pen and paper, blogs and micro blogs, online document management and journalling tools, office tools such as Excel, and even professional content management systems

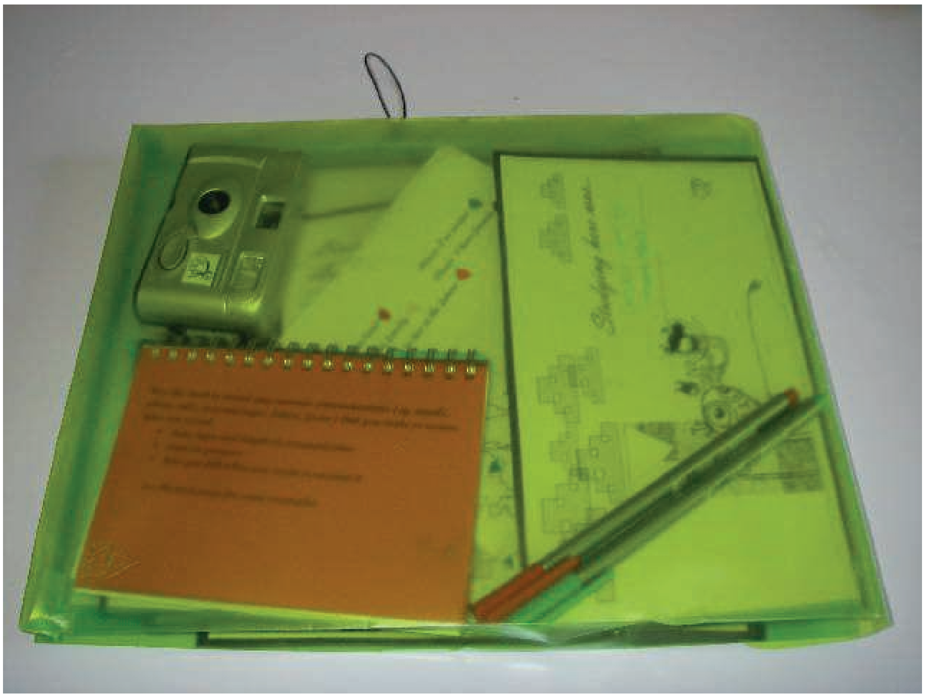


Figure 2.3: A cultural probe consists of various self-reporting and self-observation tools, such as cameras or paper diaries. Image was taken from [Rob08].

[MBP+12, SM13]. This means that there are often very specific needs for a diary implementation and consequently not many frameworks and tools exist.

The Affective Diary[1] by Lindström et al. is a diary implementation which is able to record various media, e.g., text messages, audio recording, photos (see Figure 2.4). In addition, it can also record physiological data, e.g., galvanic skin response, and other bodily experiences through, e.g., accelerometers. The idea of the Affective Diary is to use these sensor-recordings to integrate an emotion representation into the diary. Eventually, this emotion representation should be re-used for later reflection on the diary entries. The idea is to increase people's awareness of their own physical reactions and stress levels. The final concept to visualise emotion in the reflection phase consists of seven different body postures, which are coloured according to the user's arousal. The work started in 2006, when the idea was initially proposed and a concept for emotion representation was created [LSH+06]. In a long-term study with 4 users Ståhl et al. found that the affective diary enables users to reflect on the past [SHS+09]. Further, two participants reported that they discovered new patterns in their own body reactions, which made them learn something about themselves.

[1] `https://www.sics.se/projects/affective-diary`, last visited January 5, 2015.

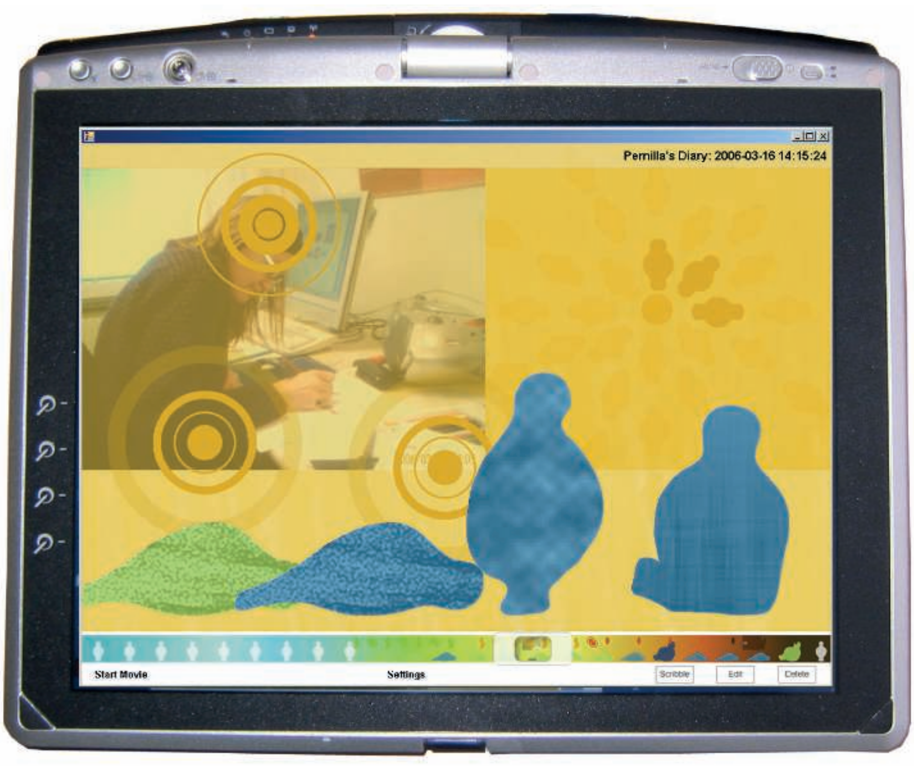


Figure 2.4: The Affective Diary visualises, e.g., the mood over the day, nearby Bluetooth devices, and photographs. Image was taken from [SHS+09].

In the last few years, more and more researchers picked up the idea of a distributed 'cloud diary'. The idea behind such a diary is that many users already use a variety of different devices and services. For example, people post their daily experiences on their favourite social network and wear an activity tracker, which uploads information about taken steps and walked distance. Combined with other non-personal sources, such as weather information, detailed information about a person's day can be reconstructed. However, at the time of writing, the idea of a 'cloud diary' remains in a conceptual stage, e.g., [GBL06, KME+08, FMZ11], with a lot of research going on in this field and the closely related lifelogging area, e.g., [SFA+07, HHS09, LDF11]. Some of today's influential phone manufacturers already took up the idea and started the implementation of cloud-supported diaries, e.g., Samsung's LifeDiary [YJJA07].

2.5 Experience Sampling Method

The Experience Sampling Method (ESM) was proposed by Csikszentmihalyi and Larson [LC83, CL87] and is similar to a diary study. A study participant is

Figure 2.5: Early realisations of ESM used a stopwatch to remind users to fill an inquiry form. Today, ESM is often realised on mobile phones, which can remind users and capture multimedia feedback.

also asked to create diary entries in situ. However, in contrast to traditional diary studies the user is actively reminded to create an entry. It is argued that thereby temporal things, like emotions or feelings, can better be recorded than with other observation techniques. In earlier studies a stopwatch or alarm served as reminder and diary entries were created on paper (see Figure 2.5). Nowadays, mobile phones can do both: remind and record feedback.

The reminders are meant to make sure that participants don't forget to fill out a diary entry, which would result into a high situatedness. The downside of active reminders is that they might be perceived as annoying or interrupting, increasing the method's obtrusiveness. Nowadays, the created entries can be recorded in a digital format, which allows a simple post-processing and makes the methodology quite scalable. Overall, the ESM can be adapted to various study settings quite well. Further, the type of recorded information can be changed from study to study or even from question to question. This allows an experimenter to capture, e.g., photos, text, or videos, and use these for follow-up elicitation studies.

In 2003 the method was introduced to the HCI community by Consolvo et al. [CW03]. Initially they proposed that this method is particularly useful for the evaluation of ubiquitous and pervasive applications, but shortly after the technique was also applied and used in mainstream HCI [IRK$^+$03, CSM$^+$05].

Until today it has been applied widely and in various contexts, e.g., to investigate personal fitness tools [CMT+08] or to understand daily routine activities [KHL+08]. However, it was also tried to overcome some methodological limitations of ESM, particularly to reduce obtrusiveness. Further, frameworks were developed to make it easier for other researchers to apply this methodology in their own projects.

2.5.1 Existing Modifications of This Technique

The original ESM was designed and implemented with a piece of paper and a stopwatch, which served as reminder. Early fully digital implementation of ESM stuck to this limited sampling, but already incorporated different kind of textual feedback, e.g., full text answers, and nominal/ordinal scales [BB01]. With the release of smarter phones, research that applied ESM already sampled multimedia contents. Similar to multimedia diaries, the additional content enables users to reach a new level of expressiveness, and researchers are able to gather much more diverse insights [FCC+07, GDS+10].

One of the early modifications that was proposed was context-aware ESM [IRK+03]. Primarily the intention was to sample experiences in particularly interesting situations, e.g., ask the user where s/he is going while travelling. However, this technique can also be used to not ask questions in disruptive situations, e.g., sampling the experience while sleeping at night. A big disadvantage of this proposal was that the context needs to be defined in advance, which makes it hard for the methodology to adapt to very individual habits. Consequently, it was proposed to adapt the underlying models dynamically over time and individually for each user. Kapoor and Horvitz investigated four different sampling strategies which can be used to train such models in a stationary setting [KH08]. In a study they found that sophisticated sampling strategies are indeed capable to minimise interruptions and maximise the classification accuracy of underlying models, which predict the cost of interruption.

2.5.2 Frameworks and Tools

One of the most popular ESM frameworks is MyExperience[2] by Froehlich et al. [FCC+07]. Originally this framework was developed for Windows Mobile operating systems (see Figure 2.6). Although ESM is the focus of this framework it also is able to record various other information. It can observe and record device usage, i.e., the used applications, the user context, i.e., whether there are any appointments in the calendar or not, and raw sensor information, i.e., the user location via GPS. One of the big advantages of the framework is that it can be

[2] `http://myexperience.sourceforge.net`, last visited January 5, 2015.

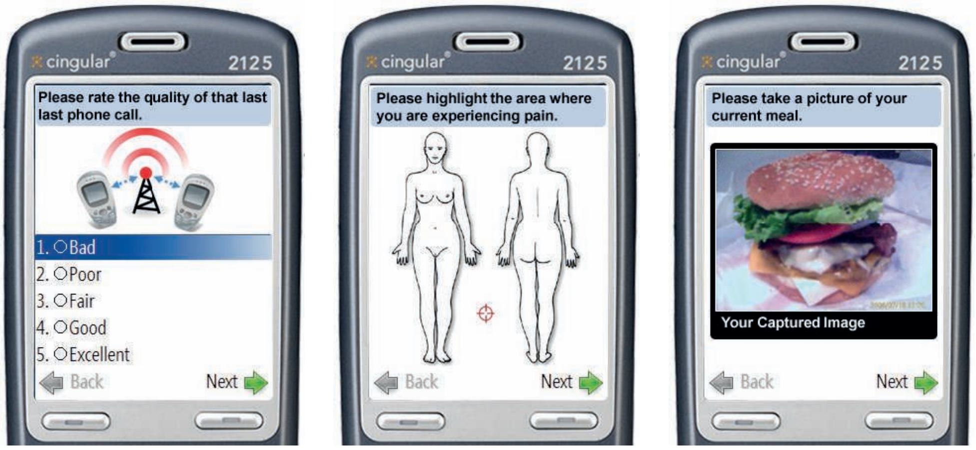


Figure 2.6: The MyExperience framework [FCC+07] is able to capture various information, e.g., answers to questions, photographs, or voice snippets. Images were taken from the project website (`http://myexperience.sourceforge.net`)

configured via XML, allowing an easy integration into any apps. The benefits of MyExperience were demonstrated in various settings, e.g., [CMT+08, KHL+08]. The last change to the framework was made in 2010 and it was not ported to newer platforms, like iOS or Android.

The MAESTRO framework by Meschtscherjakov et al. investigated context-aware ESM in mobile, non-stationary settings [MRT10]. They conclude that recorded long-term user behaviour and usage patterns can be used to shape personalised ESM questions and show them in suitable contexts. In their paper, they present the MAESTRO framework itself and illustrate the concept in a pilot study. Although it was promised that MAESTRO will be ported and released on different platforms, this did not happen until the time of writing.

The CAES framework[3] was started with the intention to implement a context-aware ESM methodology. It was originally designed and developed for Microsoft PocketPCs. However, when the MyExperience framework [FCC+07] gained importance, the developers decided stop their development, merge core parts with MyExperience, and continue contributing to this framework instead. Consequently, the CAES framework itself is outdated and no longer maintained.

2.6 Logging

In logging, the user observation is completely passed to a variety of different data sources, such as sensors or the user interface. Thus, logging is typically used

[3] `http://web.mit.edu/caesproject/`, last visited January 5, 2015.

to record quantitative information instead of qualitative insights, such as the user's exact location based on GPS. Earlier, this information was used to evaluate the usability of desktop software components [HCKN96, HR00]. Further, it was applied to track and understand visitors of websites [HHWL01]. In recent years, logging gained momentum under the umbrella terms 'big data' and 'crowdsourcing'. These terms describe the analysis of huge amounts of user-generated, quantitative data, e.g., search queries in search engines, which became possible because of the continuously improving data storage and processing capabilities of modern server infrastructures. In the field of HCI, the analysis of lots of users and a huge amount of data is also known under the term 'research in the large' [MMB+10].

In the last years there has been minor efforts to investigate and understand logging as observation method. Jensen and Larsen report from a three-month longitudinal field study which they conducted with a single participant [JL07]. They wanted to explore to what extent logging data is valuable at all and found that the recorded quantitative data allowed them to learn a lot about the study participant. Instead of fundamental research, the community has studied the applicability and versatility of logging in the wild. In fact, 'research in the large' was a dominant topic for research communities around HCI and ubiquitous computing in the last few years. Here, logging has been used to conduct long-term user studies through mobile application stores, such as Apple's App Store or Google's Play Store [MMB+10, SSR+11, HP13, RMCB13]. With this approach researchers were able to reach hundreds of thousands of users [HRB11], which used their study apparatus for long periods of several months [CRB+10], and thereby created insights of high external validity.

This enormous momentum spawned a lot of discussion about logging as observation technique. In fact, there are several things to consider when applying logging to conduct a user study, either in the market or with fewer participant in the field. A user needs to be aware of the logging and has to agree to the analysis of the data, i.e., sign an informed consent [MMB+10, KVVVM10]. For outdoor or long lasting studies, a user might not want to be equipped with several additional hardware sensors [CMT+08]. Further, combinations of device-equipped studies should rely on the same mobile device to avoid confusion [CMH07]. Coming from the experimenters point of view, the observation system should support qualitative data, quantitative data, and context information. Ideally the system supports the full experimental life-cycle. Tools for the identification and extraction of salient features are required, as well as a synchronization of different data streams [CBG+06].

Other research on logging deals with the capturing and replay of data to, e.g., support people with dementia. Further, event detection can be applied to logging data to automatically determine specific reoccurring events. Activity recognition

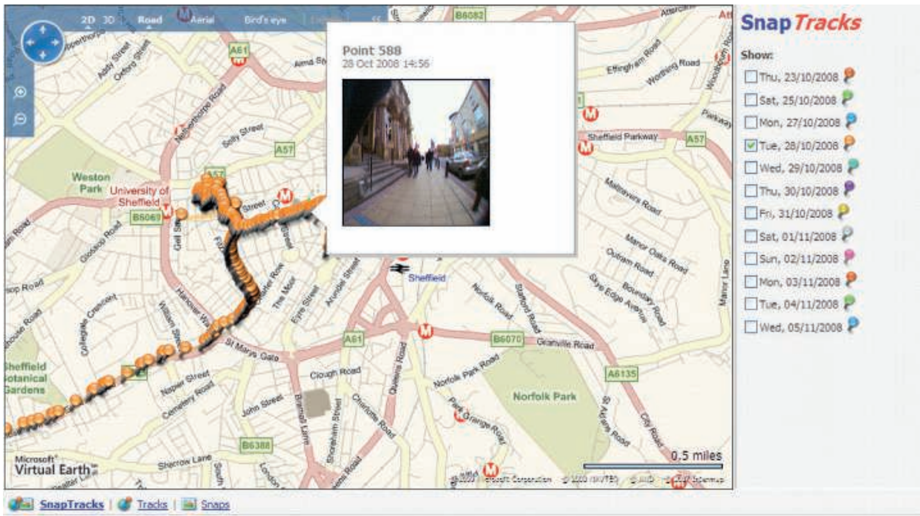


Figure 2.7: Kalnikaite et al. found that a combined visualisation of logged position information with photographs from a wearable camera leads to more recalled events. Image taken from [KSWK10].

is also working with activity recognition and tries to infer a users activity from the recorded sensor data.

2.6.1 Existing Modifications of This Technique

Logging as observation techniques was essentially modified in two ways in the last few years. First, it was studied how automatic data processing can improve the versatility and situatedness of logging. Therefore, various algorithms were applied to raw logging data to derive higher-level information about an observed subject. The used algorithms have their origin in the field of machine learning and are typically used to recognise and classify a human's activities, such as standing, walking, cycling. This modification lead to two frameworks, i.e., Context Toolkit and Context Phone, which are presented in Section 2.6.2.

Second, it was studied how logging can be combined with other observation techniques. Most notably, it was studied how logging can be combined with wearable cameras, like Microsoft SenseCam or Autographer. Kalnikaite et al. recorded users' location tracks, while wearing a camera, and compared different data visualisations [KSWK10]. They found that a combination of both visualisations (see Figure 2.7) leads to more recalled events. However, a visualisation of solely images lead to more recalled details. Further, this visualisation got a higher emotion rating, meaning that images are more emotionally evocative.

Sas et al. combine logged, physiological information on arousal, i.e., GSR (Galvanic Skin Response), with images from a wearable camera [SFR+13]. They found that arousal, which is probably caused by emotions and not by physical activity, improves the quality of memory recall. Photos taken at high arousal lead to more recalls of emotion, places, and episodic memories. Sas et al. conclude that this concept might be an effective and valuable approach to reduce the huge amount of images, which are produced by a wearable camera over a day.

The Footprint Tracker by Gouveia et al. is a tool that combines and visualises lifelogs with SenseCam images [GK13]. Thereby, the lifelogs consist of location information, and communication information, which is received from a mobile phone. They studied how visual, location, temporal, and social cues support the recall of recent events and associated emotions. They used an eye-tracking device to quantify their observations and found that visual cues got the most visual attention by participants. Further, similar to the work from Kalnikaite et al. [KSWK10], visual cues were rated as most helpful in recalling events. Nevertheless, other cues were also of importance, e.g., to resolve perceived conflicting information between different cues, or to maintain an awareness of the temporal order of events.

2.6.2 Frameworks and Tools

Funf[4] is an open source sensing and data processing framework for Android. It records values from available sensors and other data sources, and is able to do some on-device processing. Recorded data can be kept on the device or send to a server as single file or a coherent database. Funf comes with scripts, which allow users of the framework to do a basic analysis and visualisation of the recorded data. At the time of writing, funf is actively maintained and updated.

The Context Toolkit[5] is a framework which extends the idea of traditional logging through context detection [SDA99]. The framework receives sensor information from different, distributed sources, e.g., mobile phones, environmental sensors, and infers higher-level context information. All the information is stored and accessible to researchers, which allows a thorough understanding of the users and their contexts. This framework has been applied and used in numerous publications, e.g., [DAS01]. The Context Toolkit is available upon request.

Similar to Context Toolkit is Context Phone[6] [ROPT05]. This framework also aims to observe the user and infer context information, but focuses on highly integrated Nokia Series 60 mobile phones instead of a distributed environment. Context Phone is available for download on the project website. It has never

[4] `http://www.funf.org/`, last visited January 5, 2015.
[5] `http://www.cs.cmu.edu/~anind/context.html`, last visited January 5, 2015.
[6] `http://www.cs.helsinki.fi/group/context/`, last visited January 5, 2015.

been ported or adapted to newer mobile phones, and consequently is considered outdated.

Logging is also applied in commercial contexts to understand how users use and interact with an arbitrary application. Popular frameworks used in these settings are Flurry Analytics[7], Admob Metrics[8], and Google Analytics[9].

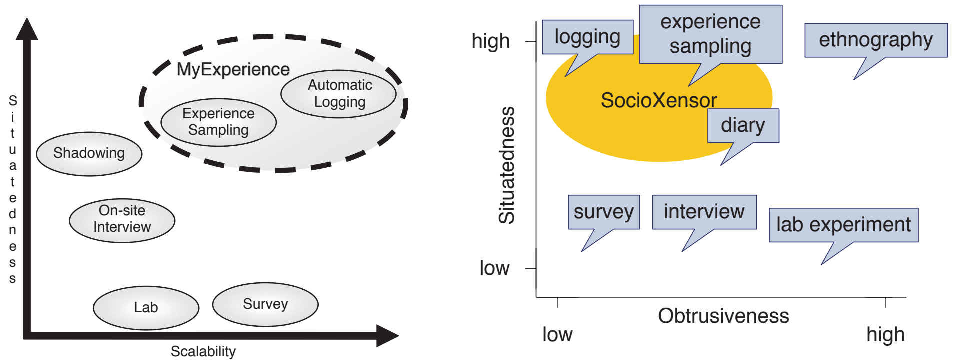
The MSP (Mobile Sensing Platform) is a wearable device, which comes with several integrated sensors [CBC+08]. These sensors allow to track and observe various parameters of the users. Further, the device itself is able to apply machine learning and do activity recognition. The objective of this platform is to enable and support researchers in the domain of ubiquitous computing to create context-aware computing systems and user experiences. MSP was applied in various different settings, e.g., [CMT+08], but is not yet available commercially.

2.7 Summary

In this section, we provided an overview of the most frequently used field observation techniques in HCI: shadowing, interview, questionnaire, diary study, ESM, and logging. For each of these techniques we explained the key design rationales, and pointed out advantages and disadvantages in common applications. We put emphasis on unsupervised techniques, i.e., diary study, ESM, logging, where we additionally presented existing technique modifications as well as frameworks and tools.

Each of the presented methodologies has it's right to exist and there is no single methodology for all user observations. However, there is ongoing work to assess the strengths and weaknesses of observation techniques, which simplifies the method selection for future studies. Today, observation methods are assessed by three key criteria: *situatedness*, *obtrusiveness*, and *scalability* [FCC+07, MHK05]. Made observations should reveal as much insights as possible. An observation technique with high *situatedness* covers many details, resulting in a better understanding and analysis of the study insights. Further, the observation should also be as *unobtrusive* as possible, because otherwise the observation itself can become an unwanted influence, which probably makes users change their natural behaviour. In addition, the observation should be as *scalable* as possible, meaning that many participants can be observed without spending too much resources, like time or money.

[7] `http://www.flurry.com/`, last visited January 5, 2015.
[8] `http://metrics.admob.com/`, last visited January 5, 2015.
[9] `http://www.google.com/analytics/`, last visited January 5, 2015.



(a) Assessment of scalability and situatedness by Froehlich et al. [FCC+07]

(b) Assessment of obtrusiveness and situatedness by Mulder et al. [MHK05]

Figure 2.8: Literature identified three key criteria which can be used to assess the overall performance of an observation technique: situatedness, scalability, and unobtrusiveness.

Froehlich et al. assessed the *situatedness* and *scalability* of observation techniques [FCC+07]. They observed that lab studies perform worst regarding situatedness, shadowing has the worst scalability, and logging performs best in both of the two aspects (see Figure 2.8(a)). Mulder et al. provide an estimation of the methods' *situatedness* and *obtrusiveness*. They conclude that ethnographic studies have the highest obtrusiveness, but provide situated insights. Lab experiments provide researchers with least situated insights. Again, logging performs best and provides situated insights without being obtrusive (see Figure 2.8(b)).

3 Improving the Information Gain of Logging

Logging describes an observation technique, where a device, e.g., a computer or a mobile phone, is recording quantitative information. Depending on what is actually of interest for a researcher, this quantitative information can be everything, e.g., the user's location or which button was pressed in a graphical user interface. Logging has its origin in software usability, where it was used to identify and understand usability issues with software [HCKN96, HR00]. With the advent of the Internet it is also used to understand how users visited and interacted with websites [HHWL01]. In the last few years, researchers started to apply logging in mobile or ubiquitous settings [FCC+07].

For mobile and ubiquitous settings the logging observation technique comes with two major limitations [FCC+07, MHK05]. The first problem is the limited **situatedness** of the observations. Traditional shadowing allows the experimenter to observe and understand almost anything a user is doing. With logging, many subtle aspects are unable to be sensed and logged. Thus, important details about the context and the situation are lost. Further, although there might be very detailed logging data available, this data is likely unable to represent a holistic situation, because each aspect of the data is analysed on an individual basis and possible connections between the data remain unnoticed.

The second limitation is a non-optimal **scalability**. While the observation and data recording itself typically scales well, the post-hoc analysis of data requires a lot of effort and processing time. Some information might need to be normalised before analysis, and sanity checks on the data need to be performed to identify and remove invalid or erroneous data samples. The actual analysis and used methods is highly depending on the observed scenario and application, often requiring a researcher to design and implement custom tools and procedures. Consequently, the data analysis takes time and limits overall scalability of the method.

In this chapter, we investigate how both, the situatedness and scalability of logging, can be improved through the sensing and processing capabilities of a mobile phone. In detail, we study how *environmental and physiological information* can be used to extend logging and gather a more holistic user understanding (Research Question RQ1). Further, we investigate if and how *rule-based analysis* of logging data can reveal valuable insights for an experimenter, particularly in the light of evaluations of user interface prototypes (RQ2).

In the beginning of this chapter, we introduce the PocketNavigator, a pedestrian navigation aid, and the Virtual Observer, a logging observation framework (see Section 3.1). The PocketNavigator serves as timely and relevant apparatus for the logging observation technique, the Virtual Observer framework is an outlet for the developed observation techniques within this thesis and uses logging as a key component. We then report on two conducted field studies, of which the first study shows that environmental and physiological information can improve the understanding of users (see Section 3.2). In the second study, we show that rule-based analysis of logging allows to create taxonomies for certain interactions, which simplifies the evaluation of user interfaces through logging (see Section 3.3). We conclude this chapter in Section 3.4 with a summary of our findings.

3.1 Background

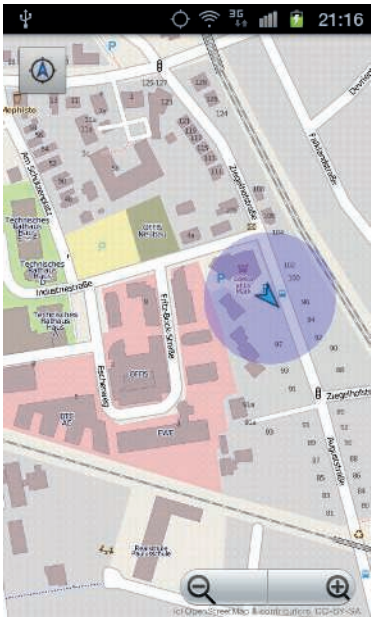
In the following, we introduce the PocketNavigator and the Virtual Observer logging framework. The PocketNavigator (see Section 3.1.1) is a pedestrian navigation system. Pedestrian navigation and orientation have been thoroughly researched in the past years, but yet are both still timely research domains. This combination allows us to thoroughly understand our logging observation approach, and at the same time show its applicability in a relevant domain. The Virtual Observer is an unsupervised observation framework, which was developed as part of this thesis and makes an essential technical contribution (see Section 3.1.2). Both are used as apparatus for several of our studies.

3.1.1 PocketNavigator: A Pedestrian Navigation Aid

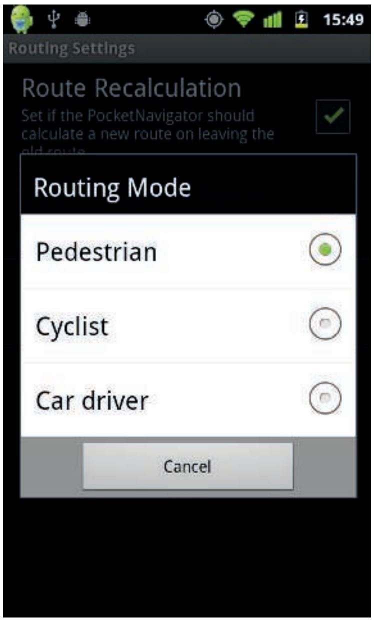
The PocketNavigator[1] (see Figure 3.1) is a pedestrian navigation system for Android smart phones. This application consists of a map layer, which shows prerendered map tiles from OpenStreetMap[2]. On top of this layer the user's location and heading is shown as a blue-coloured arrow. The position inaccuracy is indicated as a circle around the user location (see Figure 3.1(a)). The user is able to interact with the map through common gestures, e.g., pinch-to-zoom. Thereby, an arbitrary map excerpt can be displayed to the user. A button allows the user to centre the map on the user's current location. Further, the map can be set-up to rotate with the user's heading.

The application allows to set any location as destination for a navigation task by either double tapping or long-clicking the location on the map. Then, a web-based routing service is asked to provide a route between the current location and the destination. Various criteria can be communicated to the routing service, allowing suitable routed to be calculated for pedestrians, cyclists, and cars (see Figure 3.1(b)). Once a route is calculated it is visualised on the map as a grey line. The next waypoint is highlighted by a grey circle (see Figure 3.1(c)). Overall, the functionality is very similar to Google Maps and Google Navigation.

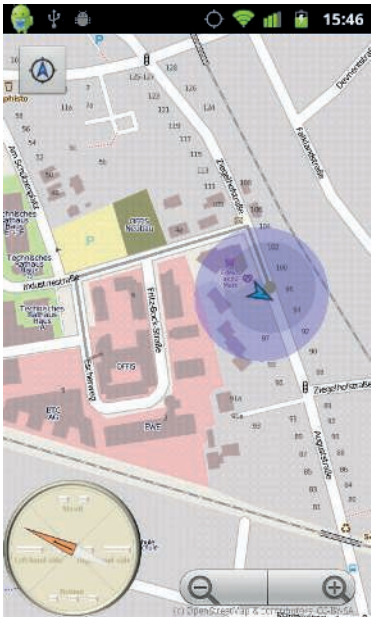
[1] `http://www.pocketnavigator.org/`, last visited January 5, 2015.
[2] `http://www.openstreetmap.org/`, last visited January 5, 2015.



(a) A map with the user location.



(b) Support for different transportation modes.



(c) Route is drawn on top of the map.

(d) Tactile feedback allows to get navigation instructions even if the device is in the pocket.

Figure 3.1: The PocketNavigator application provides similar functionality to existing navigation solutions, such as Google Maps. The user location is shown, and a route to-be-followed is drawn on top of the map representation.

The PocketNavigator was developed as part of a large, publicly funded EU project called HaptiMap[3]. The scope of this project was to investigate how map data can be made accessible through multimodal user interfaces. The PocketNavigator served as basis for vibrotactile interaction and therefore is able to convey the navigation instructions, i.e., the direction to the next waypoint on a calculated route, through vibration signals. If the way point is ahead, two vibration pulses of equal duration are presented. If the way point is on the left of the user the first pulse's duration is increased depending on how far left the point is. Accordingly, the right pulse's duration is increased if the way point is on the right. If the way point is approximately behind the user, three short vibration pulses of equal duration are presented. A visual representation of these tactile cues was available to the users (see Figure 3.1(d)).

The application supports two different operation modes. In scanning mode the device's compass is used to determine the direction to the next way point and adjust the vibration feedback accordingly. This enables the user to have a high precision about the direction of the next waypoint. Scanning mode is automatically enabled if the roll and pitch angles of the mobile phone, which can be derived from the accelerometer, are between $-16°$ and $16°$, i.e., the device is held in hand and almost parallel to the ground. The second mode, pocket mode, is enabled if the device is not in scanning mode. Then the GPS heading is used to determine the direction to the next way point. In contrast to scanning mode, the pocket mode comes with a lower precision, but allows to receive navigation instructions if the device is kept in a pocket.

The PocketNavigator is available in the Google Play store. There it can be downloaded and tested for free. It has been installed over 20 000 times and achieved an average rating of 4 out of 5 stars. The application served as apparatus for various studies. Related papers should be considered for in-depth information about the prototype, e.g., [PPB10, PPH⁺11, PPHB11a].

3.1.2 Virtual Observer Logging Framework

The Virtual Observer[4] is a logging framework for the Android platform. A key component of the framework is the Android logging library, which can be integrated into any Android projects. The integration allows to use the provided logging functionality and, thus, log device or user-related data. The library has been developed in an iterative way and has been deployed and tested in various settings and projects, partly within the research conducted in this thesis and partly in other research projects. The library and all related tools that are

[3] `http://www.haptimap.org/`, last visited January 5, 2015.
[4] `http://virtualobserver.benjaminpoppinga.de/`, last visited January 5, 2015.

presented in the following are open source and released under the Lesser GNU Public License (LGPL).

From a technical perspective, the logging library consists of three core elements, i.e., a set of ContextRetriever classes, a ContextSenser class, a ContextLogger class (see Figure 3.2). Each ContextRetriever is responsible for one type of sensor, e.g., accelerometer, compass, and comes with sensor-specific configurations and calls. The framework provides a ContextRetriever for all common sensors that can be found on Android devices. Further, also ContextRetrievers for other, non-sensor data sources are provided, which allow to record, e.g., user interactions. Custom ContextRetrievers can be created, e.g., to integrate external Bluetooth sensors.

These ContextRetrievers are linked to the ContextSenser, which manages the individual ContextRetrievers and provides an application developer with a central point of control. For example, the whole logging can be started, paused, resumed, and stopped. Further, individual ContextRetrievers can be added and removed at runtime, which allows a great adaptability and flexibility to different use cases and applications. This loose coupling further allows the logging to degrade gracefully and save battery life, if needed. A logic, which is integrated in the ContextSenser class, further allows to handle similar readings of different sensors. This allows a combination of sensor readings by type, e.g., to calculate the average on two speed sensors to increase reliability and precision.

The ContextSenser supports two sensing modes. In the first sensing-mode, which we also refer to as pull mode, the recording is triggered by time. Thus, logging information in fixed time intervals, e.g., every second, is possible. In the second sensing-mode, the push mode, the sensor itself triggers a recording. If a sensor is read with a frequency of 100 Hz, ideally 100 values per second will be processed and recorded. However, the push mode is meant to allow the fastest logging of sensor values, no timing guarantees are given.

The ContextLogger is responsible to log the gathered sensor readings from the ContextSenser to a file on the device or to the Internet. It is initialised with a specific sink, i.e., FileSink and HttpGetSink, and a ContextPlayer instance. The ContextLogger caches the data on the device and takes care of re-sending it if an error occurs or the sink is currently unavailable. The sink takes care to deliver the actual data. Part of the framework is a FileSink, which writes the data to a mobile device's external memory card, if available. The HttpGetSink sends the data to an arbitrary web server or web service. Other sinks can be developed if needed, e.g., to allow a data storage in cloud services or in large-scale database management systems.

A server-side script is provided, which is able to handle data received from the HttpGetSink. We decided for HTTP GET as appropriate methodology to send the data, because other protocols are frequently blocked or delayed by mobile

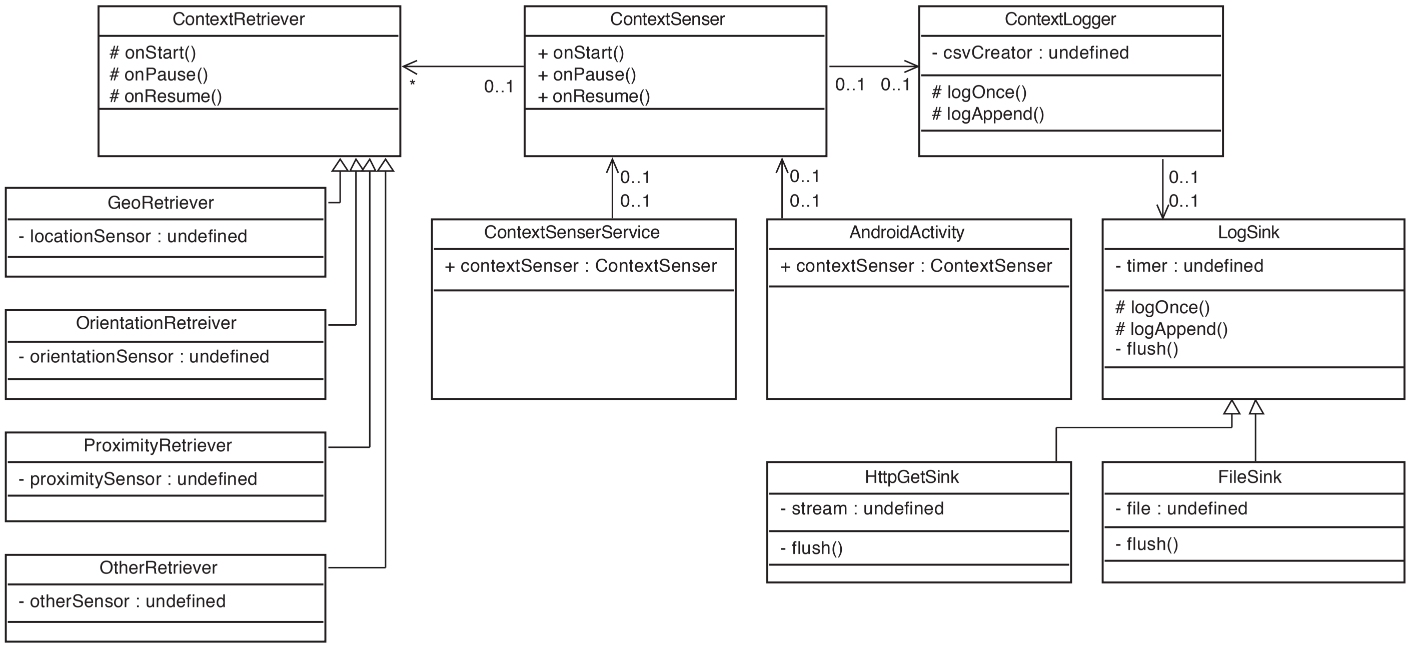


Figure 3.2: The Virtual Observer logging framework consists of three core classes: ContextRetrievers retrieve context information, ContextSenser contains the logic, ContextLogger saves the information. The ContextSenser can be used application-wide within a service or be integrated into an arbitrary Android activity.

Value	**Description**
Timestamp	Time of the recorded logging data in UNIX default representation.
Logging Version	The used release of the logging. This allows to refine the presentation of a data package if necessary.
Activity State	The state of the currently present activity, e.g., started, paused.
Location Provider	The current location provider, e.g., GPS, Network.
Latitude	The current latitude of the user.
Longitude	The current longitude of the user.
Position Accuracy	The position accuracy of the provided location.
Speed	Information about if and how fast a user is moving.
Heading	The direction in which the user is moving.
Compass Heading	The direction in which the device is oriented.
Roll	The tilt angle of the device.
Pitch	The pitch angle of the device.
Proximity	Indicates whether the device's screen is covered or not.
Light Level	The light level in SI lux units.
Extension	The extension field, which can contain any additional, application-specific information.

Table 3.1: An overview of all data values which can be logged by the Virtual Observer observation framework. The extension field allows to log various application-specific information.

service operators. The script uses PHP and allows to handle and process input from several users at the same time. For this purpose, a unique user id is generated on the mobile phone and send along the actual data package. *Salted* data and a CRC checksum are used to ensure that the server script only accepts data, which comes from a Virtual Observer instance.

The Virtual Observer file format is proprietary, but human-readable and well-defined. It addresses the need to record common data, which is relevant for most mobile applications, but nevertheless allows a certain flexibility and extensibility. Because storage on mobile devices is limited and sending data via wireless networks is time consuming, the data overhead is minimised. The file format is based on standard comma-separated values (CSV) files. Each line in a file represents a data record, containing several individual data values that are separated by defined separators. In total the file format comes with 15 data values per record, of which the latest is the extension value (see Table 3.1). The extension value itself contains a set of sub values, which are again divided by a defined separator.

However, in contrast to the overall file format, the content of the extension value is flexible. Thus, various values of various formats can be stored. Overall, this format allows to quickly and easily access the well-defined values in the beginning. Further, the extension field provides the needed flexibility to handle and log custom data.

Also part of the framework is a console-based post-processing engine. This engine is written in Java and runs on all common operating systems. It allows to import and process data, e.g., according to a set of given rules or an algorithm. Further, it is possible to export data into other formats, which are used by external statistic or data mining tools, like R[5] or Weka[6].

3.2 Understanding Users Through Environmental and Physiological Information

It is very important to have a detailed understanding of the user before a potential solution can be designed. This idea is manifested in various design approaches like the Human-Centred Design (HCD) process [ISO10] and applied widely in HCI. A popular example for the consequences of insufficient knowledge about the users are pedestrian navigation systems. Early pedestrian navigation systems used exactly the same navigation instructions, algorithms and modalities as car navigation systems did. Consequently, they were mostly unable to fulfil the pedestrians' specific needs, e.g., navigation in open areas, and the uptake was quite low [SRK07].

In research, pedestrian navigation has become a very prominent topic in the last few years. There are many groups working on the improvement of specific disadvantages with concrete novel interaction concepts, e.g., reducing cognitive workload or attention through tactile [PPHB12a] or audio [HMG02] orientation support. Thereby, potential solution for quite specific problems of pedestrian navigation are provided. However, with this approach only some very few and specific bits of the pedestrian's context and needs are understood and addressed. We argue that the most crucial and essential user requirements are until now still undiscovered.

Logging is known to be a valuable technique for understanding users. However, it is also known that logging has a lack of situatedness and many contextual information about the users and their behaviour can not be covered and therefore is lost. Consequently, logging is unable to provide researchers with a thorough and detailed understanding of the users, which is important to design and develop prototypes and products that meet the user requirements.

[5] `http://www.r-project.org/`, last visited January 5, 2015.
[6] `http://www.cs.waikato.ac.nz/ml/weka/`, last visited January 5, 2015.

In this section, we study how logging can be used and improved to provide a better understanding of the user needs. We investigate which data of a mobile phone contributes to the understanding of the user. Further, we investigate to what extent environmental or physiological information can further enhance or refine this understanding. We decided for these two add-ons because they were positively studied in other contexts and found to be promising for observations in the field [MM10, HNS+10].

Environmental information can be used to understand pedestrian walking patterns and route decision behaviour, which is known under the term 'spatio-temporal behaviour analysis' [RS99]. Here, environmental information are used to understand the navigation process as a whole. For example, it is applied to understand shopping behaviour, public transport infrastructure or for the development of social services [MM10, WSK+11, RS99].

Millonig and Maierbrugger [MM10] shadowed pedestrians in public transport infrastructures. They implemented a multidisciplinary observation approach and employed self-reporting, eye-tracking, shadowing and physiological measures. Their work is focusing on the methodological aspects and they conclude that it is crucial to have several data sources available, if comprehensive and valid interpretations should be made. In contrast to our work, they are discussing the behaviour at a very detailed, per person level, whereby we are interested in more general findings among users.

Webber et al. investigated how pedestrians use a mobile smartphone with Google Maps to navigate [WBM12]. Similar to this work, they also investigated how often the device is considered by measuring the glance behaviour. They identified three strategies to deal with the information provided by the mobile device. The *constant support & information* group sought for information at all stages of the route. *Independent & attentive* users are using the device to retrieve key tactical information regarding decision points. The last identified strategy is *least effort & inattentive*, where users hardly engage with the navigation and would rather be guided by GPS. Based on these findings, they propose when and how information should be presented to encourage environmental engagement.

In contrast to earlier work in this domain [MM10, WBM12], which investigates the navigation process as a whole, we try to understand and discuss problems, patterns and behaviours on a decision point level. If we are able to identify problems and patterns at this fine granularity it would be much easier to provide concrete design recommendations, which could lead to an improvement of navigation performance. Further, existing work always differentiated between individual users and did not consider the data of all users together to derive insights about the general pedestrian behaviour. It was not tried to identify patterns on a sub-route level, e.g., between two decision points. In contrast, we will analyse the data mostly independent of individual users and investigate behaviour patterns on a decision point level.

The use of physiological data gained attention in the last few years, because sensors have reached a wearable form factor and integration into all kinds of systems became easy with modern wireless communication protocols [Var07]. Physiological data was used to understand users [PVH01], to replace traditional usability measures [LOHI05], and to control interactive systems, such as a games [NKLM11].

Eisenman et al. observed cyclists through various sensors [EML+10]. They deployed a GSR (Galvanic Skin Response) sensor to measure a cyclist's stress level, which they assumed correlates with the pedalling frequency. They were neither able to find a correlation nor did they find the GSR values useful in any other way. They argue that their used sampling rate of 1 Hz was insufficient to capture the skin resistance dynamics of a human body.

Healey et al. equipped 19 participants with various physiological sensors, e.g., GSR and heart rate sensor, and collected data for about five days [HNS+10]. They found that the analysis of the results was significantly dependent on the selected analysis parameters. They further observed that dealing with physiological measures and self-assessments was challenging because of disparities between the individual measures and the complexities of emotional states.

In our own, previous work we used GSR to measure a user's stress level [PTHB11]. This stress measurement was permanently shown to the user through various different visual representations on a mobile phone, i.e., numbers, colours, trends. We found that none of the participants actively used the stress representation, e.g., to stop working and relax if the stress level is too high. Nevertheless, participants appreciated the permanent presence of the measurement, arguably supporting their overall stress self-assessment.

Related work made mixed impressions with physiological measures, such as GSR, for the observation of study participants. For dynamic outdoor or longitudinal settings, like biking, it comes with a lack of precision and, thus, a limited value. In this section, we study how physiological measures perform in a semi-controlled field study and in combination with several other sensor recordings.

We contribute to the logging methodology by showing that it can be valuably extended and enriched through environmental and physiological information. We show that particularly environmental, static information is helpful to take another perspective and investigate the dynamic logging data under much more relevant aspects. We further found that physiological sensors are promising, but there is a strong need for further practical research to bring them to actual value in log data analysis.

We further contribute to the overall understanding of pedestrian requirements. Thereby, we don't follow the common problem-based approach, i.e., proposing a solution for a specific problem. Instead, we study the pedestrians' desires from the ground up and at a very general level. Our observations show that users have different approaches to interact with a pedestrian navigation system. While some users consider the device frequently and for a short period, some users consider the device very rarely, but for longer periods. We further identified that most users consider the device nearby to a decision point and only a few users consider it between two of those. At decision points, which were rated to be subjectively and objectively difficult, the device is particularly often considered and walking speed is reduced. These findings can significantly change the way how we think about and design future pedestrian navigation systems.

This section is structured as follows. First, we explain how exactly we plan to advance the logging observation technique. Then, we outline the used observation tools and the apparatus, which we used to study the methodology. Further, we present insights from an evaluation of the methodology. We close this section with gathered key insights and conclusions.

3.2.1 Method Approach

Existing studies which investigated pedestrian walking patterns and route decision behaviour used *shadowing* as observation technique. That means that an experimenter spotted and followed an arbitrary person and took notes about their behaviour. This observation technique has the disadvantage that it is relatively imprecise and subjective, as the data needs to be recorded by a human being [MM10]. To overcome these limitations we use the *logging* technique, which is known to be objective and situated [FCC+07].

Our Virtual Observer logging framework is able to observe many device parameters at a high frequency. We found that this is already very valuable to understand users and evaluate user interface concepts [PPHB11c], particularly for challenging settings or user groups, like people with disabilities. However, as explained earlier, situatedness is still an issue, as some user behaviours and decisions are still hard to follow and to understand. In this section, we extend logging with environmental and physiological information. We think that this will give us further insights into the user behaviour and a more detailed understanding of the motivations, problems, and decisions that users have or make.

Environmental data will allow us to relate all recordings to certain, fixed elements, provided by the environment. In the context of navigation, we use the route and its waypoints as additional input. This reference material allows to have other perspectives on the data. Instead of traditional, holistic means, such

Observed Aspect	Inspiring Questions (Examples)
Screen On/Off	When does the user actually uses the navigation aid? How often is it used? How frequently is it used between two decision points? How many seconds before a decision point is the device used first?
Position	How far away from a decision point is the user first interacting with the device?
Speed	Is the user increasing/decreasing speed when interacting with the device? Is the user decreasing/increasing the speed before a decision point?
Heart rate	Is the heart rate increasing/decreasing before a decision point? Is the heart rate increasing/decreasing when the user interacts with the device?
Skin conductance	Is the skin conductance increasing/decreasing before a decision point? Is the skin conductance increasing for more complex decision points?

Table 3.2: The following information are logged and used to investigate when and how certain aspects change during a navigation task. The provided questions inspired our research, but are not entirely covered.

as the average walking speed for the route [PPHB12a, OTRK05], we can now investigate and discuss aspects like walking speed on a sub-route level. This allows us to draw more detailed and thorough conclusions, e.g., that walking speed is lower for more crowded areas.

Physiological data will further allow us to observe and understand aspects that even the users might not really be aware of, e.g., arousal or stress. This could lead to insights, which change the way we think about navigation and city planning, e.g., if overly crowded and channelised areas lead to stress. Logging extended with environmental and physiological data is the first technique, which would be able to show these aspects at a sufficient level of detail to allow informed decisions and conclusions.

The application of the enriched logging observation technique will help us to investigate the spatio-temporal behaviour of pedestrians, which use a common pedestrian navigation system, such as the PocketNavigator. With the help of the recorded logging data, we will do a holistic analysis to identify sub-route behaviour patterns in the wild. An overview of the observed aspects and which example questions could be answered with those, is shown in Table 3.2. We argue that our findings will eventually allow us to give generic design recommendations for future pedestrian navigation solutions.

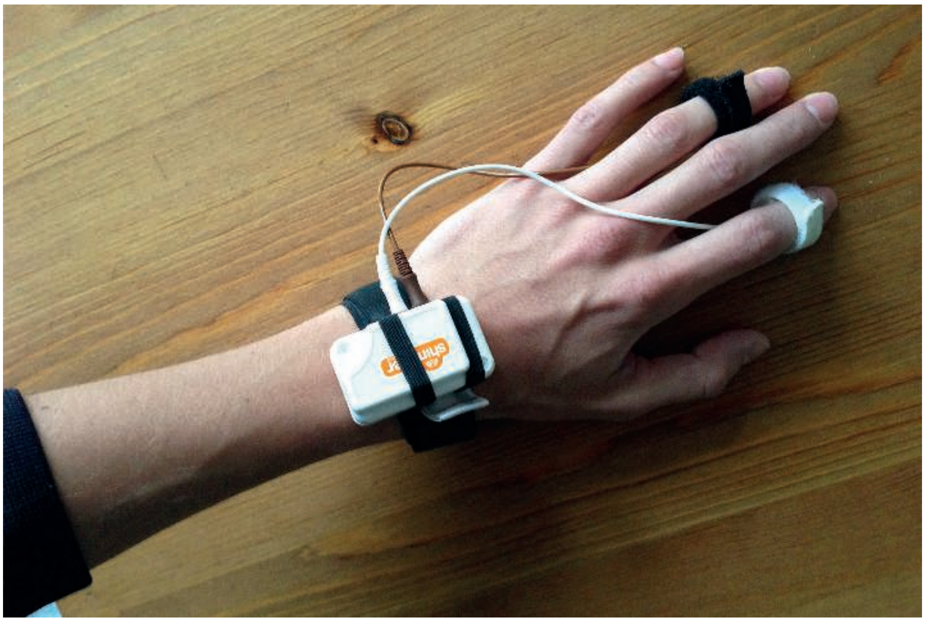


Figure 3.3: We used a Shimmer GSR (Galvanic Skin Response) sensor to monitor the user's arousal while navigating.

3.2.2 Observation Tools and Apparatus

The method approach is about understanding users. Consequently, we wanted to minimise all unfamiliar aspects, which might unpredictably effect the user's natural behaviour. At the same time, we wanted to have maximum control over the application, to log data as precise as possible. Thus, we decided against state of the art navigation solutions, like Google Navigation, and used the PocketNavigator, our custom navigation aid (see Section 3.1.1), as apparatus for our study.

We modified the PocketNavigator to not provide any tactile information, because today's common navigation systems don't provide this feature. Further, we revised the tutorial screens and other visual appearances accordingly, so that they also didn't refer to any tactile navigation instructions. Instead of dynamic route calculation we integrated a predefined route, which is displayed similarly to Google Maps. After these modifications the PocketNavigator was mostly behaving similarly to common navigation solutions, like Google Maps. Thus, no aspects were present, which could potentially influence the user behaviour.

We integrated the Virtual Observer framework (see Section 3.1.2) into the PocketNavigator. We configured the framework to run in push mode, thus, it is tried to sample and record every sensor information at very high frequency. The framework was set-up to save all information to the device.

Because we wanted to include physiological data in our observations, we connected a Shimmer GSR sensor (see Figure 3.3) and a Zephyr HxM BT heart rate sensor to the Virtual Observer logging framework. The Shimmer GSR[7] is a wearable sensor which measures Galvanic Skin Response, which is a reliable indicator for arousal and stress. The Zephyr HxM BT[8] is a sensor, which records the heart rate, also being an indicator for stress. Both sensors were connected via Bluetooth. The sensor readings were stored in the extension field of the Virtual Observer.

3.2.3 Evaluation

The primary purpose of the evaluation is to give an understanding on how environmental and physiological data can contribute to the analysis of logging data. The secondary purpose is to understand how pedestrians use a pedestrian navigation aid, like the PocketNavigator, in the wild. In the following, we will explain the exact methodology, combine the results and discussion in a findings section, and give insights into the limitations of this study.

3.2.3.1 Method

11 volunteers participated in the study, of which 7 were male and 4 female. They had an average age of 23.00 years (SD 2.10), ranging from 19 to 27 years, and were mostly university students. 7 participants stated that they own their own smart phone or are familiar with the related touch interaction principles.

For the actual observation we decided on a static route which all participants had to follow (see Figures 3.4 and 3.5). The route has a length of about 1.70 km and leads through an organically-grown area in the city of Oldenburg, which has about 160 000 inhabitants. The route leads through residential as well as commercial areas and tries to cover a variety of different navigation situations, e.g., different types of turns, varying street sizes and traffic density. Overall, the route comes with 23 decision points, i.e., places where a user has to make a decision on where to go (highlighted in Figure 3.4). The average distance between decision points is 77.52 m (SD 42.56 m), ranging from 23.52 m to 191.80 m. The route was saved in high resolution in a separate file, which we used as key environmental information for the post-hoc log data analysis.

For the study we provided a Nexus One smart phone, running Android 2.3. The phone comes with the PocketNavigator navigation application and the observation framework Virtual Observer installed. Further, we provided a Shimmer

[7] `http://www.shimmersensing.com/`, last visited January 5, 2015.
[8] `http://www.zephyr-technology.com/products/fitness`, last visited January 5, 2015.

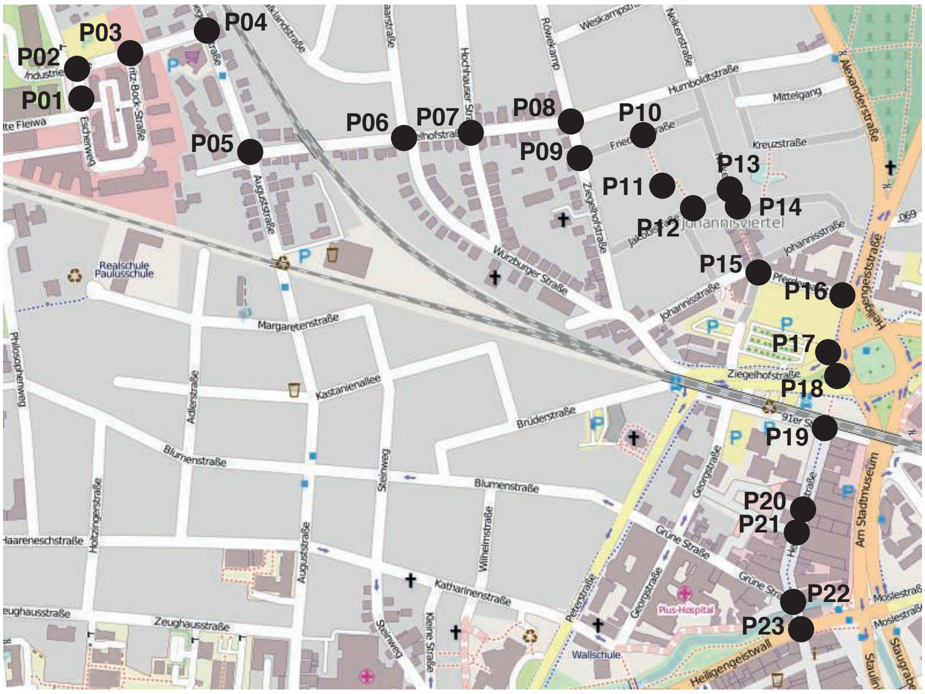


Figure 3.4: In our field study, users were asked to follow a given route through an organically-grown area in Oldenburg. This route comes with waypoints of different complexity.

(a) Waypoint P5 (b) Waypoint P14

Figure 3.5: The route leads users through an organically-grown, mostly residential area in Oldenburg.

GSR sensor to measure Galvanic Skin Response (GSR), and a Zephyr HxM BT sensor to measure heart rate. Both sensors are connected via Bluetooth.

We met the participants in our research institute, mainly because of the privacy need when attaching the physiological sensors to the body. We clarified on the overall scope of the study and asked for the participants consent. The task we asked the participants to do is to follow the pre-defined route, as described earlier. All participants were asked to walk the same route in the same direction. We asked them to behave as they would naturally behave when faced with a navigation task. We clarified that there is neither a need to follow the route with an enormous precision nor should the participants walk particularly fast or slow. The participants were asked to switch off the device screen when they don't use the device. The route ended in a cafe in the city centre. There the experimenter welcomed the participant, asked to return the phone and sensors, and conducted a short post-hoc interview. In the post-hoc interview we presented each of the decision points of the routes as a photograph and asked the participant to subjectively rate their complexity on a scale from 1 (low complexity) to 10 (high complexity). The experimenter debriefed the participant and thanked for the participation. The participants were not paid for participation.

After each study, we obtained and saved the logging data from the mobile device. When the last participant finished the study, we started with the post-hoc analysis of the data. Although we analysed all available data, we focused on the environmental data, i.e., the route with its waypoints, and the recorded physiological data, as both are key factors for this study. The data was analysed by following the common standards for spatio-temporal analysis and pedestrian navigation.

3.2.3.2 Findings

All participants reached the destination and did not deviate significantly from the predefined route. On average, each participant walked 1895.86 m (SD 108.51 m) in 20.31 min (SD 1.77 min).

We found that the participants rated the complexity of the decision points with 2.34 (SD 1.88), which is a quite low complexity. Particular decision points where the participants were asked to go straight were often rated with the lowest possible complexity (see Figure 3.6, Waypoints P01–P03 or P20–P21). Nevertheless, there were also three decision points with a high complexity, mostly those where the users had several options to continue, i.e., P05, P10, P14. The perceived complexity of decision points shows a strong positive correlation ($r = 0.62$) with the objective complexity, which we derived and calculated based on Duckham et al. [DK03].

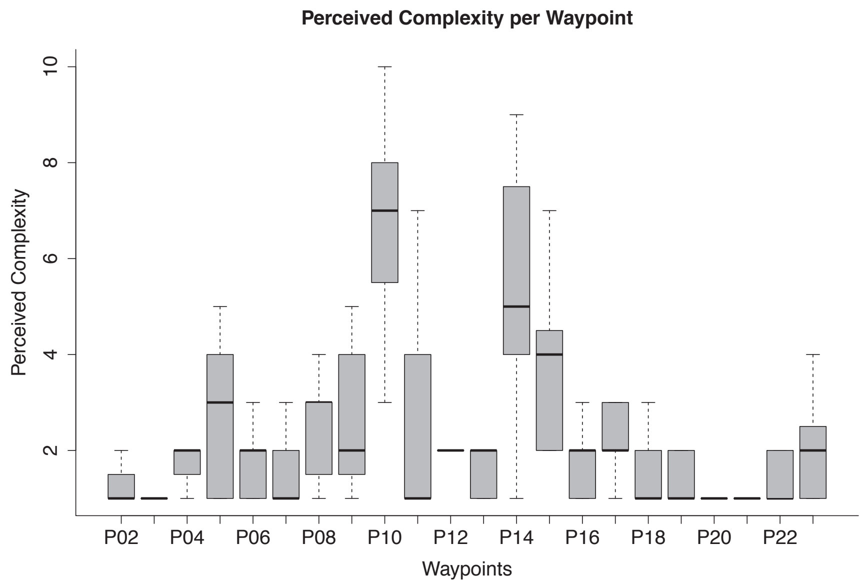


Figure 3.6: The average perceived waypoint complexity as rated by the participants on a 10-point scale (1=low complexity, 10=high complexity) after they finished the route. We observed that waypoints where participants had to turn received a higher rating.

The Less Often a Device is Considered, the Longer it Takes

We investigated how often and for how long the screen is switched on. This is an interesting indicator when the navigation information are needed. Two participants were excluded from this analysis because they didn't follow our instruction to switch off the display if they don't need the device. The remaining 9 participants considered the phone 14.44 times (SD 5.46) on average, whereby a consideration took about 27.97 s (SD 36.92 s).

We were able to identify a strong negative correlation of $r = -0.63$ between number and duration of device uses. That means that the more often a user considers the device the less time takes the consideration. This is mostly in line with earlier findings by Webber et al., although we were unable to identify the *least effort & inattentive* orientation behaviour. Future systems could use this information and provide individual feedback depending on the user type, e.g., provide abstract, at a glance feedback for the frequent device users. We assume that optimised feedback can lead to shorter interaction intervals and therefore to less distraction from the environment.

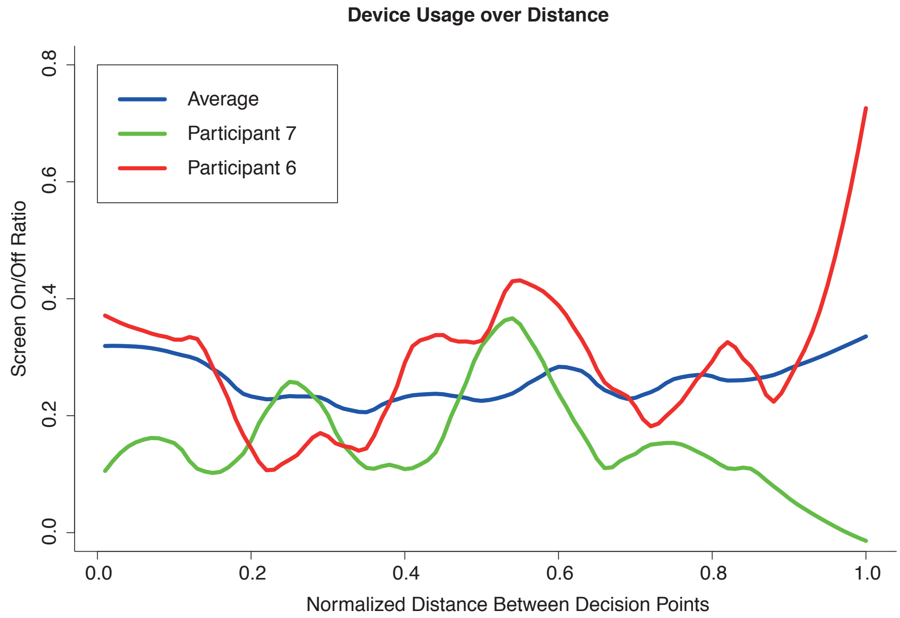


Figure 3.7: On average, the likelihood when the device is considered is quite consistent. Participant 7 considered the device between two decision points. Participant 6 shows the most frequent pattern, i.e., that the device is considered before and after a decision point.

A Device is Often Considered Nearby to a Decision Point

Further, we analysed where exactly a user considers the mobile phone between two decision points. Because the distances between the various decision points differ, we normalised the length of all sections ahead of our analysis. On average, the device is considered at each point between two decision points with about 30 % likelihood (see Figure 3.7), whereby a minor increase can be observed shortly before a decision point is reached.

Detailed investigation shows that there are two different kinds of strategies when a device is considered. Five participants watched the display shortly after and shortly before a decision point. Two participants considered the device right in the middle between two decision points. A representative graph for each user type is shown in Figure 3.7. For two participants we were unable to identify such a clear usage pattern among all decision points. They seem to consider the device at random points. In a complementary time-wise analysis of device usage we were unable to identify similar patterns. That means that future context-aware systems should present the information depending on the user location relative to a decision point and not depending on time.

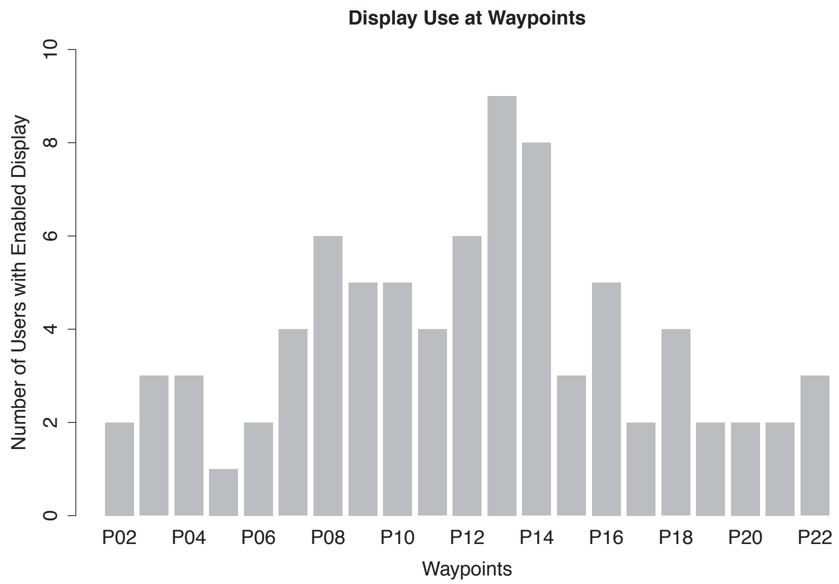


Figure 3.8: Device use at waypoints varies across users. Most users use their device at waypoints which are rated particularly complex, e.g., P14.

Difficult Decision Points Lead to Higher Device Usage

Because most participants consider the device close to a decision point, we were interested if this is true for all decision points or if decision points which were rated more complex stand out. We observed that device use varies across users, but most users use their device at waypoints with high complexity (see Figure 3.8).

We calculated a correlation coefficient of 0.68 for the subjective complexity assessment and a correlation coefficient of 0.65 for objective complexity assessment. That means that there is a strong relationship between the subjective and objective complexity rating of a decision point and if the device is considered, i.e., the more complex a decision point is perceived/rated the more likely the device is considered. Future systems could use the objective complexity assessment to avoid complex decision points in routings. Thereby overly frequent device considerations and cognitive load can be reduced. In practise, also large-scale information about a decision point's subjective complexity could be collected via community-based routing applications, similar to Waze[9].

[9] `http://www.waze.com/`, last visited January 5, 2015.

Difficult Decision Points Lead to Lower Walking Speed

In our exploratory analysis we observed that on average the walking speed of the users was varying between $3.09\,\mathrm{km\,h^{-1}}$ and $4.94\,\mathrm{km\,h^{-1}}$ (mean $4.01\,\mathrm{km\,h^{-1}}$, SD $0.29\,\mathrm{km\,h^{-1}}$). In our analysis we observed that there is a strong negative correlation between the objective and subjective complexity of decision points and the walking speed ($r = -0.58$ and $r = -0.40$). Thus, the more complex a decision point is, the slower the participant was walking. This observation can most likely be credited to the fact that participants tend to walk slower once they interact with their device. This has been observed in earlier research [PPHB12a] and is also supported by our data. In fact, we observed a significant difference ($p < 0.05$, Student t-test) in walking speed when the screen is on ($3.83\,\mathrm{km\,h^{-1}}$) or off ($4.05\,\mathrm{km\,h^{-1}}$).

No Generalisable Observations for Heart Rate and GSR

We also analysed how the heart rate and the GSR changed over time. We observed that the average heart rate per user is ranging from 89.57 beats per minute to 135.90 beats per minute, resulting in an overall average of 110.20 beats per minute. We studied if the heart rate rises or falls before or after a waypoint is passed. Here, we found that the average heart rate before a waypoint is 111.81 beats per minute and 111.92 beats per minute after a waypoint. A Student t-test indicated that this difference is not significant ($p = 0.34$).

For four users the measured GSR value reached the sensor maximum after a few minutes of walking and remained there until the end of the route. This can have two reasons. First, the users had a very strong transpiration and, therefore, a very good skin conductance. Second, the electrodes were pressed against each other, which can—despite the good isolation of the electrodes—cause an electrical shortcut.

For the remaining GSR measures we observed an almost continuous, but slight increase throughout the trip. Initially, we assumed that this increase is credited to the more crowded and complex environment in the latter third of the route. However, a separate trial with a single participant, who walked the route in opposite direction, also indicated that continuous increase. Consequently, we argue that this increase might be credited to the increasing transpiration of the study participants. Altogether, we don't think that the recorded GSR values are very reliable and, thus, we don't conclude anything from them.

3.2.3.3 Limitations

This study comes with two notable limitations. First, our conclusions are made on the basis of logging data, which we interpreted as accurately and conscientiously as possible. Nevertheless, there might be unexpected glitches in the data, which potentially accumulate and lead to invalid conclusions. Because there is no ground truth, the validity of the study insights remains questionable and they need to be handled with care.

The second limitation is the generalisability of the results. We ran the study with only 11 participants, of which a few need to be excluded from certain analysis. Further, participants only walked a single, static route. Nevertheless, our analysis aims at very essential, groundbreaking insights and is technically optimised for large data sets. Thus, it can certainly be questioned if our findings are valid and apply for the whole population.

3.2.4 Conclusions

In this section, we studied how traditional logging can be extended through environmental and physiological values in the context of spatio-temporal analysis of pedestrians. We found that particularly the environmental information, i.e., the waypoints, were a helpful addition to analyse the available logging data. In contrast, the physiological information were not that responsive as expected or unsuited for mobile contexts for various reasons. Overall, they didn't allow us to conclude anything of relevance.

We conclude that it is worth to extend and enrich logging with other, external information, as this potentially allows interesting analysis, which can reveal valuable insights. We argue that the available environmental information allowed us to significantly increase our situatedness and better understand the users' motivations, decisions, and concerns. Unfortunately, the physiological information did not contribute to this understanding, although this can be credited to the novelty of the sensors and the limited physiological dynamics of the human being. We argue that an existing body of research, like we had for the spatio-temporal analysis, is valuable to guide the researcher in the analysis process. Given that research on physiological sensors for interaction purposes has just started, we assume that their understanding will improve and they, thus, can contribute to the user understanding in the future. This argumentation is in line with previous impressions from Healey et al. [HNS+10].

Regarding the spatio-temporal behaviour, we observed that users have different approaches to interact with a navigation system in the wild. Some users consider the device frequently and for short periods, others consider the device rarely, but for longer periods. Most users consider the device at a decision point, only a few consider the device between two decision points. We found that particularly difficult decision points cause an intense device usage among all participants. A strong negative correlation shows that difficult decision points further lead to a reduced walking speed.

Our broad, initial investigation shows that there are many aspects and further open research questions that can be considered and investigated in future navigation systems. We think that the objective assessment of decision points by Duckham et al. [DK03] can be a promising starting point to detect complex decision points. Once they are accurately detected, we can think of future pedestrian navigation systems, which try to avoid overly complex decision points. Eventually, this could lead to less device considerations, a continuous high walking speed and, overall, to a more efficient and relaxed navigation experience.

3.3 Evaluating UI Prototypes Through Interaction Taxonomies

Today's smart phones often come with a digital compass, which allows to measure the device's heading. In situations where the user holds the phone in his/her hand, the device heading equals the pointing direction of the user (see Figure 3.9). Pointing over time and with varying direction is often referred to as *scanning*. Scanning is an emerging, often whole-body interaction technique. The actual scanning can be done without spending visual attention on the device, which facilitates the use of, e.g., audio or haptic feedback. The interaction technique is most prominently used in the domain of user orientation and navigation, e.g., [MRGS10a, REJ09, PPH+11]. There it is applied to, e.g., convey the direction to the next way point of a route by, e.g., presenting a tactile cue once the user points in the right direction. It is argued that this interaction technique supports exploratory navigation [Rob10].

The design space for scanning-based interaction concepts is huge, as it has to be decided which feedback is presented for which angles. Consequently, several different multi-modal interaction concepts have been investigated and published. Most of these concepts come with studies, which show that a scanning-based interaction concept is, e.g., less distracting or more efficient than a traditional interaction technique. More rarely, scanning-based interaction concepts are compared against each other or with other novel techniques. Unfortunately, most studies and comparisons use different, imprecisely defined measures to assess the advantages and disadvantages of the individual interaction concepts. There is no standardised approach to analyse the qualities and drawbacks of a certain scan-based interaction concept. Consequently, it is hard for a researcher to relate a novel scanning-based interaction technique to existing ones.

A reason why there are no standardised studies and comparisons is that existing observation techniques fail to provide comparable insights. This flaw limits the scalability of almost all observation techniques, particularly affecting those which are known to provide only quantitative insights, like logging. Further, the lack of formal comparability also threatens the validity of the made observations, which

Figure 3.9: Scanning is a novel interaction technique and often used in the domain of pedestrian orientation and navigation systems. A user holds a mobile phone almost parallel to the ground, points it to varying directions, and receives feedback.

in the end affects the situatedness of the used observation technique. Methodologically, most techniques are affected by these problems, but logging—given it's strict quantitative character—would profit most from an improvement.

In this research, we investigate if and how logging can be adapted to allow such a unified, and therefore scalable and situated, observation and analysis of scanning. We decided for logging because of the truly quantitative and objective nature of the recorded data. We plan to enhance this data with a rule-based analysis and interpretation of the data, which should reveal high-level information about how a certain scanning interaction technique performs or compares to existing techniques. Given the quantitative nature of the logging data and the strict rule-based analysis this would lead to a scalable, unbiased and situated analysis.

Related work provides various approaches how the interpretation of data can be simplified to improve scalability. Filtering of data is among the most prominent approaches. Instead of confronting a researcher with all the data, a reduced set is shown. Filter criteria can be based on previous knowledge or can be created on-the-fly by dynamically adapting the filter parameters during analysis. Another approach is the combination of data with additional information, like we did in Section 3.2. Typically, additional information is added to the data [EML+10], broadening and complicating the analysis, and rarely used to pre-separate the data, like we did. In the end, these approaches can only have a minor effect on scalability, as it is still up to the experimenter to study and analyse the data. In contrast, the rule-based analysis will provide the researcher with pre-interpreted high-level results.

In this section, we approach the problem that existing observation techniques have limitations to provide a unified assessment of novel interaction techniques, such as scanning. We study how logging can be advanced through rule-based analysis to approach this problem and provide researchers with scalable, situated insights. To do so, we start with an investigation of which measures define a scan interaction. We use these insights to define a set of rules, which automatically identify, characterise, and assess a scan movement. The derived rules cover the frequency and duration of a scan event, the walking speed while scanning, and which angle span is covered. In a field study with 15 participants we test our rule-based approach and compare our findings against findings from literature and earlier work. We conclude that rule-based logging analysis can reveal valuable, highly detailed insights, that allow a scalable and situated assessment and comparison of novel interaction techniques.

We begin this section with background information on scanning and how it has been studied and compared in earlier work. We then describe our rule-based approach and the automatic analysis of logging data. We explain how we integrated our approach into the Virtual Observer framework and how the PocketNavigator served as apparatus for our study. We present the evaluation of the observation technique and finish this section with concluding remarks.

Parts of this work were published in Poppinga, B.; Pielot, M.; Heuten, W.; Boll, S. Towards an Objective Comparison of Scanning-Based Interaction Techniques. Workshop on Haptic and Audio Interaction Design (HAID), 2012, Lund, Sweden.

3.3.1 Background: Scanning

In the following, we investigate how scanning has been used in related work and which measures were used to evaluate the applicability and performance of the interaction method.

Robinson et al. used a scanning-based interaction technique to browse the environment for geo-located content [REJ09]. Tactile feedback is received, if content has been discovered in a scan movement. They found that participants felt familiar with the scanning technique from the beginning and mostly interpreted the feedback while walking. They used the scan duration as measure for the effectiveness of finding content.

The PocketNavigator (see Section 3.1.1) also comes with a scan-based interaction. In a first study we compared the concept with a map and measured the task completion time, disorientation events, and distractions from the environment [PPH+11]. In a follow-up study [PPHB11a] we extended our measures by navigation errors, occurred orientation phases, and by measuring the overall walking speed. We further discuss the task completion time more detailed and report total scan and interaction times. We concluded that future work on scan-based interaction is needed to investigate if this technique can be made more intuitive.

Magnusson et al. [MRGS10b, MRGS10a] used audio feedback to communicate whether a user is pointing at the correct direction. Varying angles, ranging from 10° to 180° have been studied. The time to reach the destination, i.e., task completion time, has been used as effectiveness indicator. They found that a narrow 10° angle and wide 180° angle lead to rather long completion times. For exact track following 30° to 60° are recommended. If low cognitive load is important, angles between 60° and 120° should be chosen. They further report about different scan techniques, i.e., wrist flex, arm scans, and whole body rotation.

Rümelin et al. [RRH11] compared their vibro-tactile navigation system NaviRadar with the PocketNavigator and regular spoken instructions in an outdoor study. To compare the systems they measured disorientations, i.e., stopping for more than 2 s, and navigation errors, i.e., travelling in an incorrect direction for more than 5 m.

As the related work shows, scan-based interaction techniques are typically evaluated or compared in a case-by-case field experiment. This is sufficient to show the effects of a novel technique in a real context, but requires researchers to reinvent data analysis for each case and hinders comparability of individual studies. We aim to provide a rule-based logging approach, which provides the research community with a unified analysis technique.

3.3.2 Method Approach

For this method approach, logging is used as baseline, because its inherently objective and accurate nature. Particularly logging of sensor data is unbiased and provides a researcher with valuable measures and insights. Literature shows that logging can be flexibly extended and improved in various regards, making it a promising platform for future applications. The method approach presented here is to extend logging with a rule-based automatic interpretation, which provide researchers with high-level insights that are easy to interpret and to understand.

We empirically realise and study this method approach with the scanning interaction technique. In scanning a user points a device in various directions and is confronted with feedback based on the scan direction. This can be used for, e.g., outdoor navigation. Pointing in directions is one of the most prominent and

Figure 3.10: A scan event was defined if roll and pitch angles both are between $-16°$ *and* $16°$*. This visualisation shows how the pitch angle is defined.*

natural gestures known, which makes scanning a very intuitive and relevant interaction technique. In practise, scanning is recently experiencing its breakthrough from research into commercial applications.

When a user scans, the scanning device, e.g., a mobile phone, is typically held almost parallel to the ground. The most appropriate sensor to measure this state is a three-dimensional accelerometer. Through combination of the acceleration values, the device posture can be extracted. The device posture consists of three angles, i.e., pitch, roll, and yaw. Pitch describes how the device is pitched towards or away from the user. The roll angle describes how the device is tilted to the left or right side. Yaw defines the direction the device is pointed at. Pitch and roll are sufficient to determine if a device is parallel to the ground, whereby yaw serves as the actual input for the scanning interaction.

Consequently, to automatically identify a scan event in a data stream, roll and pitch need to be considered (see Figure 3.10). Through literature analysis and previous experience we defined that a scan event is present if both angles, roll and pitch, are between between $-16°$ and $16°$. Through empirical testing we found that this definition alone is not sufficient. We also removed the first second of an identified scan event, as here sensor values are tending to be imprecise. A scan event ends, if these angle intervals are exceeded for more than 3 s. This threshold helps to smooth shaky user hands, resulting in a strong variation of recorded pitch and roll angles. To cover only real and intended scan interactions we excluded scan events with a duration of less than 1 s and more than 120 s.

From literature review we know that the following aspects of a scan motion are of interest: frequency, duration, walking speed, and angle span. Scan frequency can be measured by counting the number of detected scan events. To measure scan duration, time needs to be recorded in the observation. Walking speed can be determined if the GPS location of the user is tracked or GPS speed is recorded directly. Angle span can be derived by measuring the pointing direction, i.e., the yaw angle. For lack of precision this should not be an accelerometer, but a gyroscope or compass.

3.3.3 Observation Tools and Apparatus

We used the PocketNavigator (see Section 3.1.1) as apparatus for our study, as it also makes extensive use of the scanning interaction technique in several regards. The PocketNavigator was prepared to run in three different modes, which later serve as conditions for our study. In the *tactile only* condition, the map was not shown to the user (i.e., a black screen was shown instead) and only tactile feedback was available. In the *map only* condition, the map was shown and no tactile feedback was provided. In this condition the map was by default rotating, i.e., aligning to the user's heading. The rotation of the map could be turned off by the participant, which resulted in a north-up oriented map. We also had a *combined* condition, where tactile and map feedback were both available. We decided on these conditions to ensure a maximum comparability to our earlier studies [PPH+11, PPHB11a].

The Virtual Observer logging framework (see Section 3.1.2) was used to observe the user via the logging observation technique. It was configured to run in push mode and save the sensor data to the device's memory card. We implemented the described rule in the Virtual Observer post-processing engine. Thereby, all scanning events are extracted and prepared for later analysis. The rule-based analysis was applied to all conditions in an equal manner. We argue that the remaining data set covers only intentional scan interactions. Further, the values recorded by the Virtual Observer are sufficient to argue about the identified values of interest, such as scan duration or angle span.

3.3.4 Evaluation

In the following, we investigate how accurate and comparable the observations made by the rule-based approach are, when compared to a traditional shadowing observation approach. We explain the used methodology, report our findings, and discuss the limitations of this study.

3.3.4.1 Method

We conducted a user study to record a typical set of scan movements in the field. 15 volunteers, 8 females and 7 males, participated in our study. The participants were aged from 20 to 29 years (mean 23.6, SD 2.5). Thirteen participants were university students, two were part- or full-time employed.

We used a HTC Dream smart phone, which was running the PocketNavigator application. The study was conducted in the city centre of Oldenburg, a city with about 160 000 inhabitants. We used the Virtual Observer framework to record several sensor values (i.e., accelerometer, compass, GPS) to the device's memory card.

We decided for a within subjects, i.e., repeated measures, design. Each participant was asked to walk three pre-defined routes of approximately equal length (i.e, about 500 m, containing 10 to 11 way points) and complexity (i.e., number of decision points, like crossings). For each route one of three conditions was assigned: *tactile only*, *map only*, and *combined*. To cancel out potential sequence effects, we counter-balanced the conditions.

Initially we asked each participant to sign an informed consent. Before we started with the study we made the participants familiar with the device and our navigation software. We explained all the functions and the participants were able to explore how the device behaves in various situations. On a short training route we answered remaining questions. We particularly emphasised that the participants can grasp, hold and interact with the device as ever they like. We started the study with one of the three conditions enabled. The experimenter followed each participant and video recorded them. We did not assist or influence the users during the actual navigation and orientation task. Once a participant reached the destination of a route, the experimenter changed the condition and the user was then able to go on with the navigation. At the final destination we conducted a brief semi-structured interview, where we asked the participants about their subjective impressions on the different interaction methodologies. Finally, we thanked each participant and handed an USB stick as a reward for participation.

3.3.4.2 Findings

In the following we present the results and discussions in combined paragraphs. We report our findings on the scan frequency, the scan duration, the walking speed while scanning, and the covered angle span of scan movements. An tabularly overview of the results is shown in Table 3.3.

Measure	**Tactile**	**Map**	**Combined**	**Pot. Impact**
Frequency	16x	13.5x	17x	Physical demand, context switches
Duration	21.05 s	17.35 s	18.87 s	Physical demand, mental demand, intuitiveness
Speed	$0.66\,\mathrm{m\,s^{-1}}$	$0.93\,\mathrm{m\,s^{-1}}$	$0.71\,\mathrm{m\,s^{-1}}$	Efficiency, mental demand, training level
Angle Span	136°	106°	122°	Physical demand, conspicuousness, insecurity

Table 3.3: An overview on all identified measures, their values for our exemplary study, and their potential impact on the user.

Scan Frequency

Based on our data set we investigated how often a user scanned. This measure is named frequency and is already used as measure how attentional resources are spend [OTRK05]. In the *tactile only* condition scanning means that the user made use of the higher precision of the compass and therefore perceived more responsive and accurate tactile feedback. In the *map only* condition scanning means that the user was probably looking at the display. In the *combined* condition it is unclear, whether a user relied on tactile feedback or was watching the screen.

We found that in the *tactile only* condition a user scanned approx. 16.0 times in average. For the *map only* condition a user scanned 13.5 times, while in the *combined* condition 17 times. Given the complexity of the route that means that a user scanned every 31.25 m (*tactile only*), 37 m (*map only*), or 29.5 m (*combined*). A conducted ANOVA indicated that no significant differences can be found ($F(2, 42) = 0.71$, $p = 0.50$).

Each started scan interaction means a context switch, i.e., the user starts to pay more attention to the mobile device and less attention to the environment. Obviously the context switch itself and the following scan action takes time. Therefore, the scan frequency is a very efficiency- and performance-critical measure. We further observed that users are most likely performing a scan movement if they feel insecure on how to proceed in the way finding/navigation task. Thus, scan frequency is an indicator on how often a user needs reassurance, probably because of insecurity.

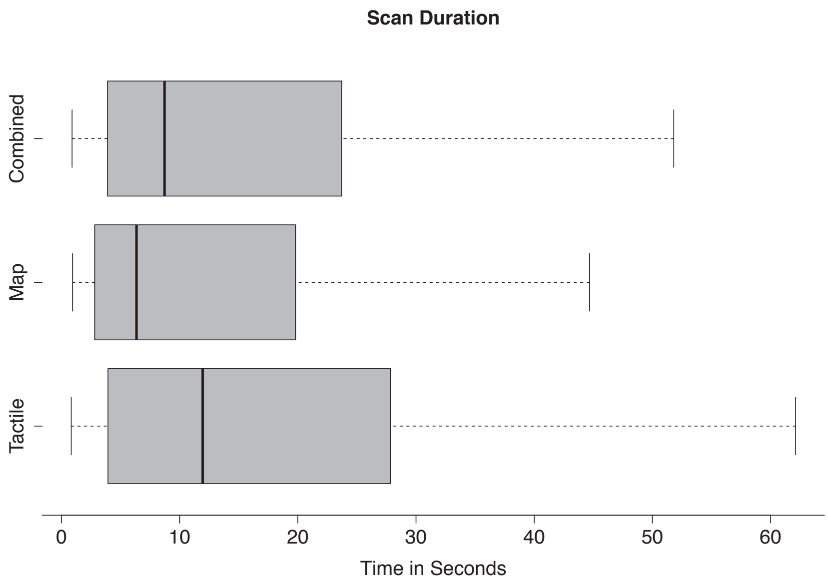


Figure 3.11: Scan events in the tactile only condition take 21.05 s *in average. We observed no significant differences between the conditions.*

Scan Duration

We also investigated the scan durations and found that for the *tactile only* condition, an average scan event took 21.05 s. For *map only* it took 17.35 s and in the *combined* condition 18.87 s (see Figure 3.11). A conducted ANOVA indicated no significant differences between the three conditions ($F(2, 694) = 1.30$, $p = 0.27$).

The duration of scan events is an important characteristic of an interaction technique. We assume that the longer a user scans the greater the physical demand is. Further, the duration could give insights on how intuitive a technique is. I.e., a short scan duration could be an indicator for an intuitive technique. Missing intuitiveness means that a user has to actively think about what is perceived, which could be interpreted as an increased mental demand.

Walking Speed

As shown by other papers, the walking speed is an interesting measure not only for pedestrian navigation [BYJS05]. We found that the walking speed in the *tactile only* condition was $0.66\,\mathrm{m\,s^{-1}}$. For the *map only* condition the speed was $0.93\,\mathrm{m\,s^{-1}}$, while it was $0.70\,\mathrm{m\,s^{-1}}$ for the *combined* condition (see Figure 3.12). A conducted ANOVA ($F(2, 694) = 5.53$, $p < 0.01$) and a Bonferroni-corrected post-hoc t-test showed that the walking speeds for the *combined* and *tactile only*

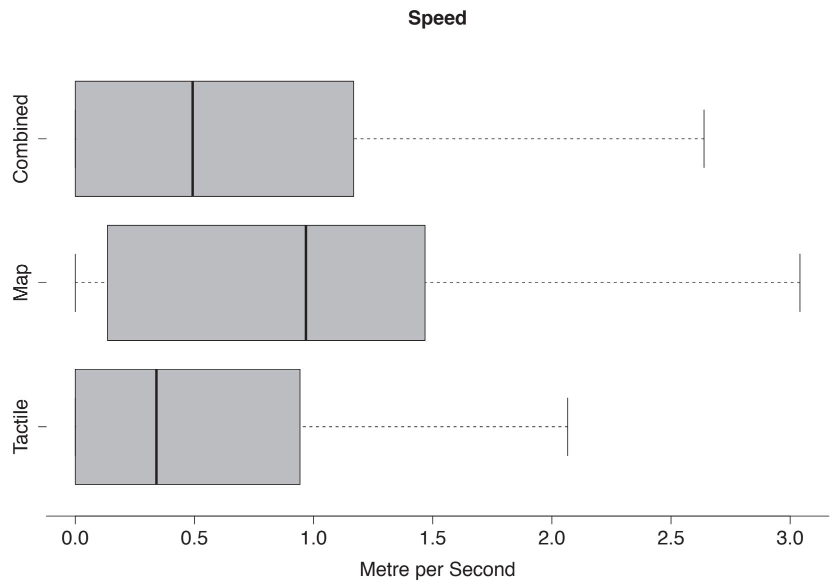


Figure 3.12: The average walking speed is significantly lower in the combined and tactile condition.

condition are both significantly lower than for the *map only* condition ($p < 0.05$ and $p < 0.01$).

We observed that a user often continues to walk while scanning. Thus, the walked distance over time, i.e., walking speed, is also a criteria to characterise a scan interaction technique. We argue that if a user walks, compared to a non-scanning situation, significantly slower, this could be an indicator for the induced mental demand of the scanning-based interaction technique. We further argue that this value might be an indicator on how trained a user is to the interaction technique.

Covered Angle Span

We further investigated the span, i.e., the overall covered angle, of a scan event (see Figure 3.13). We found that an average scan motion in the *tactile only* condition spans 136°. For the *map only* condition we observed 106° and for the *combined* condition 122° in average. Compared to the *map only* condition that makes a difference of about 30° for the *tactile only* condition. A conducted ANOVA ($F(2, 694) = 4.17$, $p < 0.05$) and post-hoc Bonferroni-corrected t-test ($p < 0.05$) indicate that this difference is significant.

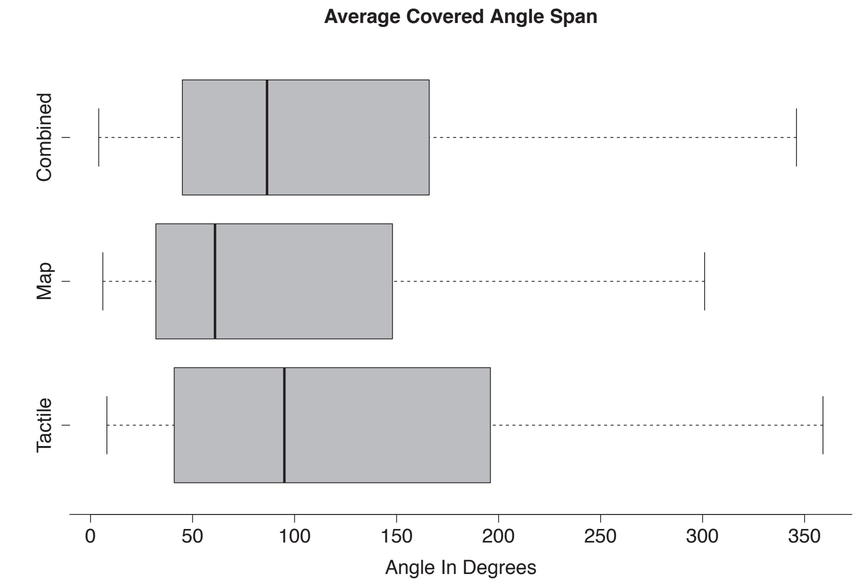


Figure 3.13: When compared to the map only condition the angle span is significantly wider than in the tactile only condition.

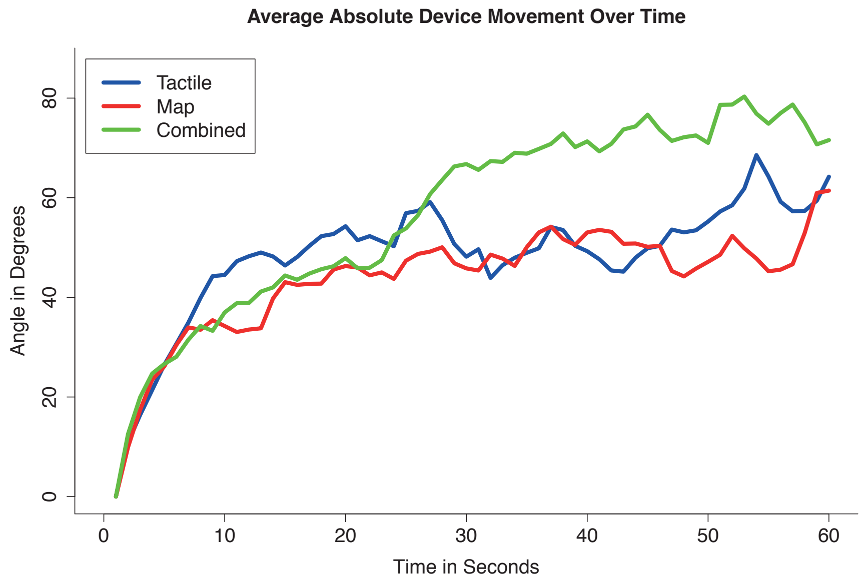


Figure 3.14: Within the first $10\,\mathrm{s}$ *of an average scan event the device is moved rapidly, after that the angle remains almost unchanged.*

We analysed how the scan angle varies over time (see Figure 3.14). To make the following analysis independent of the direction of scanning, we used the absolute value for each recorded angle. We found that the average angle dramatically changes in the first 10 s of a scan event from 0°, i.e., straight ahead, to about 50° for the *tactile only* and *map only* conditions, and about 70° for the *combined* condition. After the initial increase, the angle mostly remained unchanged.

A broad angle means that the user has to turn significantly towards left or right. At a certain point the user is unable to cover this angle with wrist or arm movements and inevitably needs to do full-body movements. If these obvious movements are necessary it is also more likely that passers-by notice the interaction process. We further observed that users tend to cover broader angles if they feel insecure with the provided feedback. Therefore, we argue that the covered angle span could give insights on how physical demanding and conspicuous an interaction technique is.

3.3.4.3 Limitations

One technical limitation is that out application used predefined angles to trigger the scanning mode in the *tactile only* and *combined* condition. This limited the user's interaction space in advance, but doesn't limit the potential set of measures itself. The automatic detection of scan events is another limitation, as this technique might come with inaccuracies, i.e., the detection of a scan event where is none. We further want to point out that the described study is by design not capable to give any insights on how effective the derived metrics are to actually distinguish between different scanning techniques. Finally, we want to emphasise that the vaguely illustrated potential impacts have been derived from mostly subjective impressions, i.e., the experimenter's observations and the participants' comments during the interview.

3.3.5 Conclusions

It is common practise to study and compare novel interaction techniques on a by-case basis. This limits the scalability and validity of such studies. In this research, we did a rule-based interpretation of logging data to exemplarily study scan interactions. We identified an initial set of objective measures, i.e., frequency, duration, walking speed, angle span, with which a scan event can be technically defined independently of any concrete scenario. In a study we found that observations made through the rule-based interpretation are similar to earlier findings [PPHB11a, PPB10, PHB09]. Further, thanks to the precision of logging, very detailed analysis are possible that were impossible with traditional supervised observation techniques. Altogether, this makes us argue that rule-based interpretation of logging data is a valid and valuable approach, and allows a systematic and objective analysis and comparison of various interaction techniques.

For researchers the rule-based logging analysis will lead to observations with increased scalability and improved situatedness. Scalability will be improved, because researchers can apply the rule-based analysis automatically and gather high-level observations. Thus, no costly manual data analysis is needed. In addition, the automatic interpretation will lead to an increased situatedness, because predefined parameters are analysed and displayed. These measures might have been overlooked or considered unimportant otherwise. A predefined set of measures can support a researcher in understanding what are the essential bits of a scan movement or any other interaction technique.

We want to emphasise that this research used scanning as an example interaction technique. Although quite likely, the generalisability and applicability of the method approach to other interaction techniques is unclear and needs to be studied in future work. Further, we identified the rules based on literature review and background knowledge. Because we found this to be a complex and time-consuming task, we propose the simplification of rule creation, e.g., by demonstrating the interaction to be extracted. We envision that a researcher demonstrates the interaction of interest and a logic unit transfers this demonstration into a set of rules. Depending on the complexity of the interaction, the logic might be an unsupervised data mining algorithm or a complex machine learning engine.

3.4 Summary

In logging, quantitative data about the users, their actions, the environment or whatever else is of interest for a researcher is recorded for later analysis. It comes with two limitations. First, the gathered insights have a limited **situatedness**, which means that not all information about the user context and situation can be covered. Second, logging comes with a non-optimal **scalability**, because the analysis of raw logging data typically happens on a by-case basis, which is time consuming. In this chapter, we studied two approaches to deal with these limitations. First, we investigated to what extent *environmental and physiological information* can be used to enhance the information gain of logging. Second, we researched how rule-based processing of logging data can lead to high-level insights, which ease the overall analysis process.

We found that external cues, such as environmental or physiological information, allow a researcher to study logging data in higher detail and from various perspectives. In our concrete example, waypoints allowed to study spatio-temporal pedestrian behaviour on a sub-route level. Where earlier approaches allow to study aspects like walking speed per route, our approach enables to study these between individual waypoints, i.e., in a much higher granularity. Consequently, a researcher gets more detailed insights, leading to an improved situatedness and a better understanding of the user. This fine-granular data further allows a more focused analysis of the data, positively affecting the scalability of the method.

We further showed that the rule-based analysis of logging data can lead to valuable high-level insights. In an empirical study, we found that the popular scan interaction technique can be analysed in logging through a set of rules. Thus, researchers don't have to focus on details and understand every bit of the data. Instead, they can focus on an aggregated data set, which allows a simplified analysis. The core advantage of this approach is that it increases the scalability of logging. In addition, it contributes to a standardised understanding of interaction techniques, making comparisons and studies more traceable and valid.

Both, the extension of logging through environmental and physiological data, and the rule-based analysis are valuable additions to the logging technique. We have shown to what extent they advance the methodology and illustrated how they can be adapted to domains beyond pedestrian navigation and orientation. Consequently, these improvements make a methodological contribution on a general, application-independent level, which can improve future studies significantly.

Figure 4.1: Notifications appear on the display, trigger the vibration motor and play a sound. If they are not attended after a while, reminders are repeated. This behaviour makes notification-based observation techniques, such as the Experience Sampling Method, perceived as obtrusive.

4 Reducing Self-report Obtrusiveness

Today's smart phones often use notifications to attract the user's attention on the device to indicate that something more or less important has happened. Typical reasons for a notification are, e.g., an incoming text message or a set reminder. While some notifications have informative reasons, e.g., an available application update, and can be handled with delay, most notifications require immediate user attention and action. For this reason, most notifications are presented in an obtrusive way, e.g., by visual screen appearance, short vibration, and by flashing an LED (see Figure 4.1). If not attended, some notification reminders are repeated.

Notifications are also used for user observations. They can either serve as a reminder for the user to do something study-related, e.g., call the experimenter for an interview. Further, they can inherently serve as mean for user observation, where a notification is linked to a self-reporting procedure. The most prominent technique, which uses notifications, is the Experience Sampling Method (ESM). It was originally proposed to be a pen and paper technique [CL87, LC83], but was adapted to the world of smart phones and notifications [FCC^{+}07]. In ESM, a smart phone issues a notification to initiate a self-reporting procedure. Each notification is linked to a short questionnaire or something similar, where the users enter information about their current situation. Consequently, the researcher gathers in situ data about the users' experiences through self-reporting.

Reoccurring notifications from the Experience Sampling Method are perceived as **obtrusive**. It is undoubtedly valuable to have means like notifications to quickly reach the user and, in case of a study setting, to sample in situ user experiences. However, each issued notification distracts a human from their ongoing activity. Depending on the situation a user is in, an interruption is perceived as particularly annoying and obtrusive. For example, a user will feel disturbed when a notification is issued while driving a car or while sleeping. It has been shown that unwanted notifications lead to stress, increased frustration, time pressure and effort [MGK08].

With technological advance there are more and more machines touting for our attention, causing external interruptions. Because machines are insensitive to whether a human is able to attend a notification or not [AFH07], active interruption management is needed. In theory four design solutions to cope with interruptions were identified: immediate, negotiated, mediated, and scheduled interruption [ML02]. Mediated interruptions use indirect information, e.g., a human's digital calendar or environmental sensors, and are recognised among the less interruptive techniques [McF02].

Creating context-awareness by modelling the user's interruptibility has been proposed to be a solution to this problem [FHL04]. The idea is that a context-aware phone somehow identifies if it is appropriate to trigger a notification and thereby reduce obtrusiveness. While it has been shown that such opportune and non-disruptive moments can be reliably identified for workers on stationary desktop computers in [FHL04, FHA$^+$05], research is still ongoing for mobile and dynamic outdoor scenarios.

While there are many factors that influence a person's interruptibility, e.g., social engagement, literature found the user activities to be very relevant. Interruptions between coarse breakpoints, i.e., major changes in the work flow, produce less annoyance and users assessed them as being more respectful to their ongoing activity compared to other, e.g., random moments [AB04, BI08]. To detect and differentiate between these breakpoints in interactive tasks, models can be created. Such models for stationary computers in office contexts reach an accuracy of up to 78 % [FHA$^+$05].

However, in comparison to stationary computers, mobile devices are used in much more dynamic ways and in by far more complex contexts. Fischer et al. argue that the endings of mobile interactions, like making phone calls and receiving SMS, denote a coarse breakpoint and are opportune moments to attend notifications. They found that a user deals with a notification more quickly if it is triggered at coarse breakpoints. Further, they found that the type of task has a significant effect on acceptance time and completion rate. Despite the triggering in coarse breakpoints, some users still noted that notifications and subsequent tasks were annoying in certain situations [FGB11].

Instead of using user-entered and phone-specific information, like a received SMS, also sensors could be used to identify opportune moments. Ho and Intille used two accelerometers and activity recognition to trigger interruptions at moments when user changes activity, e.g., from sitting to walking [HI05]. They showed that users are significantly more receptive to interruptions compared to a random condition. Kern and Schiele used wearable accelerometers, audio, and location sensors to decide whether a user should be notified and which modality to use. They argue that no advanced model is needed, as the combination of tendencies is already sufficient [KS03].

Earlier work [FGB11, HI05, FHA+05] studied about 20 users for a maximum of 14 days, whereby they investigated a maximum of 2000 notifications and achieved quite artificial response rates of up to 90 %. Most of the existing related work focused on specific predictors, such as calendar entries, or concrete settings to predict interruptibility. Consequently, a from the ground, longitudinal investigation of the matter is still missing. In this chapter, we do such a structured investigation, where we aim to model interruptibility with sensors and technologies that are available in today's commercial smart phones.

To study interruptibility and opportune moments for notifications we designed MoodDiary, a mobile diary application for mood tracking, which makes intense use of notifications. For every issued notification the whole context information, based on the mobile phone's sensors, and whether the notification is answered or not is recorded. MoodDiary was published in the Google Play store, where we collected 6581 notifications from 79 users. A created model is 77.85 % accurate in predicting opportune moments to issue notifications. In contrast to earlier work, our observations are made in real life, in the wild and over a long period of time, which results in high external validity and broad applicability of our findings.

We argue that our insights can help designers and developers of mobile software to trigger notifications in less disruptive moments. Our model can further be applied to all self-reporting observation techniques, like ESM, also making their inquiries able to appear in opportune moments. Consequently, we conclude that the application of the model will make self-reporting techniques less obtrusive. We further argue that the appearance at opportune moments will likely increase the quality and richness of provided insights, positively affecting the overall situatedness of gathered data.

This chapter is structured as follows. First, we describe our method approach in Section 4.1 explain how exactly we want to create a sensor-based model for predicting opportune moments. Then, we present how we designed and developed the MoodDiary application, how the sensor data is recorded and how we distributed in in the Google Play store (see Section 4.2). We then present a statistical analysis of the recorded data and how we processed the data to create and evaluate a model (see Section 4.3). We discuss our findings in relation to their practical relevance and applicability and conclude this chapter with implications of this work on self-reporting techniques in Section 4.4.

Parts of this work were published in Poppinga, B.; Heuten, W.; Boll, S. Sensor-Based Identification of Opportune Moments for Triggering Notifications. IEEE Pervasive Computing Magazine, 13(1), pp. 22–29, January–March 2014.

4.1 Method Approach

In this chapter, we study the contexts in which notifications are considered and in which they are rejected or ignored by the user. To do so we design an application, which logs the context a user is in when a notification is triggered and if this notification is answered or rejected. We will use the gathered data to develop a predicting model on what are opportune moments to issue a notification in mobile contexts.

A critical decision to make is which information should be considered in the to-be-created predictive model. Previous work used either user-entered information, like calendar entries, or sensor data. User-entered information has heterogeneous levels of availability and quality, as it might be sloppily maintained and is therefore incorrect or outdated. Further, this information is personal and users might not want to share it with an unknown application or third party. In contrast, sensor data is homogeneous as most smart phones come with similar sensing capabilities and the measured values are often normalised and well defined.

Consequently, we follow the sensor-based context observation approach as we think this is more reliable and promising. Our approach incorporates sensors which are available in nowadays smart phones: GPS, accelerometer, gyroscope, compass, microphone, and a proximity sensor. These sensors can obtain various measures, which can be used to observe various aspects of the users' notification answering behaviour. In fact, the recorded context features are: timestamp, location provider, position accuracy, speed, GPS heading, compass heading, roll, pitch, proximity, and light level. We neither collected location information nor assessed the microphone for ethical reasons. Although we use state of the art sensors, we want to emphasise that these can only observe a tiny fragment of what makes a holistic mobile context.

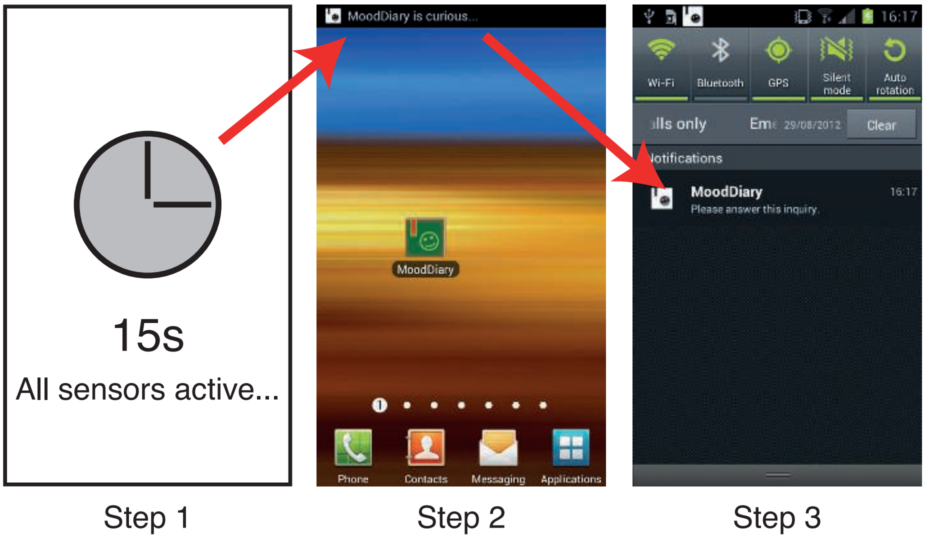


Figure 4.2: The process how a notification is issued and answered consists of six steps. The first three steps cover the capturing of sensor information (Step 1), the actual triggering of notification (Step 2) and the user's notification handling (Step 3).

4.2 Observation Tools and Apparatus

To collect context information and trigger notifications on the users' mobile phones, we needed an apparatus. For our study we decided on a self-tracking application, which tracks the user's mood at regular intervals via notifications. This decision is based on the fact that a human's mood can change frequently over a day, which justifies short intervals between notifications. Basically, any notification-triggering application could be used for this research; actual mood assessments are not relevant for this work and will not be presented or discussed.

The application is called MoodDiary, runs on Android, and is separated into two parts. The first part is a background service that regularly triggers notifications and asks the users to assess their current mood. The second part is an activity that provides an overview of all mood assessments and thereby creates an actual value for the user. We paid attention that the application is of good quality, stable, and reliable. We created an appealing logo and application description to attract many users. The application description makes the user aware that it collects anonymous sensor information for study purposes.

4.2.1 Trigger New Notifications

The background service triggers a new mood assessment notification at regular intervals of about three hours and fifteen minutes. The regular sampling strategy

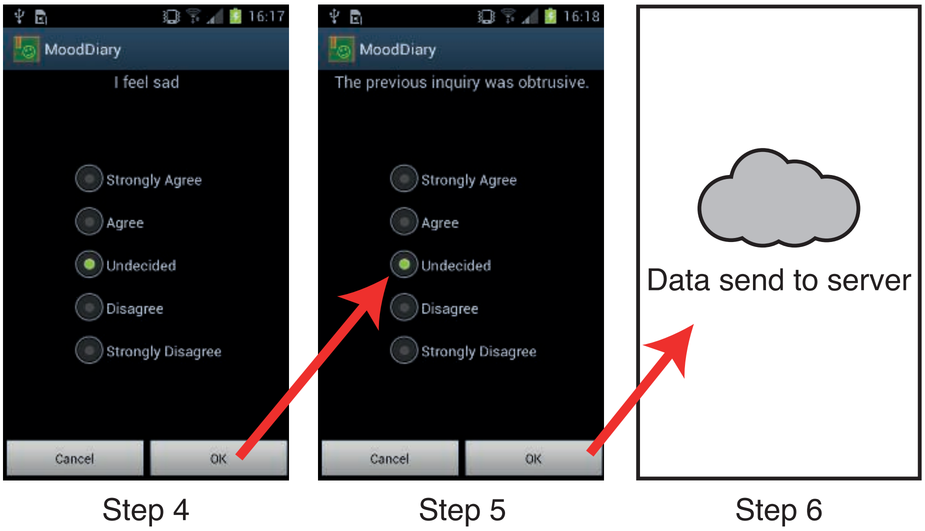


Figure 4.3: After the sensor data is recorded and the notification is issued (see Steps 1–3 in Figure 4.2) the actual mood assessment starts. In Step 4 the user has to rate a mood-related statement. Step 5 records the perceived obtrusiveness of an inquiry. In Step 6 all recorded information are transferred to our servers.

is realistic for such an application and eventually allows us to cover a whole day. The service runs continuously and automatically restarts on phone reboots. For study purposes the service invisibly infers the user's context 15 seconds before the actual notification is triggered (see Figure 4.2, Step 1). After context gathering, a snapshot of the context is taken and the actual notification is shown, which basically looks and behaves similar to, e.g., traditional SMS notifications (see Figure 4.2, Step 2 and Step 3). The notification takes the phone profile into account, i.e., no audio alert is played if the phone is muted. A user can either answer a notification or reject it. After one minute the notification is automatically dismissed and will no longer be shown in the notification bar.

If a user opens a notification, a sequence of two dialogues comes up (see Figure 4.3, Step 4 and Step 5). Those dialogues consist of a statement at the top and a 5-point Likert scale in the middle, which ranges from strongly agree to strongly disagree. The first dialogue gives a statement about a certain mood. The supported mood types are based on findings of Charles Darwin and Paul Ekman: angriness, fearfulness, happiness, sadness, how disgusted and how surprised someone feels. Thus, a possible first statement could be "I feel surprised". The second statement assesses the inquiry's obtrusiveness: "The previous inquiry was obtrusive". After the user clicks the OK button in the second dialogue, the dialogue disappears and the assessment is stored in a local database.

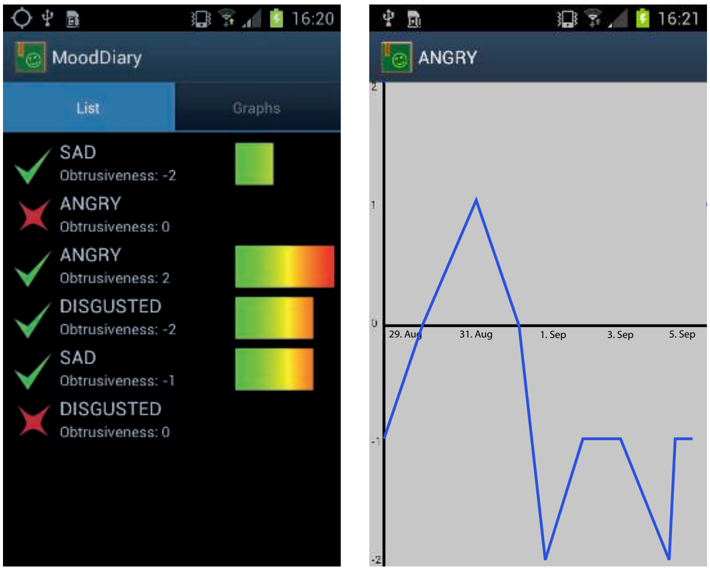


Figure 4.4: MoodDiary also provides a value for the user. All answered and ignored assessments can be shown and it can be visualised how the user mood changes over time.

Independently of whether a notification was attended or not, we log information to our servers (see Figure 4.3, Step 6). We log the inferred context information (as described earlier), if a notification has been answered, the mood type, the given answer, and the rated obtrusiveness of each inquiry. We further log information like the device type, the set locale etc. All information is collected anonymously with the users consent.

4.2.2 Value for the User: Overview of Mood Assessments

Beside the background service, the application also provides an activity that shows earlier assessments. This activity is decoupled from the background service and consists of two tabs (see Figure 4.4). The first tab contains a list of all answered or rejected inquiries. Each list element shows which mood type was asked for, if this inquiry was answered, how the user self-assessed his mood and how obtrusive this inquiry was rated. The second tab provides the user with graphs how the individual mood types have changed over time. However, this part of the application is not of relevance for this work.

4.3 Evaluation

We released the MoodDiary application, to the Google Play store in March 2012 and did our analysis six month later, i.e., in September 2012. In the following, we investigate the data from two complementary perspectives. First, we present the most relevant descriptive statistics in which contexts a notification is typically attended or ignored. Second, we create a model to predict opportune moments and investigate its performance.

4.3.1 Statistics

The application was installed by 314 users over a period of five month. Overall 15 926 issued notifications were recorded by our server. To exclude users that installed the application without really using it, we applied a filter. We excluded all users that used the application less than 1 day and had an answering rate, i.e., ratio between answered and issued notifications, of less than 10 %. After filtering, a set of 79 users and 6581 issued notifications remained.

About two thirds of the 79 contributing users had an American time zone (GMT/-4 to GMT/-9). Further, 73 users (92.40 %) had set an English locale. Of the 6581 total notifications 1508 were answered and 5073 were not answered, which results into an overall answering rate of 22.91 %. On average, each user was confronted with 83.3 notifications (SD 108.30, median 36.0), of which 19.09 (SD 23.09, median 8.0) were answered. These results indicate that almost every fourth notification is answered. The application was used between 1.01 and 76.63 days, which is about 11.01 days (SD 14.87, median 4.74) on average.

4.3.1.1 Time of Day

We were able to collect equally distributed data for the time of day (skewness $s = 0.42$, excess kurtosis $k = -0.72$). In the following, we investigate if the time of day has an effect on the notification answering behaviour. We found that only 0.08 % of the notifications are answered at around 04:24, and 0.31 % are answered at 22:18 (see Figure 4.5). In relation to all sampled notifications this answering behaviour is skewed to the right, i.e., towards the evening ($s = -0.61$, $k = -0.68$). The mean time for answered notifications is 14:16 (SD 06:36, median 15:06), the time for unanswered notifications is 12:12 (SD 07:00, median 12:22). A Welch-adapted two-tailed Student t-test indicates that this difference is significant ($p < 0.01$). We interpret these statistics in a way that people mostly attend notifications in the later hours of the day.

4.3.1.2 Location Provider

The location provider can give insights whether the user is situated indoors or outdoors, as a GPS location is most likely to be inferred only in outdoor situations. Consequently, we split our data into an indoor data set (no location information and network location) and an outdoor data set (GPS locations only). We observed

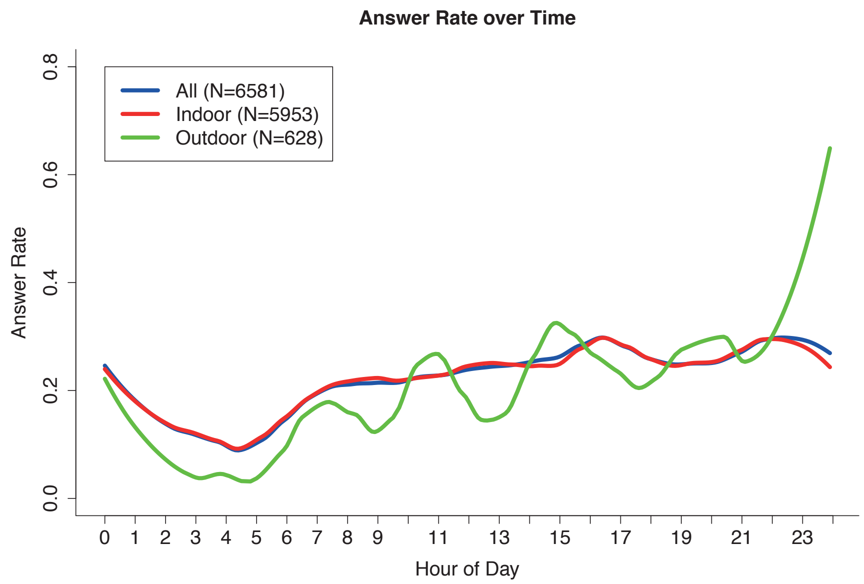


Figure 4.5: Despite a major drop to 8 % at night times the answer rate is more or less increasing constantly, ending in a global maximum of about 30 % in the evening hours. For outdoor notifications we further observed some local minima, which can most likely be credited to commuting.

5953 (90.46 %) indoor and 628 outdoor notifications. For both settings the distribution of observed notifications looked similar, i.e., less notifications at night (see Figure 4.5). However, for outdoor notifications we identified noticeable local minima throughout the day, which probably can be credited to commutes (09:00 and 17:30) or going for lunch (13:00). 1352 (22.71 %) of the indoor notifications and 156 (24.84 %) of the outdoor notifications were answered. A Chi-squared test indicated that there is no significant difference ($\chi^2 = 1.12$, $p = 0.29$). Thus, we cannot argue that outdoor notifications were answered more likely than indoor notifications or vice versa.

4.3.1.3 Position Accuracy

Similar to location, the position accuracy can indicate whether a user was located outdoors or indoors when attending or ignoring a notification. The average position accuracy for answered notification is 851.02 m (SD 1174.14 m, median 96 m), whereby it is 1017.41 m (SD 1219.38 m, median 672 m) for ignored notifications. This difference is significant ($p < 0.01$).

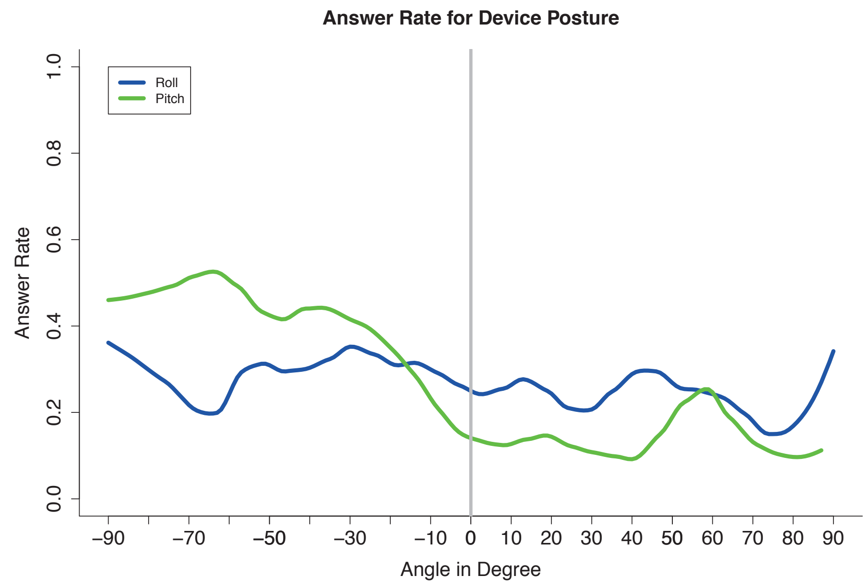


Figure 4.6: About every second notification is answered if the device is pitched towards the user for about $60°$. $50\,\%$ *of all answered notifications have a pitch angle between* $-0.50°$ *and* $-55.93°$.

4.3.1.4 Device Posture

Roll and pitch angles, which we both recorded for each notification, can be used to estimate the device posture. Roll represents the left/right tilt of the phone, and pitch indicates the up/down tilt. The posture could give indications on how the phone was located, e.g., lying flat on a table, when a notification was answered. The average pitch angle for answered notifications was $-31.82°$ (SD $53.58°$, median $-28°$), while it was $-10.89°$ (SD $67.51°$, median $-0.39°$) for ignored notifications. A t-test showed that this difference is significant ($p < 0.01$). The roll angle for answered notifications was $0.31°$ (SD $22.42°$, median $0°$), for unanswered $2.20°$ (SD $22.28°$, median $0.02°$). This difference is also significant ($p < 0.01$).

A visualisation of the answer rate over roll and pitch (see Figure 4.6) showed that the answer rate is about $26.46\,\%$ for almost any roll angle. However, for pitch a major peak at around $-64°$ could be observed, where $52.58\,\%$ of all notifications were answered. That means that more than every second notification is answered if the device is tilted towards the user by about $60°$. Interestingly, these pitch angles are representatives for how a device is typically held in a user's hand [HPSH00, HFG+98].

4.3.1.5 Proximity

The proximity sensor indicates whether the display is covered by something or not, which is typically used to avoid unintended touch interactions, e.g., during a phone call. However, it can also be used to understand where a phone is probably located, e.g., the sensor is covered in a pocket and not covered in the users hand. 1332 (88.26 %) of the answered and 3816 (75.22 %) of the unanswered notifications came with the information that the display was not covered. A Chi-square test indicated that there is a relation between the facts that a notification was answered and that the proximity sensor was covered ($\chi^2 = 116.48$, $p < 0.01$). Thus, the display was more likely to not be covered for answered notifications.

4.3.1.6 Heading, Speed, Light Level

The headings for answered notifications were 54.46° (GPS, SD 102.60°) and 176.89° (compass, SD 111.82°) and for ignored notifications 55.82° (GPS, SD 98.42°) and 170.65° (compass, SD 107.67°). The speed had a notable variation, but a median of $0.00\,\mathrm{m\,s^{-1}}$ for answered and unattended notifications. The median light level for answered notification was 4.09 lx and for ignored notifications 4.00 lx. None of these differences were significant, so we cannot conclude anything.

4.3.2 Data Mining

Beside the statistical investigation of individual measures we processed the obtained data in Weka, a data mining tool. We wanted to create a classifier, which ideally will be able to predict, based on sensor data, if it is likely that a user will answer an inquiry or not.

We investigated different classification algorithms, which we provided with nine attributes to build the model, i.e., location provider, position accuracy, speed, roll, pitch, proximity, time, light level, and whether the request was answered. The full set of 6581 notifications was used to train the classifiers and we used a ten-fold cross-validation procedure for evaluation. We used a trivial classifier as baseline, which predicted that each element belongs to the largest class in the sample, i.e., false, and thereby reached an accuracy of 77.08 %.

Our empirical investigation showed that the tree-based C4.5 classifier (see Figure 4.7) performed best and classified 5114 of the given 6581 notifications (77.85 %) correctly. The precision of the classifier was 0.74 and recall was 0.78. We got 257 true positives, 207 false positives, 1251 false negatives, and 4866 true negatives. This showed that particularly situations where the user would actually answer a notification were misclassified.

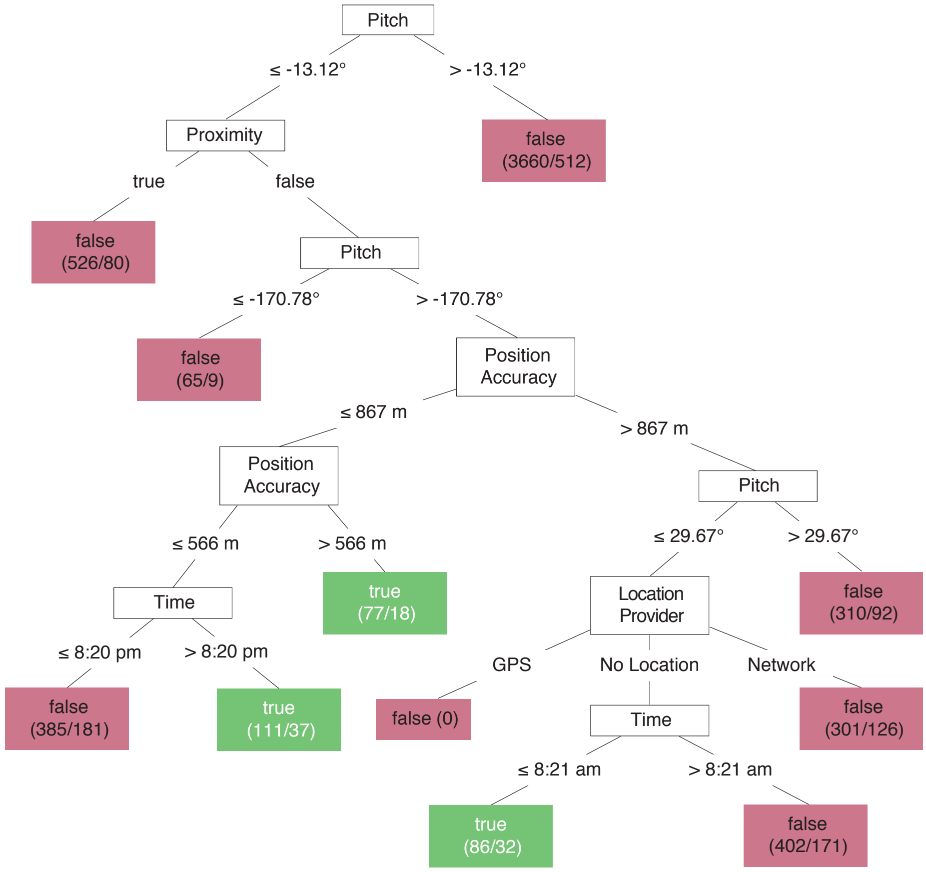


Figure 4.7: The resulting C4.5 decision tree consists of 20 elements and 11 leaves. It allows to classify whether a notification should be issued (true) or not (false) with an accuracy of 77.85 %.

Although the accuracies of the baseline classifier and the C4.5 classifier look quite similar, the C4.5 classifier has the overall better performance. While the baseline classifier is unable to predict a single opportune moment, the C4.5 classifier predicts 257 opportune moments accurately. The calculated Kappa statistic of 0.17 and the receiver operating characteristic, area under curve (ROC AUC), of 0.72 indicate that this classifier is not performing outstandingly well, but is definitely better than the baseline condition.

Alternative classifiers didn't perform better: a JRip classifier reached 77.48 % (precision 0.74, recall 0.77), and a neural network reached 76.84 % (precision 0.73, recall 0.77) in a similar evaluation procedure. The overall accuracy of all classifiers is comparable to existing classifiers for office settings [FHA+05].

4.3.3 Discussion

In the following we discuss how these results should be interpreted and what are the practical and theoretical limitations of the identified model. Again, we want to emphasise that our set of sensors is unable to measure all aspects that make a mobile situation. Consequently, our findings are only valid for this simplified view on the context.

4.3.3.1 Trigger Notifications at the Right Time

In the statistical analysis we could observe that the average time of day for answered notifications, 14:16, is significantly later than for ignored notifications, 12:12. Further, the answering rate decreases significantly during night time. Two leaves of the decision tree, which classify a notification likely to be answered, have a time node as predecessor. Thus, time seems to be a good predictor.

A combination of both, statistics and decision tree, implies that the ideal time to trigger a notification is before 08:21, after 20:20, but not during night time. These times appear logical if we compare them with a typical 9-to-5 working day. Notifications are likely to be answered after getting up, during breakfast or during the commute in the morning. Further, they are likely to be answered when the user is back at home, and the working day is over.

4.3.3.2 Trigger Notifications When Users Hold the Phone in Their Hands

The identified average pitch angle for answered notifications indicates that the users often already held the phone in their hands when answering a notification. Further, the proximity sensor is significantly less often covered for answered notifications. The derived decision tree is able to reach a high accuracy with pitch and proximity as two important identifying measures. Particularly the fact that 3660 (55.61 %) notifications with a pitch angle greater than $-13.12°$ are classified as unlikely to be answered is particularly remarkable (see Figure 4.6). We interpret these measures, i.e., pitch and proximity, as an indicator that notifications are more likely to be answered if the users already hold the phone in their hands. This is the most relevant aspect to predict whether a notification will be answered or not.

4.3.3.3 Relevance and Applicability of the Model in Practice

From a technical perspective, the derived model can be transferred into simple comparisons, which can be executed on today's mobile phones. This makes an application feasible and probable. The presented model gives a basic recommendation in which situations a notification is likely to be answered and in which not. As explained, the approach is limited by the fact that we just measure and include a tiny fragment of the overall context. The identified performance measures indicate that the predictions of the simplified model are only a medium improvement over random predictions. That means that still a considerable number of notifications will probably occur in inopportune moments. Consequently, we think it will be beneficial to further improve the model before application.

We think that our model can serve as a starting point for less simplified models, which incorporate more complex relationships and constraints between individual predictors. These advanced models will probably lead to an increased prediction accuracy and, eventually, to a transferability among use-cases and an applicability of the approach in everyday life. In addition, given how different a daily routine can be among users, we think it would be beneficial to have user-specific models.

Further, also notifications have different characteristics that can be considered. One example is the relevance of a notification for the user. For particularly relevant, important or time-critical notifications the model, as presented here, is hardly applicable. To make it applicable, much more information about the users and their tasks need to be collected and incorporated. Further, the model needs to be integrated into a complex notification management framework, like sketched by Iqbal et al. [IB08], that is responsible to issue a notification at opportune moments, but follows a complex set of dynamic constraints.

We think that the resulting model can also be applied and used beyond mediated interruption techniques. For example, we can think about a negotiated interruption technique, where the user's choice if a notification should be handled right away or later can be used to re-fine the model at runtime. This approach would allow the model to adapt to changing user contexts, e.g., if the user starts a new job or moves to a new city and then follows other daily routines.

4.3.3.4 Limitations

One limitation for large-scale studies via mobile applications is the data validity. Although we did sanity checks, erroneous data samples might still affect the analysis. Further, we cannot say anything about the users' motivation to answer or ignore/reject notifications. Thus, it could be the case that a user ignores a notification although it was issued in an actual opportune moment, thereby influencing our findings. However, we argue that this is regularly not the case and for the majority of the recorded data the user is behaving as initially expected.

4.4 Summary

In this work we have studied in which contexts a mobile notification is attended and in which it is rejected or ignored. We did so by publishing the mood-tracking application MoodDiary in the Google Play store, where we collected 6581 notifications from 79 users over periods of up to 76 days. This approach gives our studies the advantage that they are done in a real, longitudinal, and large-scale setting, which is beneficial for the results' validity and which is something that has not been done before in interruption research. The presented approach was demonstrated in the mood monitoring domain, but does not consider the actually reported mood information. Thus, it can be applied in any other domain, where sensor information of similar kind is available.

We analysed the notifications from a statistical perspective and did a data mining analysis. We identified that notifications should be triggered at the right time and when the device is already at hand. We illustrate that our derived classifier can be applied already, but a surrounding notification management framework and less simplified models would be helpful to achieve a major practical value in day-to-day use. We envision that the applied model can lead to a substantial decrease of notifications in inopportune moments, which therefore reduces the number of annoying and unpleasant interruptions at no costs. This eventually leads to fewer context switches and therefore to less stress, frustration, time pressure and effort [MGK08].

With respect to self-reporting user observation techniques, which make use of notifications, our findings will lead to more targeted, less obtrusive and consequently less missed inquiries. We argue that an intelligent notification management framework will likely lead to an increase in the number of gathered user experience insights. Further, we expect that the provided insights will increase in their richness and quality. These improvements will allow researchers to conduct much more detailed and fine-granular analysis, revealing a deeper understanding of the users, their wishes, desires and problems.

Until now, we only discussed the potential attendance of a notification as model criteria. However, a future notification management system for self-reporting techniques could also use several other criteria to define whether a notification should be issued or not. For example, researchers might be interested to sample user experiences in a situation where the user is probably unable to attend a notification, e.g., to record stress-related measures. To do this, inverting the proposed model would be a good starting point. However, future work could also consider other not directly measurable, but predictable parameters, such as frustration, happiness, or tiredness.

Future research should investigate the actual application of the derived generic model and compare it against user-specific models. Further, non-sensor information, like phone activities or calendar entries, should be incorporated in future model revisions. Eventually, the model should be embedded into a holistic notification management framework to study how inquiries of various importance or type could be managed and issued.

5 Enhancing the Situatedness of Diary Studies with Storytelling

A diary typically is a paper-based booklet and allows users to record everyday impressions of various kinds and nature. Typically, a diary is not focused on any specific aspects to be recorded. Moreover, the users decide on their own which aspects go into the diary and which not. There are both, open diaries, which just provide plain and empty space to write on, and diaries with pre-defined questions, e.g., about today's most positive/negative experience. It is up to the users if and when they want to write something in the diary, there are no reminders or defined intervals in which an entry has to be made. There are different motivations to keep a diary, although all have in common that users want to remember their experiences at a later point in time.

Although diaries were not created as an observation tool for studies, they can be applied in study contexts as well. Diary studies are user observations which are done through a diary. They have a long-lasting history [Rie93] and are widely applied in HCI research, e.g., [CM05, Hyl06]. Given their nature they are mostly used for longitudinal user observations, lasting a week or longer, and to collect qualitative insights. Diary studies can be used to understand the daily routine of a user or to evaluate the day-to-day experiences with a novel user interface. Although a diary itself can be valuable for a researcher, diaries are often used as a starting point for post-hoc discussions, where the recorded material can support the participants memories [LG06, SFA+07].

Unfortunately, such diary studies come with a challenging disadvantage: a limited **situatedness**. Most frequently, people simply forget to create a diary entry. Further, diary entries are often reflected on several days after they have been initially created. That makes it hard for the user to remember the context and motivation behind an entry and the potential interconnection to earlier or later diary entries. In fact, the entries lose their currentness and contextuality over time, which was definitely present during creation. From a researchers' perspective both problems eventually result into missing or partly distorted information, which are threats for the overall study outcomes and validity.

Literature essentially proposes two concepts to address these problems in similar domains and applications. One concept aims to increase the amount of information of an recorded event. This can be done by enriching an information with additional media, e.g., a text is accompanied by an image. Another concept tries to encourage users to re-use and process recorded information. Thus, users reflect on individual entries by interacting with them in certain ways, which arguably supports them in remembering important details.

Brandt et al. [BWK07] addressed the problem that people forget about creating entries, probably because of the burden to create these in situ. They proposed to record small snippets in situ, which serve as prompts for participants when completing full and complex diary entries at a convenient time. Because they used first-generation smart phones when they conducted the study, their research was mostly focusing on text entries as reminders for a later extension.

Carter and Mankoff [CM05] investigated the role of different media in diary studies and how they support memories in post-hoc settings. They analysed three diary studies which they conducted, whereby they discuss how each media type can support participants' recall process. They found that photos are very supportive to recognise the *who* and *where* of a situation. They further found that the timing and sequencing of events is very important for activity reconstruction. Based on their insights they designed and proposed an ideal diary study pipeline, which foresees to handle and modify diary entries after their initial creation.

Bonsignore [Bon11] presented a study on StoryKit, which investigates how storytelling is used in practice to create and share stories. She was further interested how created stories might inform the design of mobile storytelling applications in general. They found that a multi-modal input mechanism and a simple mobile interface with a storybook template enable people to create expressive and vibrant stories. They further observed that their studied prototype already resulted into a surprising volume and variety of different stories.

Both concepts, i.e., enrichment of information and re-use of information, were investigated in related work. Altogether, most findings indicate that each concept is promising and could result into a notable increase of situatedness. Nevertheless, none of the related work studied the combination of both approaches. Further, there are no insights how these concepts could improve a diary study technique in a dynamic field study setting.

In this chapter we present a mobile diary application, Storyteller, that takes up and combines both concepts from the related work. Storyteller allows a quick, in situ creation of multimedia entries. Further, Storyteller allows to compile individual multimedia elements to a holistic story using the storytelling approach. Thus, users are able to reflect on individual diary entries at an early stage and combine them to a holistic story. This story is then able to represent and convey the overall user experience and feelings beyond individual entries. We argue that this would lead to more and better qualitative insights and thereby overcomes the two most crucial problems of traditional diary studies.

This chapter makes an incremental contribution, which supports and extends existing knowledge about diary- and storytelling-based observation techniques. To our knowledge, we are the first who combine the diary study and in situ reflection with the idea of storytelling. We do our research in a realistic, midterm field study setting and we are using the most recent smartphone generation, which provides various new ways to capture and deal with recorded media. Some of our observations contradict existing findings, which underlines that we make a significant contribution.

This chapter is structured as follows. In Section 5.1, we explain the method approach, i.e., how exactly the method works and how we think it will change future diary studies. Then, we present a conducted requirements study, which we did to collect ideas how a concrete realisation of the method should look like from a user's perspective (see Section 5.2). Based on these requirements, we designed and developed a tool, Storyteller, which gives a reference implementation for our proposed methodology and is described in Section 5.3. We studied our tool and the underlying method approach in field study, which we report in Section 5.4. We close this chapter with a summary, concluding remarks and ideas for future work.

Parts of this work were published in Poppinga, B.; Oehmcke, S.; Heuten, W.; Boll, S. Storyteller: In-Situ Reflection on Study Experiences. Proceedings of MobileHCI, 2013, Munich, Germany.

5.1 Method Approach

The approach of this method is two-fold and aims to increase the situatedness of diary studies. First, our approach allows a quick capture of rich multimedia material, which sum up to highly dense individual diary entries. Second, the recorded material can be compiled to a coherent story, making inherent relations between individual entries present and easy to remember.

The improvement of in situ data through input, processing, and output modalities of a mobile phone is one of the two key approaches of this thesis. The first part of this method approach is in line with the overall approach and aims to improve the in situ capturing of diary entries. Instead of a single and limited diary entry, e.g., a text note, as they are typically recorded in traditional diary studies, we allow users to record rich multimedia diary entries. Further, and as an extension to related work, we provide users with a quick mean to record individual entries, making the recording less complicated and as timely as possible.

The second part of this method approach is the introduction of storytelling as a means to connect and relate individual diary entries. Already during the capturing of material, various contextual information, e.g., location, is recorded implicitly. Further, the quick creation mechanism encourages users to extend recorded single elements with other, additional information of a different format, e.g., an image is accompanied by a textual explanation. Through the storytelling concept we aim to encourage users to combine and relate these individual fragments to holistic, comprehensive stories. An authoring tool, which implements the storytelling approach, is necessary to realise this idea. On a conceptual level, the authoring tool will be accessible in a similar quick and easy way, encouraging users to engage with recorded material ideally at an early stage of the study. We argue that this timely engagement with material will persist many of the contextual information in a story, making them available to participants as well as to experimenters in post-hoc settings.

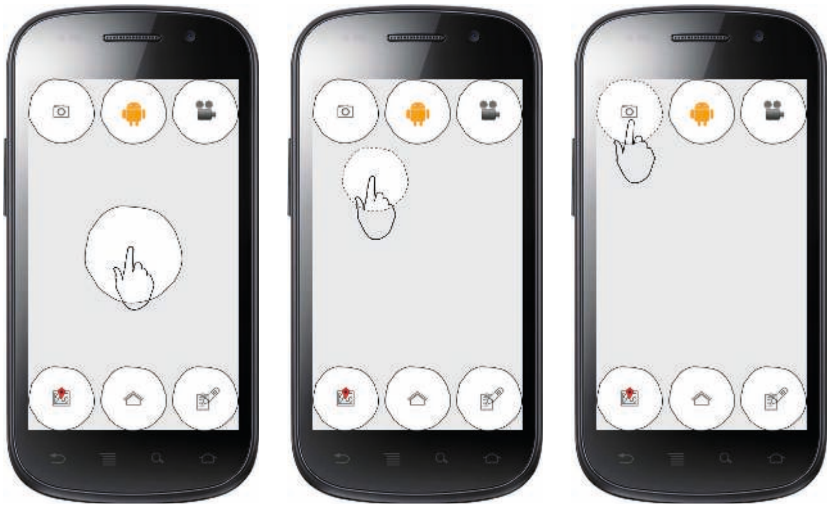
We argue that, altogether, this will lead to more recorded situation details and a clearer relationship between those. Users will remember these relationships and contextual details easily and will be able to communicate these to bystanders and experimenters. Eventually, this will allow an observer to get a detailed understanding of the overall situation, leading to an increased situatedness over traditional realisations of the diary technique.

5.2 Requirements Study

The design of the Storyteller is inspired by literature and by insights and inspirations we got from a small requirements study. The purpose of the requirements study was to find out what are the important elements that a participant wants to capture, and with which methodology they would like to do so. Further, we wanted to find out how they report on earlier made impressions, e.g., whether they report in chronological order or report by importance.

In our study we asked 5 participants to walk for about 30-60 minutes to a certain, pre-defined place and return afterwards. We asked them to take notes on their way on *what* would be interesting elements to record for a diary and *how* they would like to record these. When they returned we asked them to tell what they've experienced. Then we discussed about their impressions *which* and *how* things could be recorded. Participants were allowed to use various coloured pens, paper and scissors to better express themselves. After they finished their report, we asked them to assess a set of recording techniques, like photographs, notes, or voice memos, and estimate the perceived usefulness of each technique. If possible, these techniques should be prioritised for the later implementation.

We found that all participants used a chronological order to tell their story. Some participants used a map to put their impressions in an order. There was a consensus that photographs are the most suitable media to record most elements of relevance. It was emphasised that photos are fast and easy to take, nevertheless they still provide a lot of insights. Video recordings were rated second place, whereby they agreed that most of the made videos would be quite short. Textual notes are the third most important media, whereby some participants commented that longer text will be most likely entered afterwards, i.e., when the diary entries are compiled to a story. However, participants agreed that shorter notes would



(a) Step 1: Select centred icon.
(b) Step 2: Drag bubble towards desired functionality.
(c) Step 3: Release when finger is above desired functionality.

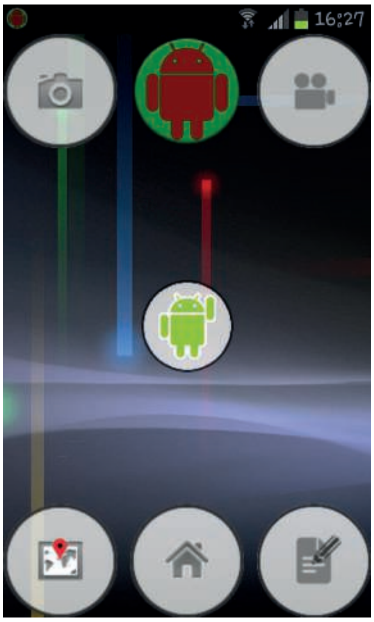
Figure 5.1: A custom lock-screen allows the users to quickly capture various multimedia elements, such as photographs or notes. In contrast to traditional diary applications this reduces the burden to create a new diary element.

be helpful to enrich individual photographs or videos, e.g., to point out the most relevant aspects.

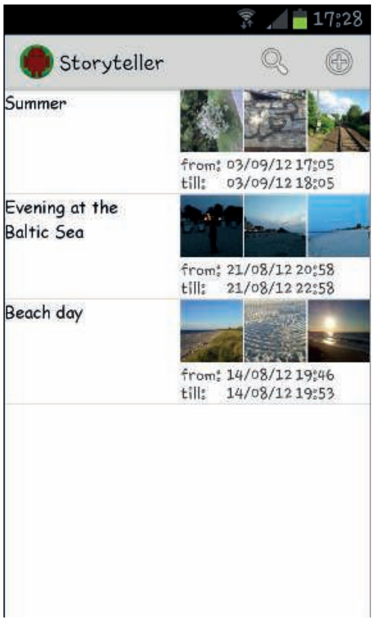
5.3 Observation Tools and Apparatus

Based on our findings from the requirements study we decided to focus on a quick capture of photographs, videos, notes, and the user location. Our conceptual unlock pattern is similar to established patterns for lock-screens. A user has to drag an icon, localised in the centre of the screen, over a certain distance and release it afterwards to either unlock the phone or trigger a certain functionality, e.g., taking a photo (see Figure 5.1). All available actions are located at the border of the screen to prevent from accidentally unlocking the phone or triggering any other action. With this replaced lock-screen a user can get the phone out of his/her pocket and, e.g., take a note or take a photo instantaneously.

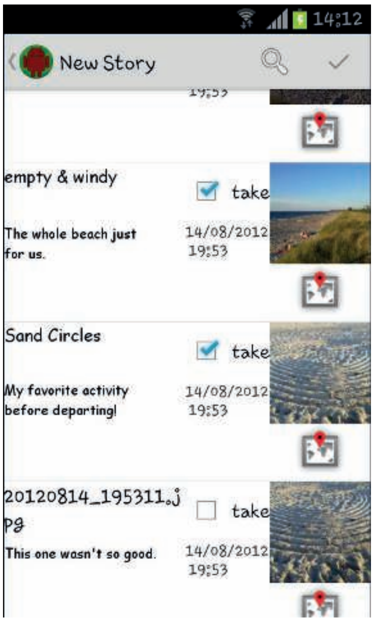
The interviews made clear that both, a chronological and map-based representation of the diary entries, should be available. It further should be possible to add and remove elements from a story, and change the order of the entries. We prototyped and refined the application several times, before we reached consensus about the final design.



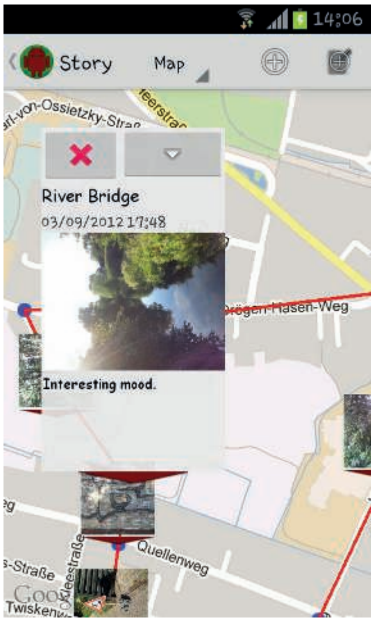
(a) Diary entries can be created straight from the lockscreen.



(b) Storyteller can be used to manage existing stories.



(c) New stories can be created by selecting individual diary entries.



(d) Stories can be visualised in a list or on a map.

Figure 5.2: Storyteller is a mobile application which allows convenient capturing of diary entries. Further, these entries can be combined to a holistic story, which we argue can better represent and convey the feelings and impressions beyond a single diary entry.

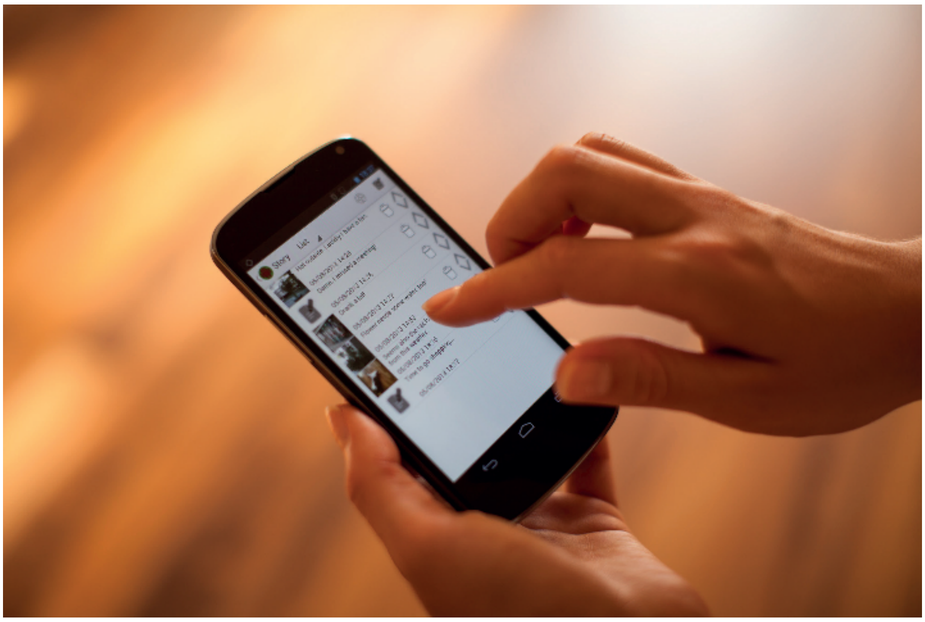


Figure 5.3: Drag and drop allows users of Storyteller to create stories of chronological or thematic orders.

The final Storyteller application runs on Android and consists of 5 different views, shown in Figure 5.2. We realised the illustrated unlock concept in a custom-made lock-screen (see Figure 5.2(a)), which allows the users to quickly capture various multimedia elements, such as photos or videos. Further, most of the other conceptual ideas were implemented. Figure 5.2(b) shows the application's entrance point, which contains an overview of all available stories. In addition, this view allows the user to trigger the creation of a new story.

The activity shown in Figure 5.2(c) allows then allows a user to select the elements, which are of relevance for the to be created story. All stories can be visualised as a list or on a map (see Figure 5.2(d)). By default all elements are arranged in a chronological order, although this order can easily be modified through drag and drop (see Figure 5.3). All stories are stored persistent as XML. Consequently, they can be restored if the application is closed. Further, this theoretically allows to develop a Storyteller variant, which allows to edit and compile stories on a Desktop computer.

5.4 Evaluation

In this initial user study we want to investigate the general uptake of the Storyteller approach. We are curious what the advantages and disadvantages of the overall technique are and what might be implications for future design iterations and studies.

5.4.1 Method

Our evaluation was designed as within-subjects experiment with two conditions. Storyteller, as described earlier, served as the experimental condition. For the control condition we relied on a set of pre-existing Android applications, i.e., the default camera application, Google Maps, and a note-taking application, which altogether cover similar functionality, but don't provide the option to compile diary entries to a story. The modified lock-screen was not present in the control condition.

10 volunteers participated in our study, of which 7 were male and 3 were female. The participants were aged between 21 and 33 years. We calculated an average age of 23.6 years (SD 3.23). Most of the participants were students from a local university and haven't had any prior experience with the Storyteller application so far.

Since we wanted to test both applications in a realistic and dynamic setting, we decided to test them in a mid-term shopping scenario. Therefore, we agreed on two dates before the actual study, where the participants wanted to go shopping anyway. Conditions were counter-balanced and assigned to one day. On the first day, we welcomed the participant, clarified on the motivation and task of the study, and got the informed consent signed. Then we made each participant familiar with the diary tool they were allowed to use on that day. The actual task was—unsurprising for a diary study—to capture everything interesting during the shopping trip where they would like to tell us about later. Before the participants left, we agreed on a place where we planned to meet again for the post-hoc interview, e.g., a local café. Participants were not accompanied during the shopping experience. On the second day, the procedure was similar except that we didn't asked for another informed consent.

When the participants returned from their shopping trip, we gave them as much time as needed to sort their taken diary entries and to prepare for the post-hoc interview. In the experimental condition this was the last chance to combine entries to a story, if this wasn't already done during the actual experience. Eventually, we asked the participants to tell about their experiences with the help of created diary entries. The experimenter measured how long it took each participant to tell their story. After the storytelling was finished we asked the participant a selection of questions on the methodology, e.g., if the participant liked to tell the story and if it was perceived as demanding, if they felt supported in creating a diary entry, and whether important functionality was missing. In addition we asked each participant to rate their experience on a five-point Likert

Figure 5.4: A study participant is using the Storyteller at the bus stop to compile individual diary entries to a holistic story.

scale. Further, we asked each participant to fill a System-Usability-Scale (SUS). Ultimately, we thanked for the participation in our study and, if needed, agreed on a date for the second trial. None of the participant was paid.

5.4.2 Results

8 participants went shopping in the city centre for both conditions (see Figure 5.4). 1 participant once decided to go to the city centre and once decided to go to a supermarket. Another participant went to a supermarket for both conditions. With the Storyteller the shopping took between 60 and 150 minutes (mean 82.5, SD 30.21). In the control condition participants returned after a period of 60 to 180 minutes (mean 85.5, SD 40.65). That means that the shopping experiences took similarly long with both conditions.

In the Storyteller condition, 2 participants already had started to create a story in situ and used the provided time to extend these. 8 participants created their stories after they returned from the shopping and before the actual interview. In the control condition, none of the participants had already started to prepare themselves for the interview. All used the time given immediately before the interview to prepare.

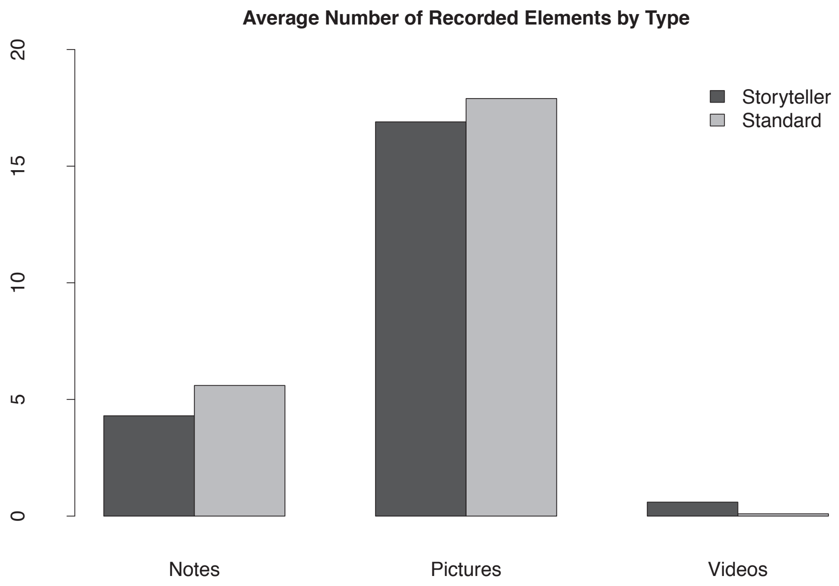


Figure 5.5: No significant differences could be found regarding the average amount of captured elements, i.e., notes, photos, and videos.

In total, each participant recorded 21.7 (SD 8.08) elements with the Storyteller and 23.6 (SD 20.62) elements in the control condition (see Figure 5.5). We observed that in both conditions, photos were captured more than any other elements. In average a participant took 16.9 (SD 8.07) photographs with the Storyteller and 17.9 (SD 17.86) in the control condition. Thereby, many photographs depict objects or visited places (see Figure 5.6). Second placed were textual notes, of which 4.3 (SD 3.69) have been created in the experimental condition and 5.6 (SD 5.99) in the control condition. Videos were ranked last. 0.6 (SD 1.2) videos were recorded with the Storyteller and 0.1 (SD 0.3) in the control condition. None of these differences is significant ($p = 0.67$ for photos, $p = 0.27$ for videos, $p = 0.22$ for notes). The Storyteller further provided the option to annotate a photo or video. Participants annotated 4 (SD 3.4) photos and 0.1 (SD 0.4) videos in average. These options were not provided in the control condition.

For the Storyteller we recorded which of the elements were added to the story, either in situ or during the preparation phase. 13.3 (79 %) of the photos, 3.9 (91 %) of the recorded notes, and 0.4 (67 %) of the videos made it to the final story. We measured how long the participants told us about their shopping experience with both conditions. We found that participants reported 5:53 (SD 2:43) minutes with the Storyteller and 7:12 (SD 5:42) minutes in the control condition. A Student's t-test showed that this difference is not significant ($p = 0.32$).

Figure 5.6: Many participants captured objects, which they found to be interesting or which in some way contributed to their overall experience.

9 participants reported on their experiences in a chronological order in both conditions.

Participants assessed different aspects of the individual conditions with the help of 5-point Likert scales. Student t-tests showed that Storyteller was rated significantly better than the control condition in every aspect (see Figure 5.7). Further, we asked them to elaborate on each of the asked aspects and give reasons for their rating. The participants agreed (mean 4.2, SD 0.6) to the statement that they feel supported in recording media in the Storyteller condition, and rated neutral (mean 2.7, SD 1.19) for the control condition ($p < 0.01$). Nine participants stated that the Storyteller's lock-screen allows a quick and easy capture of new media. Three participants liked that all important features were in one place, which leads to less effort.

The participants agreed (mean 4.4, SD 0.66) that they liked to tell the story with the Storyteller, and rated neutral (mean 2.8, SD 0.98) for the control condition ($p < 0.01$). Further, they strongly disagreed (mean 1.3, SD 0.9) that telling the story was exhausting with the Storyteller, whereby they rated neutral (mean 2.8, SD 1.4) for the control condition ($p < 0.05$). In the interview, the participants were particularly negative about the missing links between individual diary elements, both regarding time and location. That also lead participants to check some elements again to make sure not to forget about anything.

Ultimately, participants mostly disagreed (mean 1.5, SD 0.67) that any features were missing for the Storyteller. However, they agreed (mean 4.1, SD 0.94) that there were features missing in the control condition ($p < 0.001$). The most frequently asked for feature of the Storyteller was to allow a textual annotation of photos right after they've been taken. We asked the participants to fill a System Usability Scale (SUS) for each system. We found that participants rated the Storyteller (mean 85.5, SD 10.11) higher than the control condition (mean 59.75, SD 12.07, see Figure 5.8). This difference is statistically significant ($p < 0.001$).

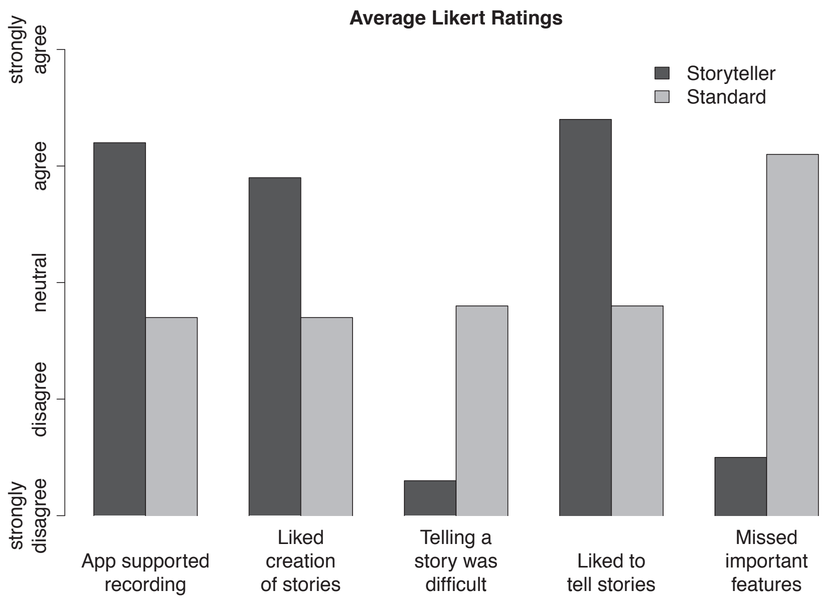


Figure 5.7: Participants rated the Storyteller significantly better than the control condition in all aspects. With Storyteller creating and telling a story was a much more pleasurable experience.

5.4.3 Discussion

Our study results are unable to support one of our assumptions, i.e., that more elements will be captured. In fact we were unable to find a significant difference in the captured elements between the two investigated conditions. However, we found that participants rated the capturing of the Storyteller to be significantly better, allowing a quick and easy capture of media. From a quantitative perspective we are also unable to support our second assumption, i.e., that participants reflect on recordings in situ and therefore more qualitative insights of increased currentness are revealed afterwards. In fact, the reported stories showed a roughly similar duration and no significant differences could be found. Nevertheless, we got the impression that participants were less interrupted and less often needed to collect themselves with the Storyteller application. We see this supported through the positive observations we made from the Likert scales.

In contrast to earlier work by Brandt et al. [BWK07], we found photographs to be more important and more frequently used than text. We think this interesting change can be credited to the technological development. While it took comparably long to take a photo on 2007's feature phones and the quality was mostly insufficient, nowadays smart phones can capture a photo within milliseconds and

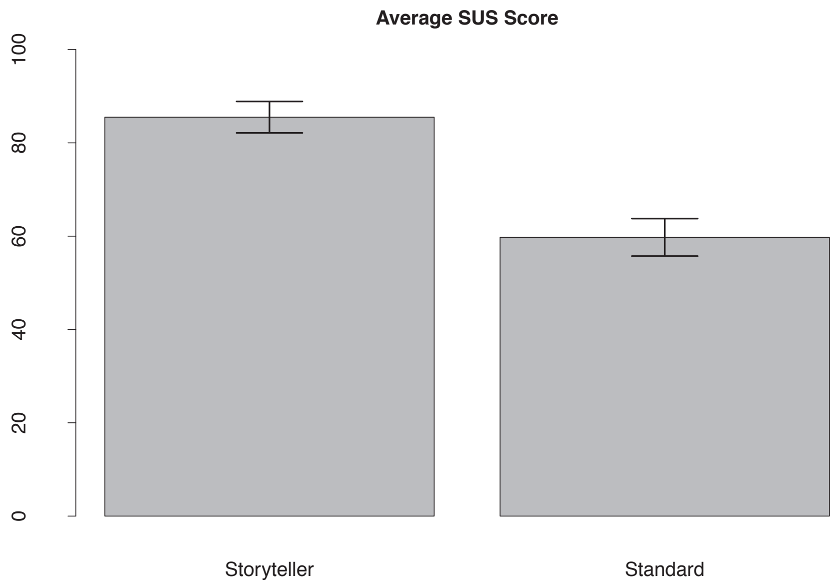


Figure 5.8: The System Usability Scale (SUS) scores indicate that participants rate the Storyteller application better than the standard system. This difference is significant. Error bars indicate the standard error.

the quality is very good. Nevertheless, also in our study textual notes were still used quite frequently.

Carter and Mankoff [CM05] report that their participants rarely used a chronological order to reflect on their diary entries. Interestingly, we observed the opposite, i.e., 9 of 10 participants reported their experiences in a chronological way. We think that this is the case because participants interpret the list or map elements with respect to time. To also allow non-chronological stories, other representations of elements, like stacks or folders, should be implemented in future diary study and storytelling tools.

In retrospect, we think that the biggest limitation of the study is that the shopping scenario wasn't time-consuming enough. While we were able to identify several benefits of the Storyteller approach, we believe that the underlying mixture of a diary study and storytelling will provide a researcher with even better results in a long-term study.

Figure 5.9: All participants used and appreciated the Storyteller to prepare and guide them through the post-hoc interview. Future work should study how a cloud diary could improve the storytelling process, e.g., by allowing an experimenter to analyse data in advance (as illustrated with the laptop).

5.5 Summary

Insights from diary studies typically have a limited situatedness, because study participants don't create sufficiently many diary entries and the resulting diary itself is unable to capture and echo all the contextual information. In this chapter, we present a method approach, which modifies a diary study in two ways: we provide participants with a quick means to capture diary entries, and we allow the combination of diary entries to a holistic story, using the storytelling approach. We describe how we studied the requirements, and how we subsequently designed the Storyteller diary application. We further report from an initial investigation of the novel approach in a mid-term field study.

Our requirements study showed that all participants used a chronological order to report on their experiences. There was a consensus that photographs are the most suitable and appropriate media to record most experiences. Further, participants would appreciate to record videos and text notes. It was envisioned that text notes would be particularly helpful to enhance other media types. The Storyteller application replaces the Android lock screen, but mimics the common slide-to-unlock pattern to trigger custom functions, e.g., to take a photo. This allows participants to quickly capture elements of interest when in the field.

Further, Storyteller enables participants to combine elements to holistic stories, which can be presented in chronological order and on a map.

The initial user study investigates the general uptake as well as advantages and disadvantages of the methodology. We found that participants liked our quick and easy mechanism to capture new diary entries, although not significantly more entries are created compared to a traditional diary approach. Participants liked to interlink diary elements with each other to create a holistic story. They agreed that they appreciate to tell the story with the help of the Storyteller and felt well supported. Interestingly, we observed that our findings differ from the literature in some aspects, which can be mostly credited to the advanced technology we used.

Overall, we conclude that the idea to combine a diary study with storytelling is very promising. Although we were unable to show significant effects in our initial field study, we think that Storyteller will result into more created diary entries in long-term studies. Further, we found clear indications that the storytelling approach eases the reporting on diary studies in several regards, which we argue will eventually lead to better qualitative insights. We conclude that the presented method approach definitively comes with several advantages, which have the potential to significantly improve the situatedness of future diary studies.

Future work should study the presented method approach in other contexts, and particularly in long-term, unsupervised field studies. Further, other ways to represent a story should be studied, e.g., a hierarchical order of diary entries. In addition, we can imagine that a cloud-based diary, which can be extended and maintained from various mobile devices, is a promising variation for future studies (see Figure 5.9). This would allow participants to maintain their diaries in a more comfortable way. Further, experimenters would be able to study and analyse the diary data before an on-site interview, and prepare questions on the available material.

Figure 6.1: Cycling as a leisure activity is often an intimate experience. Traditional observation techniques, like shadowing or diary studies, often fail in this setting as they are too obtrusive or request the cyclist to interrupt his or her experience.

6 Hybrid Observation: Using In Situ Material to Assist Post-Hoc Reflection

Most field studies use one or more observation techniques. In situ techniques allow the observation of participants during their experiences in the field, and post-hoc techniques capture participants' insights shortly after the actual experience. Nowadays studies use at least one of the two observation concepts, but typically apply a combination of both approaches.

However, both observation approaches, i.e., in situ and post-hoc, come with their limitations and problems. These problems have specific characteristics for individual observation techniques, which we initially introduced and referred to under the terms situatedness, scalability, and obtrusiveness. For some scenarios and settings these individual problems sum up to an extent, that none of the techniques could be valuably applied separately. A sequential combination of observation techniques is unable to compensate for individual differences, which in the end leads to scenarios which hardly can be studied at all.

We know from our own long lasting experience that cycling as a leisure activity during holidays (see Figure 6.1), is a representative field study scenario, which can hardly be investigated with traditional observation methods [PPB09a, PPB09b, PPHB12b]. Tourists on bicycles often cycle for several hours and try to forget about everyday routines and problems to experience ultimate relaxation and simply enjoy the landscapes and surroundings. In situ observation techniques, like shadowing, disturb the intimate feeling of cycling and can influence the participants in their natural behaviour. In addition, the permanent involvement of experimenters is costly and limits the scalability of the study. Other in situ techniques, like diary studies, don't rely on the experimenter, but require active and repetitive user feedback. Thus, they would require the cyclist to stop or interrupt the experience, which is not appropriate. Post-hoc study techniques, like interviews, involve the participant after the actual experience. Since the participants have to reflect on earlier experiences, their descriptions often lack details and situatedness as the participants tend to forget or do not report details. The community is aware of these methodological limitations and proposes to investigate better field observation methods [KG03, KCS+12, KP12].

To overcome the individual limitations of single observation techniques, hybrid observation approaches were proposed and applied, e.g., [ELF93, KSWK10]. The idea behind hybrid observation is to use recorded in situ data to drive post-hoc observation techniques. For example, the unobtrusive logging observation technique could be used to record in situ data. In a post-hoc setting the gathered in situ data is then used to provide objective insights and consequently jog the memories of the study participants. The idea is that this allows experimenters to do a more targeted observation and to obtain additional qualitative feedback.

There doesn't exist much work on hybrid observation techniques. Existing research mostly applied hybrid observation in lifelogging or in other repeating day-to-day scenarios, where study participants mostly dealt with established, reoccurring behaviours and routines. These everyday settings differ from a typical field study context, because people do their routine, perceive and deal with situations several times, and therefore inherently remember many details in a supported post-hoc session. Until now, the applicability and advantages of the approach in less familiar field study scenarios remain unclear. Most notably, it is unclear if a hybrid observation approach can provide sufficient **situatedness** when applied to a field study setting.

In the last few years, more and more studies emerged in the HCI community, which involve cyclists. It has been reported, that cycling has a distinctive, idiosyncratic and often spontaneous, explorative, but intimate nature [RFO+09, FWPM10, PPB09b]. To observe cyclists, logging [RFO+09, EML+10, RSD+10] and post-hoc questionnaires [RFO+09] have been used. Further, tourists and tourist guide applications have also been investigated for a long time now. The observation of tourists in the wild has been identified to be a challenge, as a study would clearly impinge on the leisure time of tourists [CDM+00]. Consequently, mostly post-hoc techniques are used to study tourists, whereby sometimes in situ material, like photographs [TFA10], is used to drive those. We conclude that the observation of cyclists and tourists is of interest, but there is a lack of an appropriate study technique.

Many traditional observation methods are inconvenient to use in situ, because they are obtrusive and often inferring with the participants' tasks. To better support qualitative in situ observations, there is ongoing research on methodologies. An overview on all relevant study techniques in the field of HCI has been presented by Hagen et al. [HRKS05]. The Experience Sampling Method (ESM), diary studies, and the Day Reconstruction Method (DRM) are among the mostly unobtrusive and situated observation techniques [FCC+07, MHK05]. The approach of ESM is to remind users to fill out a questionnaire on a regular basis [LC83]. Modern ESM implementations use a mobile phone to ask the participant for input, e.g., via a displayed questionnaire [FCC+07]. In diary studies, participants can freely control the timing, media, and means of a capture themselves, although it is known that there exists a burden to actually record an experience [BWK07]. In so-called elicitation studies, the diary content is used to prompt and maintain a post-hoc discussion with the participant [CM05]. DRM asks a participant to separate the previous day into scenes or episodes (like in a movie) and answer a set of questions for each scene. DRM is typically applied for broad studies, e.g., to determine how the user experience or feeling of well being changes over time [KKS+04].

Because many observation methods come with various severe limitations, there is an continuous need and effort to overcome these and come up with improved observation methods [KG03, KCS+12]. A popular and established approach is the clever combination of individual techniques to a hybrid observation technique [ELF93]. Nowadays hybrid observation techniques typically re-use in situ material to elicit post-hoc observation methods. There are several examples of techniques, which re-use videos [ELF93] or photographs [IKM02]. Brown et al. conducted a photo diary study to explore design requirements for new kinds of information capture devices [BSO00]. They used the participants' photographs in post-hoc interviews. Carter and Mankoff proposed an optimised media supported diary study approach, which they claim can reduce active user feedback to a minimum [CM05].

Most of these studies are dealing with information, which actively needed to be recorded by a participant, e.g., through photographs, voice or video. Most recently, life-logging devices like the SenseCam [HWB+06] allow a passive capturing. Most similar to our work, Kalnikaite et al. recorded GPS tracks and SenseCam images for a period of five weeks to see which data supports the human memory in a life-log scenario. They found that with a combination of both techniques, people remember most events. The SenseCam image in particular supports to remember details. Location information support inferential processes and allow a person to reconstruct, e.g., habits in their behaviour. They intentionally excluded weekends to only focus on typical weekday events. More recently, Gouveia and Karapanos presented a paper [GK13], how different cues, i.e., visual, location, temporal, and social cues, support memories and related emotions. In contrast to Kalnikaite et al. [KSWK10], they relied on quantitative user feedback and insights from an eye tracker to assess how each cue supports memories.

Related work showed that hybrid observation techniques work in long-term, day-to-day or workplace conditions [KSWK10, GK13]. Aspects like applicability, advantages and disadvantages in less predictable field study settings remain questionable. In this research, we aim to fill this gap and study hybrid observation technique in the cycling scenario, which we identified as representative non-everyday study setting. We give a detailed explanation on how we designed, implemented, and combined individual observation techniques and how we adapted the hybrid approach to this setting. We further describe our motivation behind the design of our reflection software, and discuss the role of each component from a user and experimenter perspective.

By doing so, we want to contribute to the understanding of the technique in non-everyday settings and field studies. In detail, we investigate which in situ data collection techniques are least obtrusive and disturbing, but provide sufficient insights for post-hoc settings. Further, we study if specific in situ techniques lead to discussions of certain aspects in post-hoc interviews. In addition, we investigate how the hybrid observation approach affects the relation and communication behaviour between experimenters and participants.

Altogether, this work is relevant for the continuous need to improve observation techniques for mobile scenarios [KCS$^+$12]. Our findings help to understand the advantages and disadvantages of hybrid observation in intimate, mobile field studies, where traditional in situ/post-hoc observation techniques fail or have severe limitations. We argue that our insights are beneficial for future studies in the field of cycling, tourism, hiking, or various rather intimate sport activities, e.g., running and sailing. With further adaptation of the technique our findings are also relevant for studies and applications beyond mobile outdoor activities. Our findings can contribute to future design, development, and application of hybrid observation techniques in various field study settings.

We start this chapter with information about the Tacticycle, an orientation aid for cyclists, which serves as a relevant and timely example to study the hybrid observation technique. In Section 6.2, we detail the methodological approach and present our used observation tools and apparatus in Section 6.3. Further, we report from an application of the hybrid approach in a field study with 11 cycling tourists and give insights into the advantages and limitations of the methodology (see Section 6.4). We conclude our research in Section 6.5 with a summarising assessment and ideas for future work.

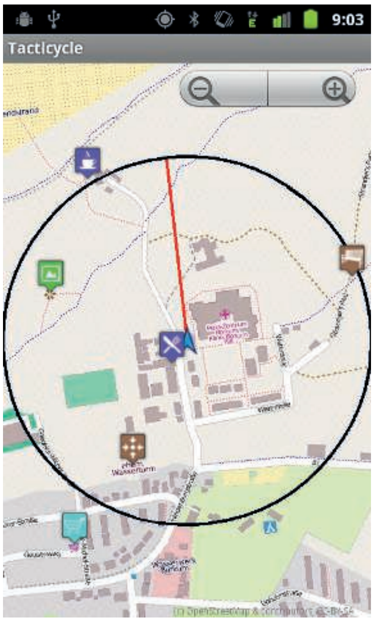
Figure 6.2: The Tacticycle consists of two core components: a mobile phone, which serves as visual display, and two vibration actuators in the handlebars. Both components are connected via custom-made hardware.

Parts of this work were published in Poppinga, B.; Pielot, M.; Heuten, W.; Boll, S. Unobtrusive Observation of Cycling Tourists in the Wild. International Journal of Mobile Human Computer Interaction (IJMHCI), 6(4), pp. 22 – 41, October–December 2014.

6.1 Background: The Tacticycle

In 2008, we observed tourists on bicycles in a typical holiday region and how they use a bicycle to explore the area. We did this as part of the user-centred design with the intention to create an orientation aid for cyclists, which was task of a funded project. We found that tourists tend to plan their bicycle trip on the fly and only have a rough idea of their destination. They do not mind if they travel partly in a wrong direction, as the way itself and the things to see along the route are the only important aspects of their trip. In consequence, we argue that for these bicycling tourist a general orientation support is needed rather then a fully equipped navigation system.

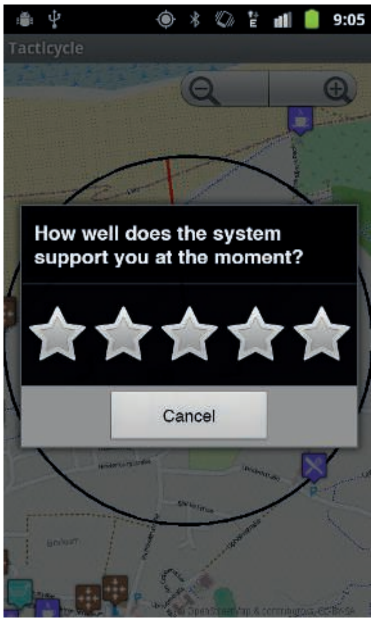
We designed the Tacticycle [PPB09b] as orientation aid instead of a fully functional navigation system (see Figures 6.2 and 6.3). It consists of a smart phone, mounted on the handlebar, and two vibration actuators, attached to each handle. On the smart phone, a map with the user's location is displayed and points of



(a) Per default the user location is centred on the screen.

(b) The map can be panned and zoomed to explore the environment.

(c) Details about the POIs can be shown.



(d) Tacticycle can assess a user's impressions through self-reporting.

Figure 6.3: The Tacticycle software shows the user's own position centred on a map, which is enhanced with nearby points of interests. On a long touch, details about the POI are shown and it can be selected as destination.

interest (POI) are shown. A user can select one of the POI and set it as destination, which will then be displayed through the vibration motors. If both vibration motors are vibrating at the same intensity, the direction is straight ahead. If the left vibration motor has a higher intensity than the right one, the destination is on the left. Consequently, the destination is on the right if the right intensity is higher. If the destination is behind, no vibration is presented. A vibration pulse has a duration of 500 ms. Once a pulse has been shown, the vibration is interrupted for 2000 ms to not annoy the user. Beside this orientation guidance, the Tacticycle also highlights close by POI through a sequence of short vibration pulses at maximum intensity. This is done to motivate the cyclist to explore the environment.

We conducted the final field trial of the Tacticycle during summer 2011 in the same area in which we first made our tourist observations. The evaluation of the Tacticycle was two-fold, i.e., we were interested in user experience-related and observation method-related aspects, whereby both aspects equally contributed to the study design and methodology. Our findings regarding the user experience are published in [PPHB12b], where we conclude that the Tacticycle orientation aid encourages the users to playfully explore the island, whereby a basic orientation is always maintained and users have an overall rich and relaxed travel experience. In contrast, this research focusses on the methodological challenges, solutions and insights, which we gathered regarding the observation of the cycling tourists during the evaluation.

6.2 Method Approach

The requirements for an ideal observation technique had been derived from our understandings of the target users, i.e., cyclists. We are interested in both, qualitative as well as quantitative data. The observation must be as unobtrusive as possible to not disturb the tourists and influence their experiences with the Tacticycle. Thereby it should be as situated as possible and give detailed insights in the user experiences. We further want to be able to assess the importance of individual experiences through qualitative feedback in face-to-face discussions.

Our design combines in situ and post-hoc study techniques. For in situ techniques (i.e., recording part) we decided to use logging, Experience Sampling Method (ESM), and the SenseCam. As post-hoc technique (i.e., reflection part) we decided to use a traditional face-to-face interview. The combination is made by using the gathered data of the in situ techniques to jog the memories of the study participants in the post-hoc setting (see Figure 6.4). With this approach we get quantitative in situ and qualitative post-hoc insights.

The three in situ techniques which we used for our design all have the advantage that no experimenter has to be present during the actual observation in the field

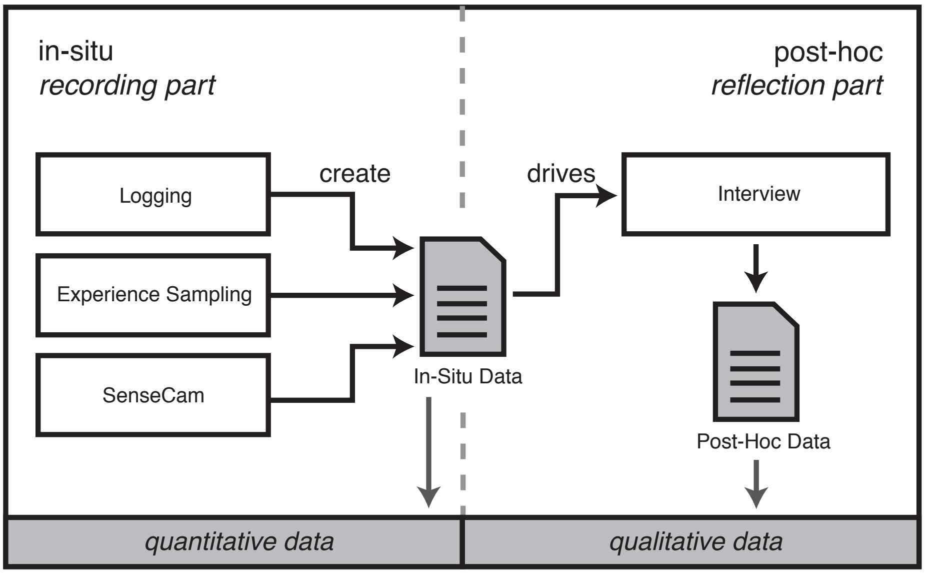


Figure 6.4: The hybrid observation approach consists of two subsequent phases. The actual observation of the participant takes place in the field (recording part), while the post-hoc reflection is done in the lab (reflection part).

[KSAH04, HRKS05, VLR+10]. That allows study participants to use the system freely and unbiased. In addition, costs to conduct the study can be reduced and the scalability of the observation improves. However, these techniques have the disadvantage that they only capture brief moments and without contextual information it can be hard to interpret the data. It remains unclear how the participant felt and what the overall experiences were.

The selected post-hoc technique, an interview, has the advantage that the experimenter can get a good overall impression on the participant's study experiences. If something is unclear, the experimenter can ask the participant to further elaborate on a specific topic. The downside is that the technique is lacking situatedness. That means that participants may forget about important details or report them with less accuracy [BGHS09, KDB+11].

With the combination of both approaches, we want to combine the advantages and cancel out the disadvantages. With the objective in situ data available in the post-hoc setting we envision that participants get additional clues on, e.g., where they have been (logging), how they felt (ESM), what they have seen (SenseCam). The objective data can further be used to validate what the participants are reporting in the interview. When talking about the study experiences, the experimenter will further get contextual information for the in situ data, which makes them more meaningful. We claim that the hybrid approach allows us to get highly situated and more trustworthy insights, whereby the technique is mostly unobtrusive, cost efficient, and allows an unbiased use of the to be investigated prototype.

6.3 Observation Tools and Apparatus

For the observations to be made in the actual study, i.e. recording part, we implemented and configured the set of three identified in situ observation techniques, i.e., logging, ESM, and SenseCam. For the reflection part we invented a software, named Context Player, which is somehow equal to the Replayer described by Morrison et al. [MTC06], and which is capable to synchronise and visualise the gathered data to a study participant. In the following we will elaborate on each part individually.

6.3.1 Recording Part: Virtual Observer

Intelligent logging describes the observation and recording of mostly technical information, i.e., application configuration, sensor information. We integrated the Virtual Observer logging framework (see Section 3.1.2) into the Tacticycle, which allows us to record location- and application-specific details, like the user's location including accuracy, speed, heading, the compass bearing, selected destination, and highlighted POI. We configured the framework to save all values every second, which allows an accurate reconstruction of the whole bicycle trip.

In addition we integrated ESM into our prototype, which directly asked the user questions about their experiences. We triggered a question every 15 minutes. No audio or vibration feedback was presented if the question pops up to not interrupt the user. Questions could be dismissed by clicking a cancel button. Questions were automatically dismissed without any notice after 45 seconds. We asked 5 different questions about how disoriented and relaxed s/he is, if the user knows where the destination is, and how supportive and distracting s/he thinks the system is. However, neither the exact questions nor the answers are of importance from a methodological perspective. To answer a question, a participant could select from 1 to 5 stars, indicating whether a participants agrees or disagrees to a statement. A screenshot of the Tacticycle application asking the user can be seen in Figure 6.3.

The SenseCam was also part of our observation system. The SenseCam has originally been developed by Microsoft Research [HWB+06] and is now available as commercial product. It is a neck-worn camera, which takes wide-angle photos of what is in front of the user. A typical SenseCam image can be found in Figure 6.5. We configured the device to capture a photo every 5 seconds.

Figure 6.5: We used the SenseCam as one of three unobtrusive in situ observation techniques. It turned out that in the post-hoc interview the visual impressions support the participants in remembering situational details.

6.3.2 Reflection Part: Context Player

The logged data, the answers on the ESM questions, and the raw SenseCam images are all individually saved in their own data structures. This makes it quite hard to synchronise, relate or compare them to each other. However, the relationship between the data is of great interest, since, e.g., a reason for a logged U-turn, i.e., participant turns around 180°, can be found in the SenseCam images. To bring all the different observations in line we developed the Context Player and used it as an essential component in the reflection part.

The Context Player is a Java-based desktop software application, which is able to read and synchronise the different data sources (see Figure 6.6). It consists of three sub-components, a map panel (labelled with A), a SenseCam image panel (B), and an ESM answer panel (C). The context can be set to a certain time stamp by using a slider (D), which ranges from the first to the last recorded context element. We use the word "player" in the name as the interaction is similar to a desktop audio player, i.e., play, pause, move the slider to a certain place.

In this study the OpenStreetMap-based map panel showed the user location, the travelled path, highlighted POI, and the destination were shown. The SenseCam panel showed the actual SenseCam image, and the ESM answer panel displayed the recently asked question and the given answer. We envision that the

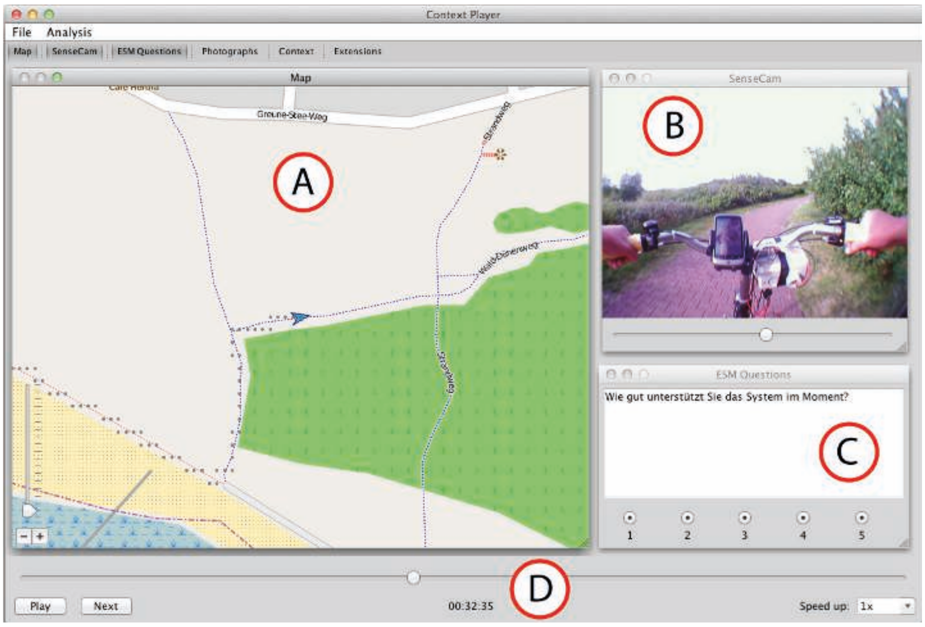


Figure 6.6: The Context Player is able to synchronise the data from the different observation techniques. It consists of a map, the corresponding SenseCam image and the latest given ESM answer.

Context Player can support the reflection process and jog the memory of the study participants.

There are several realisations of the Day Reconstruction Method. All have in common that the participant can perceive the experience without interruptions to, e.g., take notes like in diary studies. Typically they are asked the next day to elaborate on the day before. Form-like papers are used for the actual reflection, where a study participant can divide the day before into chunks and elaborate on them. To emphasise the reflection on the chronological order and location of experiences, and to make it easier to add impressions in a second reflection phase, we adapted the form-like paper. For our study we decided to use a horizontal timeline, ranging from the start to the end of the trip (see Figure 6.7). Insights can be added quickly and revisions are easily done. Further, the placement of thoughts on a timeline with a known frame of reference, allows to get impressions on chronological order and geographical distances at a glance. We think that a timeline is a well-suited tool to communicate about location-based or location-related events.

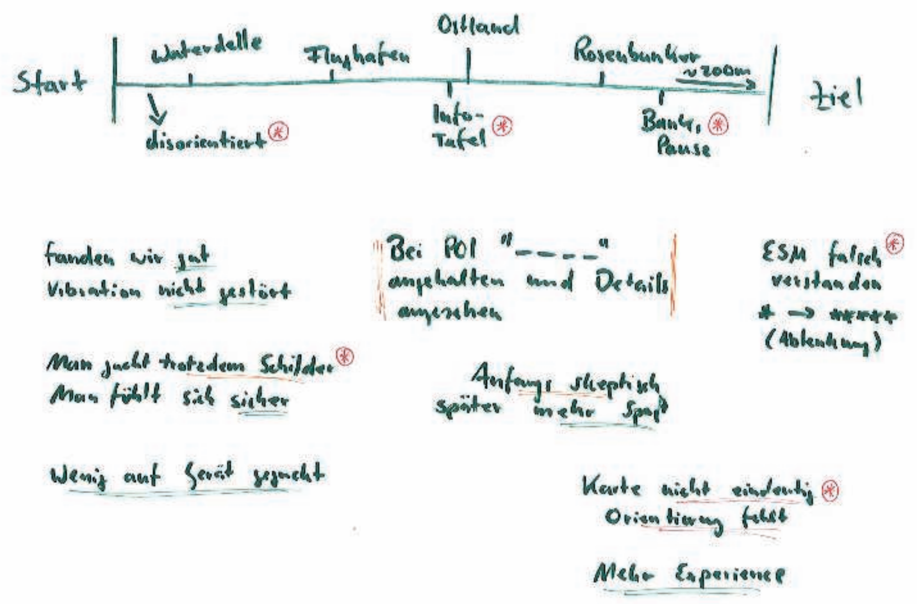


Figure 6.7: As part of the post-hoc interview we decided to use a timeline for taking notes, as it allows to easily assess and relate reported places and experiences. Insights we get during the reflection with the help of the Context Player were marked with an asterisk.

6.4 Evaluation

In our evaluation, we wanted to get initial, mostly qualitative insights into the proposed methodology. To do so, we applied the Tacticycle and the hybrid observation technique in a real scenario, i.e., with tourists on a holiday bike trip. In particular, we were interested in

- how the recorded feedback helps to remember in situ details,
- how valid the initial user reflections are,
- how the communication between experimenter and participant changes with the Context Player available,
- where the information of the different in situ techniques is used for, and
- how obtrusive the in situ observation techniques are.

We will elaborate on each of these points in the Findings section later on. Note, that this research highlights and discusses the methodological findings. Details about our findings regarding the user experience are published in [PPHB12b].

Id	Gender	Age	Duration	Distance	Reflection
P1	M	45	01:04	8.3 km	Immediately
P2	F	45	00:10	1.7 km	Immediately
P3	F	40	01:27	16.2 km	Immediately
P4	M	55	00:53	10.9 km	Immediately
P5/6	F	50	02:34	9.9 km	1 Day after
P7	F	40	01:15	14.8 km	2 Days after
P8	M	45	01:13	11.3 km	Immediately
P9	F	30	00:23	2.7 km	Immediately
P10	F	20	01:28	17.8 km	Immediately
P11	M	55	00:30	4.0 km	Immediately

Table 6.1: In total 11 volunteers did 10 *different bike trips. In average a participant cycled* 9.76 km *in about 66 minutes.*

6.4.1 Method

The evaluation took place on Borkum, which is a German island in the North Sea. The island is often visited by tourists and has several POI to visit. We cooperated with a local bicycle rental, where we got space for preparing and maintaining the Tacticycle. We used the bicycle rental for the briefing, the interview, and the de-briefing of the study participants. Two experimenters conducted this study within 7 working days in summer 2011.

We acquired the study participants from the regular rental visitors, who had the intention to rent a bike. The Tacticycle was set-up and running in demonstration mode in the entrance of the shop. In addition some posters were presented, sharing the idea of the project and the Tacticycle itself. 11 volunteers participated in our study, whereof 7 were female (see Table 6.1). The estimated age ranged from 20 to 55 years. In total the participants were observed for 97.6 km and almost 11 hours.

We approached every visitor of the shop, explained him or her the basic idea of the Tacticycle and asked if they would be interested in participation. If so, we either agreed on a fixed date, when the Tacticycle should be tried, or if it was possible handed the Tacticycle immediately to the volunteer. Before the actual study and trip began, we asked the participant to sign an informed consent. We briefed the participants that they should continue to mind the traffic rules, while using our system. We also explained all the system functionalities, in particular the tactile feedback, in detail and let them interact with the system on their own. We explained that the mobile phone as well as the SenseCam are recording data and how this can be turned off to maintain the participant's privacy if desired. All participants were allowed to keep the Tacticycle as long as they needed it to finish their trip, but for no longer than one day.

After they finished their trip, we met again in the bicycle rental to hand back the bicycle. We then agreed on a time for the follow-up interview, if it was not possible to do this immediately. The interview has been done in a partially separated room and not during the regular opening hours of the rental. In the semi-structured interview, we engaged the participants to reconstruct their trip. Our primary interest was on the users' experiences with the Tacticycle and how it supported the user in various situations. During the interview, one of the experimenters drew events and places on the timeline, following the participant's advices. At the same time, the data from the SenseCam and mobile phone was copied and loaded into the Context Player, which took about 5 to 7 minutes depending on how long the participant used the bike. Once the study participant started to repeat details or do not seem to remember any more experiences, we introduced the Context Player to the participant.

We then started the trip playback, fastened up such that 10 seconds of the trip corresponded to 1 second in the Context Player playback. We asked the participants to comment on the Context Player playback. If the participant started to explain or describe something in-depth, we paused the playback and if necessary played this part of the trip again. If one of the experimenters was unable to understand the situation, we paused the Context Player and asked the participant to clarify what was going on. However, we did not overstress certain situations to not cause any biases. After the trip had been reconstructed we continued to ask the participant more general questions about the Tacticycle and the observation techniques, e.g., how distracting the logging or the SenseCam have been. Overall, a typical reflection phase took between 30 and 50 minutes.

Finally, the participants received a voucher for the bicycle rental and a USB stick as a reward. We neglected our initial plan to ask the participants to fill out two questionnaires (System Usability Scale and an individual questionnaire on existing navigation experience) because the study itself already took quite long and we did not want to bother the participants any longer.

6.4.2 Findings

In the following we will report on the insights we made during the study. Wherever possible, we illustrate by and refer to concrete examples, quotations or figures. Overall we have the impression that all participants were quite happy with the Tacticycle itself and with the used hybrid observation technique. In detail, our key findings are:

- Our methodology allowed some participants to remember details, which they were not aware of any more.

- Many statements made by the participants could be supported by in situ data. However, we also identified some weaknesses in the participants' descriptions.

- Communication and reporting behaviour becomes lively with the Context Player.

- We got mixed insights regarding the obtrusiveness of the in situ observation techniques. Overall we would say that the SenseCam is the most obtrusive, followed by ESM and logging.

- We observed that during the reflection part the map was used to refer to places, while the SenseCam images were used to support situated experiences.

6.4.2.1 Remembering Details

Before we showed the Context Player to the participants, most of them behaved in a restrained manner and only mentioned the most prominent points of their trip. With the help of the Context Player 11 complementary and 17 completely new insights could be collected in total. Regarding the complementary insights, 6 participants remembered the name or location of a POI they visited or passed by. E.g., P1 described something very generic about "some kind of place, ... there were lots of children, maybe a cafe or ... ". With the help of the Context Player this place could be clearly identified as Ostland, a cafe with a prominent playground next to it. P3 was reporting about a viewpoint and a close by cafe, where she has been. She was unable to elaborate on any details and was unsure about the order in which she visited the places. The Context Player helped her to find the names for the POI and remembered her about the order she was visiting these places.

With respect to the completely new insights, 3 participants forgot to mention a break at all. With the help of the Context Player they were able to remember this place and tell the experimenters what happened there ("Oh, yes, we stopped there" (referring to the Context Player), "had a break and took some photos of the bike", P5). 2 participants identified that they had mentioned the visited POI in the wrong order. For 2 participants the experimenter identified that they were using paper or public maps as additional navigation aid or have asked passers-by about directions ("we stopped and asked the woman standing there if we are cycling into the right direction", P6). 3 participants indicated that they felt disoriented or made a navigation error, which had not been mentioned without the Context Player, e.g., P3: "I had some problems with the system and felt disoriented. Thus, I stopped at this bus stop" (identified with the help of the map, see Figure 6.8). "Yes, there was this bench I was standing in front of" (pointing at the SenseCam image).

Figure 6.8: In this setting P3 felt disoriented. She used the map to communicate that she stopped close to a bus station. She referred to the SenseCam image for the detail that she had stopped in front of a bench.

Further, 3 participants remembered where and which follow-up destination they set with the help of the Tacticycle software. For 2 participants a misinterpretation of the ESM questions or scale could be identified.

From our observations we can report that none of the participants seemed to conceal any of the complementary insights by intention. More, it was noticeable that the presented information somehow helped them to remember the earlier experience better. While the complementary insights typically cover minor details, like remembering a name of a cafe, the completely new insights are of high interest. It seems that in particular potentially negative insights could be revealed, e.g., that participants used additional navigation aids or that some participants felt disoriented or made navigation errors. We are undecided if the participants intentionally avoided to talk about their negative experiences to not disappoint us.

6.4.2.2 Trustworthiness of Insights

The additional reflection with the help of the Context Player mostly supported earlier made statements by the participants. However, for some participants we were able to identify differences between the participants' initial descriptions and the recorded and objective facts. One participant was telling us that the tactile feedback indicated to go straight. However, in his opinion there was no way in that direction available and he therefore turned right to continue his trip: "There

Figure 6.9: The participant reported that there was no path available to further go straight. However, the SenseCam image shows the according traffic sign and the map indicates that there is such a way (marked with arrow).

the system said I should go straight on. However, there was no path available. I could go either left or right and I decided to go right." (P8, see Figure 6.9)

In fact, and visible on the map and SenseCam, there had been a straight way, which the participant did not notice for unknown reasons. Two participants (P5, P6) stated that they were not looking at a map during their trip: "no, we did not use a map during the trip". However, the Context Player showed that they indeed stopped in front of a public map and seemed to be discussing where to go next. This, however, have not been discussed any further.

Another participant (P11) stated that he had been at a certain POI. The Context Player showed that this was not the case. When we directly asked the participant about this, he replied as follows: "Yes, now I remember, I set this monument as destination, but re-decided to go back home at some place."

We do not think that any of the participants lied to us by intention. Moreover, we think this is a weakness in the perception of the environment, i.e., maybe some plants covered the view on the straight way, the participants were overly excited or they were simply mixing up things. However, for researchers these insights are potentially of high interest and we think that they would not have been revealed without the Context Player.

Figure 6.10: The Context Player was appreciated by all participants during their reflection. One experimenter was doing the interview, while the other experimenter controlled the playback.

6.4.2.3 Communication Behaviour

When we started the post-hoc interview and reflection on what the user experienced, most users only told us very general things. Without the Context Player, typical things they reported were, e.g., where they had main difficulties using or understanding the system (e.g., "I was unable to find a location next to this bike rental on the map", and "If you are cycling straightforward for a while, the vibration feedback gets annoying"). At this point some participant already stopped talking. Most participants continued to talk, but the statements were still very generic, partly repetitive and rarely contained a concrete location (e.g., "I did a U-turn by intention to test the system", and "There was a railway crossing, where I had problems to follow the vibration signals").

When a participant stopped talking and did not want to add anything to the drawn timeline, we showed the Context Player and played the participant's route (see Figure 6.10). While doing so, most of the participants remembered various details or were able to clarify earlier statements, e.g., "there I did the U-turn" (pointing at the map), and "there is the mentioned railway crossing" (pointing at the map). Further, the experimenters were able to identify other interesting points (e.g., stops, navigation errors, U-turns) and could pause the playback to ask the participants about them.

Our impression is that the Context Player, with all the recorded data presented to the user, helps to increase the understandability and comprehensibility of the participants' comments. Further, the overall communication between the experimenter and the participant gets more inspired and focused. In our opinion the timeline also helped to keep the reflection as focused as possible. In addition, it helped to avoid misunderstandings, because the essential places of the trip can be obtained at a single glance. Therefore it was clear within seconds if a new comment belongs to a certain part of the timeline or not.

6.4.2.4 The Role of the Individual Parts of the Context Player

As described, the Context Player consisted of three individual panels, i.e., a map panel, a SenseCam panel, an ESM answer panel, shown to the participants. They were arranged as shown in Figure 6.6. During the reflections of the participants we identified some preferences on how the individual panels were used. The map was the first choice if the participants were remembering certain places or events, which they relate to places (see Figures 6.8 and 6.9). Thus, the map helped several participants to name exactly where they have been (e.g., name of a cafe or viewpoint). In contrast to this, the SenseCam helped the participants to remember more or less important, situation-specific details for a given context (e.g., "we stopped and asked the woman standing there if we are cycling into the right direction", P5/P6). Interestingly, the participants also remembered details, which were not explicitly visible on the SenseCam images (e.g., "I briefly talked to another cyclist", P8). The users mostly ignored the ESM information. Our insights coincide with earlier results in a related scenario [KSWK10].

P1 summarised our findings very well: "The map is more important for me than the SenseCam photos. Because I exactly remember that there" (pointing at the photo in the Context Player) "was a tree and I was standing right at the side of this path" (pointing at the photo again). "However, I was unsure how far away this is from my previous destination. That's what I can get from the map."

From the experimenters' point of view, the map often helped to make sense to the comments of the participants, which were often quite vague. In addition, it helped to identify, where a participant stopped, did a U-turn or a navigation error. Also it was helpful to see which destination was set or which POI had been highlighted and how the reaction of the participant on this highlight was, if any. The SenseCam images were in particular helpful to identify if and when the participants used external tools to orient themselves (see Figure 6.11). Further they showed, if and how the participants interacted with the system. They also often supported the participants' statements, e.g., that the display was hard to read in direct sunlight because of reflections. In contrast to the study participants, the ESM answers were helpful to the experimenters. They were in particular interesting because they are unfiltered and provide the raw and blunt feedback, given from the user to the system. Based on this, we found that one participant was annoyed from the overall system, because of the reoccurring ESM questions.

Figure 6.11: The SenseCam images often enabled a participant to remember some situational details. For the experimenter the images were helpful to identify, what a participant was doing, and when and how s/he interacted with the Tacticycle.

6.4.2.5 Obtrusiveness of the Observation Techniques

The obtrusiveness of an observation method is very important, because an obtrusive observation technique can affect the natural behaviour of the study participant or even passing by people. We assessed the combination of methodologies in two ways. First, we looked for indications, whether the observation was obtrusive, during context playback. Secondly, we directly asked the participants in the interview how obtrusive they think this observation technique was. From the Context Player we learned that one participant was somehow surprised that his rest was also observed ("Obviously, really everything was observed", P1, see Figure 6.12). Two participants were thinking about if there could be some embarrassing photos on the SenseCam, even if they did nothing else than cycling. We also noticed that the participants sometimes were seeking for privacy and thus, were hiding the camera (P3, P10). We take this as an indicator that especially the SenseCam cannot be completely forgotten during short-term studies.

The Experience Sampling Method was also investigated with respect to obtrusiveness, because it was the most user involving technique used. We found that all participants were eager to answer the questions. In fact 20 of 43 total questions have been answered. 19 questions have been overseen and 4 questions have been actively dismissed. We identified one participant, who accidentally dismissed a question, because of the shaking of the device while cycling. Only one participant said that the ESM questions got annoying over the time, because of reoccurring

Figure 6.12: We got different feedback regarding the obtrusiveness of the SenseCam. P1 forgot about the cam, had a break on the dike near the shoreline and was lying in the sun.

questions and the short time interval between questions. Most participants complained about that the questions were hard to read on the device because of irritating sunlight reflections (see Figure 6.11) or traces of fingerprints on the display. Two participants (P5/P6) misunderstood the scale of the questions and were accidentally thinking that 5 stars always indicate that one is satisfied with the system. When the experimenter asked, "OK, there you've rated that the system is distracting you very much." they replied, "No, we haven't been distracted by the system at all. There must be something wrong."

Without post-hoc discussion that misunderstanding would probably never been revealed. We conclude that cyclists probably need bigger and more separate buttons to allow a more precise selection of given actions. Further, the scale should always be interpreted the same way, e.g., 5 stars mean always that one is convinced by the system. For future studies we recommend to increase the time where a question is displayed to several minutes.

None of the participants commented on the intelligent logging, which we used to record some navigation and interaction information. Thus, we think that this is the most unobtrusive of the three techniques.

In summary, we share the impression that most participants were aware that they are observed, but none of the in situ observation techniques was really obtrusive. The SenseCam raised the most concerns across all three techniques, probably because it serves the experimenters with vivid photographs instead of sensor data. However, participants knew how the device could be suspended and made use of it when desired.

6.4.3 Limitations

Our study has not been designed as experiment. Thus, no formal comparison against any traditional techniques can be made. Also, no unambiguous cause and effect relationship can be made. However, we consequently followed the described study procedure. First, we let the participants reflect on their experiences without the Context Player; then, we introduced the Context Player and did the reflection again. Nevertheless, the study design allows us to make a qualitative assessment of the Context Player and it also allows us to pinpoint that and how it supported the original reflection process. We paid extensive attention that each of the claims made can be fully traced back to it's origin by giving quotes wherever possible.

We extensively used the recorded in situ materials, i.e., logging data and SenseCam images, in our paper on the user experience with the Tacticycle. However, we rarely use this material to argue about what we learned from a methodological perspective. The reason why we did this is two-fold. First, the objective data itself cannot contribute to the methodological understanding on its own. It needs to be related with the post-hoc interviews to have a methodological meaning. However, here it is impossible to determine in advance which aspects of the interview relate with the objective data. Second, even if we were able to relate objective measures with certain aspects of the interview, the following argumentation would be weak because of the characteristics of an open interview. Instead of, e.g., clearly naming a place and time, participants would typically tell a whole story, where places and times are—if at all—only vaguely defined.

6.5 Summary

The observation of cycling tourists is, similar to many other intimate settings, quite challenging. Traditional in situ observation techniques fail as they threaten the intimateness of the experience and interfere with the users' tasks. Post-hoc studies are unsuited, because participants fail to remember all the contextual details. Hybrid observation techniques were proposed a while ago and applied to lifelogging or the observation of everyday activities. Their application in regular field studies and in intimate contexts, like cycling, hiking or running, is questionable and not fully understood. In this research, we report from the application of a hybrid observation technique in a field study with intimate context, i.e., a study of cycling tourists. We used logging, ESM, and a SenseCam for in situ observation, and custom-made software, i.e., the Context Player, to visualise recorded data and assist study participants in the post-hoc interview.

In summary, we found that the hybrid observation approach is perfectly suited to observe users in intimate, mobile settings. We observed that participants continuously included the recorded in situ material into their post-hoc interview and that this led to a vivid and more precise communication between participants and experimenters. The map visualisation in the Context Player was mostly used to refer to places and SenseCam images were considered to support situated experiences. Statements made by participants were often supported through in situ data and most participants were able to remember additional situational details. Further, the in situ data could also be used to identify and clarify on weaknesses in the participants' reporting. We got mixed impressions regarding the obtrusiveness of the in situ techniques, but would argue that SenseCam is the most obtrusive, followed by ESM and logging. We conclude that hybrid observation is a very valuable approach, and its application would be beneficial in many other intimate mobile study settings, which otherwise could hardly be investigated. In the following, we elaborate on each of our research questions individually.

We conclude that none of the used in situ techniques, i.e., logging, ESM, SenseCam, is seriously inferring with the study task or cycling experience. We observed that the SenseCam raised most concerns by the participants. When we handed and explained the SenseCam most people reassured themselves how it can be switched off in case of privacy need. Nevertheless, some participants expressed vague concerns about the recorded images when they returned from their trip, mostly fearing somehow exposing pictures. However, this was not the case and participants appreciated the presence of the images in the post-hoc setting. We assess ESM to be second most obtrusive technique, although the gathered information is not used by the participants at all. Consequently, ESM would be our first technique to omit. Logging was not noticed at all, but particularly the map visualisation was really helpful. Thus, we argue it is the least obtrusive in situ observation technique.

Regarding the relation and communication between experimenters and participants, we learned that the initial briefing took a bit longer than in traditional field studies, because of the complexity of the study procedure. Nevertheless, in our impression the briefing was not sufficient to create a deeper relationship between participants and experimenters. Consequently, participants were able to try the Tacticycle mostly unbiased, and, in the post-hoc setting and compared to what we know from other studies, formulate criticism in a straighter way. While the reporting was a bit vague without assistance, it became really lively and vivid with the Context Player available. We have the overall impression that all participants appreciated that they were able to cycle without an experimenter accompanying them, and that they welcomed the assistance through the Context Player.

We further studied which in situ techniques particularly supports which post-hoc interview situations. Here we found that our observations are mostly in line with literature [KSWK10, CM05, IKM02]. We observed that both, the map view (see Figure 6.6, label A) and the SenseCam view (see Figure 6.6, label B) were used equally often by the participants to go through the whole trip again and to identify key aspects. The SenseCam images were mostly used to refer to contextual details, and the location information was used to refer to visited places and stops. The ESM view (see Figure 6.6, label C) was ignored by most participants. The information provided by ESM was helpful for the experimenters, but didn't need additional post-hoc discussion.

Given these insights, we conclude that the combination of both, i.e., in situ and post-hoc, techniques to a hybrid is a valuable approach and works well in field studies with privacy constraints, like cycling. While the in situ observation is precise, objective and unobtrusive, the post-hoc discussion provides more general qualitative insights and allows to experimenter to ask questions if needed. We argue that this leads to an overall increase of situatedness. We further argue that the technique can also be used in other field studies, where an unobtrusive, but nevertheless situated observation technique is needed, e.g., studies on hiking or running. Future work should investigate how the data presentation through the Context Player can be improved and adapted to other post-hoc study settings, e.g., focus groups. Further, it may be studied how the Context Player can be integrated into longer-term studies. Here, we think that an in situ reflection, similar to the Storyteller (see Chapter 5), can be quite beneficial to consolidate made experiences till the final post-hoc interview.

7 Conclusions

It is important to understand the users and their needs to design and develop targeted solutions. Thus, many of today's design processes, for example the HCD process, incorporate one or more steps to either understand the user or to evaluate a created design solution. In the era of mobile and ubiquitous computing, where interactions and user experiences are immersive and omnipresent, the observation of users is a real challenge. Because traditional lab-based or supervised observation techniques come with significant limitations, unsupervised observation techniques were proposed and studied. The key advantage of such techniques is that they don't require the presence of a human observer and gather insights through external sources or active user feedback.

Unfortunately, these unsupervised observation techniques only meet the three key challenges and quality criteria for observation techniques to a limited extent. They have problems in conveying *situated* insights, which is crucial for the observers to really understand the users' situations and problems. Further, some unsupervised techniques are *obtrusive* and infer with the users' tasks and activities. In extreme extent this could lead to changes in the behaviour, which negatively affects the validity of study insights. In addition, these techniques have limitations regarding the *scalability*, which negatively affects the overall efficiency of a method.

In this thesis, we studied a two-fold approach to overcome these limitations. First, we investigated *smart observation*, i.e., how the interaction and computing capabilities of a mobile phone can be used to improve unsupervised observation techniques. Second, we researched *hybrid observation*, i.e., how collected in situ data could be re-used to drive subsequent post-hoc studies. The overall objective of our research was to increase the value and applicability of unsupervised observation techniques in field studies of mobile and ubiquitous technologies.

We systematically studied both, *smart* and *hybrid observation* with a selection of the three most popular, unsupervised observation techniques: logging, ESM, and diary studies. First, we studied how *smart observation* could improve logging, and ESM. Then, we studied how *smart* and *hybrid observation* will affect diary studies. We conclude with a combined study on how *hybrid observation* can be used with ESM and logging. This series of consecutive studies allows us to give a differentiated assessment of the strengths and weaknesses of the individual approaches and observation techniques.

In the next section we summarise our key findings and contributions, and provide detailed answers to the initially raised research questions. Section 7.2 illustrates the expected impact that our contributions may cause. We discuss limitations of our work in Section 7.3 and give ideas for future work in Section 7.4.

7.1 Contributions

With our overall approach in mind, we initially identified a set of five research questions. We conducted at least a single study to approach each of these questions and contribute by providing detailed answers to these. Further, we condensed all of our methodological efforts into a single observation framework. This framework allows other researchers to apply and extend our methodologies, which also makes a significant contribution. In the following we provide a summary for each of the contributions.

RQ1: To What Extent Environmental and Physiological Information Can Be Used to Enhance the Situatedness of Logging

This research question targets the logging observation method, where quantitative data of various sources is recorded. For example, this technique is applied to record the path that a study participant walked. A key problem of this technique is the limited situatedness that it can provide.

Environmental and physiological information have both shown their value in other contexts. Consequently, we studied how these extensions can improve the situatedness of logging in field studies. We conducted a field study with 11 participants, where we recorded traditional data, e.g., user location, but also static information about the environment, e.g., locations of crossings, and physiological data, such as heart rate and skin conductance.

Environmental cues allowed us to argue about spatio-temporal behaviour of a study participant on sub-route level. We identified that pedestrians follow two strategies to interact with their mobile device between waypoints, i.e., some use their device between waypoints and some consider it right at a waypoint. We further observed that some participants tend to use the device frequently, but for short time spans, and some use it less frequently, but for longer periods. Particularly those waypoints, which are rated as difficult, lead to more device interactions and a reduced walking speed. We were unable to derive any insights from physiological data, because of lacking responsiveness and resolution.

Our answer to this research question is two-fold. On the one hand, we found that static environmental information provided us with a new way to analyse and structure the remaining data. We were able to analyse our data under environmental aspects, which led us to valuable insights that would be impossible to get otherwise. On the other hand, physiological data did not increase the situatedness. This, however, can likely be credited to the novelty of the used sensors and will change with future hardware improvements. We conclude that particularly the use of physiological information for field studies needs further investigation.

RQ2: In What Way a Rule-Based Analysis of Logging Data Can Reveal Valuable Insights

Another problem with logging is its limited scalability. In today's applications often a broad variety of data is recorded at a high sampling frequency. It is then left to the analyst to identify relevant information and do a thorough analysis. With increasing data complexity and data amount this is getting a more and more complex and time-consuming task, which limits the scalability of the logging technique. It is questionable if assisting or automatic analysis can support the analyst and reveal valuable high-level insights.

We implemented our approach by introducing a rule-based analysis which automatically analyses recorded data. In fact, all data that is not matching the rules is ignored. Consequently, the analyst is left with pre-interpreted data, which simplifies the analysis. We studied the rule-based analysis for the popular *scanning* interaction technique, where a user points the device in various directions to interact with an application.

We found that observations made through the rule-based analysis are similar to findings, which were observed by a human experimenter in earlier studies. Consequently, we argue that scan events can be reliably detected through a rule-based analysis. The high accuracy and precision of the recorded data further allowed for a much more detailed analysis of the measures. Thus, we were able to study scan frequency, scan duration, walking speed, and the covered scan angle in great detail.

We conclude that rule-based interpretation of logging data is a valid and valuable approach. Automatically extracted and pre-analysed data makes it simpler for the analyst to interpret and understand the data. This leads to reduced effort, and therefore to an increased scalability and less costs. Further, logging itself allows for a higher granularity in the analysis of the gathered insights. This leads to more accurate insights, which increase the method's situatedness.

RQ3: If Machine Learning Could be Used to Issue Self-Reporting Inquiries at Opportune Moments

Self-reporting techniques, such as the Experience Sampling Method, actively ask the user to provide feedback, e.g., to give a self-assessment on current emotions or to answer a questionnaire. This is particularly helpful since not all information could be observed by a human observer or a sensor. The problem, however, is that these inquiries often appear in inopportune moments, which leads to stress and frustration. An open question is to what extent machine learning could be used to detect opportune moments and issue inquiries when they will likely not disturb the user.

In our work, we collected context data for several thousand answered and ignored notifications of 79 users. The context data consisted of, e.g., the user's speed, the device posture and local time. We did a statistical analysis of the data and used it to create a model, which predicts whether a moment is opportune or not. The idea is that this model is a part of a notification management system on a mobile phone and effectively only issues notifications in suitable situations.

The statistical analysis showed that notifications should be issued at the right time, preferably in later hours of the day and not at night. Further, notifications are frequently answered if the device is already at hand. A cross-validation of our derived classifier found that opportune moments can be predicted with approximately 78 % accuracy. The evaluation showed that this is a considerable improvement over a baseline classification approach.

Consequently, we conclude that machine learning can be used to issue self-reporting inquiries at opportune moments. Nevertheless, before a practical application it needs to be studied how this model can be integrated into a larger notification management framework. Future research should further investigate how time-critical notifications should be handled, and whether the derived model works with non-mediated notification strategies. In the end, the application of the model will lead to fewer user interruptions, and therefore to less stress, frustration, and effort. We further assume that this will lead to an effective increase in answered notifications and in the quality of gathered insights.

RQ4: How Storytelling and In Situ Reflection on Diary Entries Affect a Post-Hoc Interview

In diary studies it is often completely left to the user when and what they should record a diary entry. Thus, people forget to create entries and, if they don't, individual entries are often unrelated to each other and not interconnected. In follow-up interviews it is therefore hard for participants to remember the rationale for creating individual entries. In addition, it is difficult for participants to recreate relations between entries to tell the interviewer a holistic, comprehensive story. In the end, the experimenter is unable to get a detailed and situated understanding on the overall user experience. It is questionable if storytelling or in situ reflection on diary entries would encourage users to improve their diaries, which in the end leads to more valuable interview insights.

We presented an approach, where the smart phone's lock screen is modified to provide the users with a quick means to capture new diary entries. Further, we designed and developed a mobile authoring tool, named Storyteller, which allows to combine individual elements to a holistic and comprehensive story. User studies were conducted to drive the design of the tools and to assess the effect of these modifications in a field study.

We identified that most participants prefer to tell their stories in chronological order. Further, multimedia content was rated more important than text notes, which were mostly imagined to be used in a supplemental way. We found that participants liked the prototype implementation, which we created based on these insights. They particularly appreciated the quick and easy media capture lock screen which, however, did not lead to an significant increase in created entries. Study participants further appreciated to tell their stories with Storyteller, although this was also not significantly different compared to the used standard tool set.

We conclude that storytelling and in situ reflection on diary entries are very promising approaches. In our study, we were unable to find significant improvements, but qualitative insights indicate that the Storyteller was more appreciated by the study participants in several regards. Future work should study if these trends can also be supported for longer-term diary studies, which is where we think the method approach will show its strengths.

RQ5: How a Hybrid Observation Approach Performs in Longer-Term Non-Everyday Field Studies

Hybrid observation techniques combine in situ material with post-hoc studies. Recorded or created in situ material is re-used and integrated into post-hoc studies to elicit further insights on the users and their experiences. This technique is comparably novel and almost no studies have shown the applicability of the approach in longer-term, non-everyday field study settings.

To re-use material in post-hoc studies is a key part of the thesis' approach. Consequently, we deployed a variety of in situ study techniques in a longer-term field environment, where participants were facing a mostly exceptional experience, like it is the case for field studies. We used logging, ESM, and a wearable camera to record in situ experiences. A custom software, i.e., the Context Player, visualised the recorded data in a post-hoc setting, aiming to elicit further insights. We compared the visualisation approach with a regular, non-supported interview in an informal experiment.

We found that all participants liked to include in situ material to recap their experiences. Overall, this led to a more lively and vivid communication between experimenter and participant. Thereby, participants preferred certain visualisations to report about episodic and semantic memories. In addition, the visualised data supported experimenters in the identification of weaknesses in given explanations. We got mixed impressions regarding the obtrusiveness of individual in situ observation techniques. The SenseCam was perceived as most obtrusive, followed by ESM and logging.

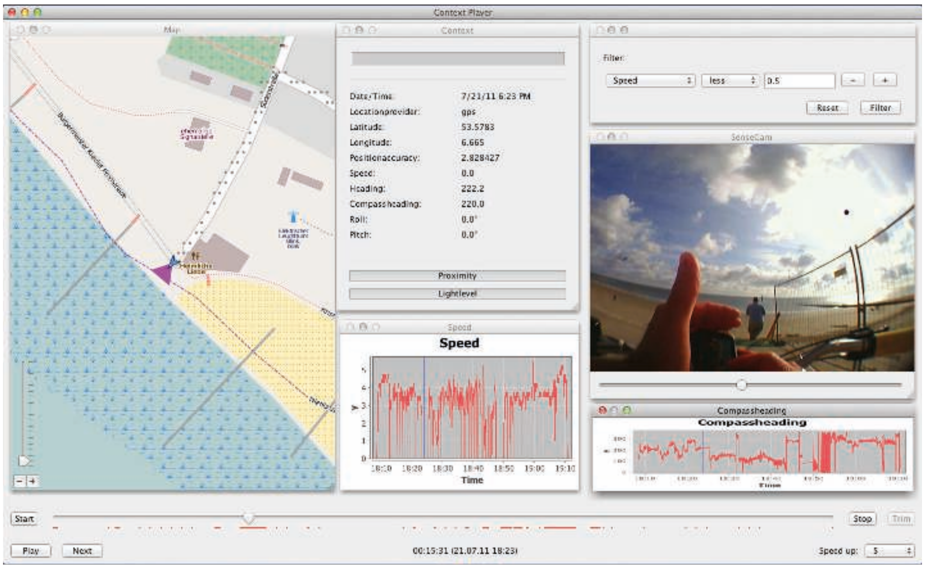


Figure 7.1: The Virtual Observer observation framework and the Context Player were iteratively enhanced over time. These tools enable other researchers to apply and further enhance our unsupervised observation methods.

Altogether, we conclude that hybrid observation is a valuable and well-working approach to perform longer-term field studies. The unsupervised in situ observation techniques are perceived as unobtrusive and provide the experimenters with quantitative insights. In addition, the post-hoc, visualisation-assisted discussion supports participants in their reflection and allows experimenters to get a detailed understanding of the participants' experiences.

'Virtual Observer' Framework

Besides the research contributions we also make a significant technical contribution. Over the several conducted studies, we continuously updated and extended the Virtual Observer user observation framework. With our work on hybrid observation, we also developed a playback tool, which allows the convenient analysis of recorded in situ data.

The Virtual Observer framework was initially introduced in Section 3.1.2. It is an Android library, which can be integrated in any Android projects. It is able to record standard Android sensor data, such as location or device posture, and extended, application-specific information. Further, the Virtual Observer provides means to issue self-report inquiries in opportune situations. All recorded data can be saved on a mobile phone or in the Internet for later analysis.

The Context Player is a visualisation tool for the recorded data (see Figure 7.1). In contrast to existing analysis solutions, the Context Player follows a playback metaphor, giving an exploratory character to the analysis. As described in Section 6.3.2 it is written in Java and therefore runs on all major operating systems. Data can be visualised in different ways, e.g., on a map, or in charts. Further, the Context Player supports the visualisation of SenseCam images or other photographs that were taken with a regular camera in, e.g., diary studies. Data can be filtered and forwarded to more specialised analysis tools, like Weka or R.

We used the framework for every study that we conducted and describe within this thesis. Further, the observation framework was made available to other researchers in an early stage. Thus, the framework was also used by other researchers in studies, which were not reported in this thesis. This allowed us to get early feedback on how the framework should be designed and how it should behave. We used our own insights as well as the impressions that we got from other researchers, to continuously improve the quality and stability of the framework.

With the public availability of the observation framework, other researchers can apply each of the observation techniques that are presented in this thesis. Further, the framework is open source, so it can be modified and adapted to eventually meet other study scopes or allow future research on derived observation techniques.

7.2 Expected Impact

In each of the presented studies, we show that our modifications have a direct impact on one or more of a techniques' key quality measures: situatedness, scalability, and obtrusiveness. We conclude that the situatedness and scalability of logging can be improved by adding environmental and physiological information, or through a rule-based analysis. The obtrusiveness of self-reporting techniques can be reduced by issuing notifications in opportune moments. Diary studies benefit from the storytelling metaphor and the resulting in situ reflection on individual entries, which both arguably lead to an increased situatedness. Finally, we show that a hybrid approach increases the overall situatedness and scalability for long-term field studies in unfamiliar environments.

Most of our findings and method improvements are published and therefore contribute to the state of the art. All researchers who want to conduct a study have the ability to apply our improved observation techniques. This will enable them to improve their studies under various aspects. They have the ability to get more situated insights with less obtrusive observations at increased efficiency and reduced cost. Further, our observation techniques will enable studies, which otherwise would be hard or even impossible to be conducted. Similar to our study on cycling tourists (see Chapter 6), our methods enable researchers to study particularly critical environments or sensitive target users, like tourists.

Because of the methods' unobtrusiveness, the participation in a field study goes almost unnoticed. This decreases the chance of participants changing their behaviour, as they don't feel continuously observed by an experimenter. In return, this increases the insights' validity. Further, observations can become as ubiquitous as the studied pervasive technologies. This allows experimenters to conduct studies over longer periods and on a day-to-day basis. Consequently, experimenters learn how the studied technology is used in everyday life, which arguably leads to far more diverse and realistic insights.

Beyond the plain observation of users and situations, we also observed the methods' suitedness and appropriateness for ideation, i.e., idea generation, and rather exploratory field studies. Logging could unobtrusively record information that not even the users are really aware of. Self-report inquiries might encourage the users to report their feelings in the most boring or extraordinary situations. Multimedia diaries could be used to capture and document unique experiences which were never were recorded or shown elsewhere. Already the plain availability of this information opens up completely new perspectives on the design and usage of a product. The use of the Context Player even multiplies the number of perspectives with it's explorative data analysis approach, which in return leads to even more insights and ideas how users could be supported by an interactive system. Although the focus of this research is on the realistic and authentic observation of users, we expect that also people in idea- and design-oriented environements will benefit from our methodologies, and use our approaches to inspire their idea generation and solution exploration processes.

Altogether, our modifications improve scientific studies of users and systems in their real environments and in real contexts. This will lead to a more valid, more substantiated, and more realistic understanding of users in everyday situations. As a consequence, this will allow designers to create the most appropriate and best suited design solutions. These design solutions will, in the end, lead to a better user experience, an improved usability, and to eventually more appreciated products and services.

7.3 Limitations

Methodological work is typically facing two key critical aspects: transferability and validity. With observation techniques being methods, they are facing these two aspects as well. A method is considered transferable if it can be applied to any setting and then performs equally well compared to these settings, where it is known to work. Typically, transferability is granted to a certain extent. For example, a questionnaire can be filled in all situations, where a participant is not fully busy with other tasks, and has cognitive and physical resources to actually interpret questions, and to phrase, and write an answer.

Another critical aspect is a method's validity. Researchers rely on methods and often blindly trust the collected insights and results. Consequently, it is important that observation techniques really observe, record, and measure the aspects, which they pretend to be measuring. This is particularly critical for quantitative data, which is processed through statistics and not interpreted and analysed by a human, like it is the case for qualitative insights. Consequently, often huge efforts are undertaken to actually validate means for quantitative measures, e.g., the NASA TLX questionnaire to assess and quantify a human's task load.

In the field of human-computer interaction, transferability and validity are less critical than for observations in clinical settings, where a lack of validity could potentially be life-threatening. Thus, HCI researchers often simply apply a novel observation method in uncritical study settings. Thereby, the participant observation is typically backed by another, established observation technique. After the study, the experiences and impressions with the novel observation technique are shared and discussed with the community. Each individual, empirical contribution shows that a technique is transferable to a given setting, and a comparison of the finding's with those from an established methodology allow to argue about the validity of the insights. In addition, each empirical trial of a technique helps to further shape and assess the advantages and disadvantages of a novel methodology. Eventually, the sum of all empirical contributions allow researchers to get a holistic understanding of a methodology that is not based on scientific evidence but on experience.

In this thesis, we modified existing observation techniques to a limited extent. Thus, the effect on a methodology's validity and transferability is likely limited as well. Nevertheless, we followed the empirical approach outlined above and, consequently, studied each of our implemented modifications in an appropriate field study setting. For each of the modifications, we provide a detailed discussion, where we analyse and share the individual limitations, advantages, and drawbacks that we observed. Thereby, we created a foundation for future empirical applications and discussions of our method approach.

7.4 Future Work

In our work, we investigated how a specific method approach affects a single, selected observation technique. We did this with two conceptionally different approaches, i.e., *smart* and *hybrid observation*, and with three of the most prominent, unsupervised observation techniques. Consequently, we are unable to argue that a certain approach works for any other than the studied techniques, which certainly is something to be addressed in future work. However, many of the made qualitative insights already suggest that also other combinations of method approaches and observation techniques could result in valuable benefits.

Consequently, future work should study which of our ideas might be helpful to further improve the scalability, situatedness, and obtrusiveness of other observation approaches. For example, we can imagine that an in-situ reflection approach (as presented in Chapter 5) can be valuable to further refine the situatedness of the Experience Sampling Method. In fact, we argue that an enrichment of ESM inquiries with earlier recorded answers and insights could significantly increase the method's situatedness.

Future work should additionally extend the idea of hybrid observation techniques, because we found this approach to be particularly helpful and valuable in several regards. We imagine that this approach would benefit from the integration of shared and distributed editing and playback. Thus, an experimenter could already see the data, while the study is still running. Further, study participants could use better suited devices to potentially elaborate on recorded insights, e.g., a summary on a day's data could be created using a tablet computer. With this modification an iterative and interactive hybrid observation approach could be created, which likely further improves situatedness, scalability, and unobtrusiveness.

Figures

Tables

Bibliography

[AB04] Adamczyk, Piotr D. ; Bailey, Brian P.: If not now, when?: the effects of interruption at different moments within task execution. In: *Proceedings of CHI*, 2004

[AFH07] Avrahami, Daniel ; Fogarty, James ; Hudson, Scott E.: Biases in human estimation of interruptibility: effects and implications for practice. In: *Proceedings of CHI*, 2007

[AM00] Abowd, Gregory D. ; Mynatt, Elizabeth D.: Charting Past, Present, and Future Research in Ubiquitous Computing. In: *ACM Trans. Comput.-Hum. Interact.* 7 (2000), Nr. 1, S. 29–58

[BB01] Barrett, Lisa F. ; Barrett, Daniel J.: An Introduction to Computerized Experience Sampling in Psychology. In: *Social Science Computer Review* 19 (2001), Nr. 2, S. 175–185

[BDR03] Bolger, Niall ; Davis, Angelina ; Rafaeli, Eshkol: DIARY METHODS: Capturing Life as it is Lived. In: *Annual Reviews of Psychology* 54 (2003), S. 579–616

[BGHS09] Bruun, Anders ; Gull, Peter ; Hofmeister, Lene ; Stage, Jan: Let your users do the testing: a comparison of three remote asynchronous usability testing methods. In: *Proceedings of CHI*, 2009

[BI08] Bailey, Brian P. ; Iqbal, Shamsi T.: Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. In: *ACM Trans. Comput.-Hum. Interact.* 14 (2008), Nr. 4

[Bon11] Bonsignore, Elizabeth: Sharing stories "in the wild": a mobile storytelling case study. In: *Proceedings of CHI EA*, 2011

[Bro96] Brooke, J.: SUS: A quick and dirty usability scale. In: Jordan, P. W. (Hrsg.) ; Weerdmeester, B. (Hrsg.) ; Thomas, A. (Hrsg.) ; Mclelland, I. L. (Hrsg.): *Usability evaluation in industry.* Taylor and Francis, 1996

[BSO00] Brown, Barry A. T. ; Sellen, Abigail J. ; O'Hara, Kenton P.: A diary study of information capture in working life. In: *Proceedings of CHI*, 2000

[BWK07] Brandt, Joel ; Weiss, Noah ; Klemmer, Scott R.: txt 4 l8r: lowering the burden for diary studies under mobile conditions. In: *Proceedings of CHI EA*, 2007

[BYJS05] Barnard, Leon ; Yi, Ji S. ; Jacko, Julie A. ; Sears, Andrew: An empirical comparison of use-in-motion evaluation scenarios for mobile computing devices. In: *International Journal of Human-Computer Studies* 62 (2005), S. 487 – 520

[CBC+08] Choudhury, Tanzeem ; Borriello, Gaetano ; Consolvo, Sunny ; Haehnel, Dirk ; Harrison, Beverly ; Hemingway, Bruce ; Hightower, Jeffrey ; Klasnja, Predrag ; Koscher, Karl ; LaMarca, Anthony ; Landay, James A. ; LeGrand, Louis ; Lester, Jonathan ; Rahimi, Ali ; Rea, Adam ; Wyatt, Danny: The Mobile Sensing Platform: An Embedded Activity Recognition System. In: *IEEE Pervasive Computing* 7 (2008), Nr. 2, S. 32–41

[CBG+06] Crabtree, Andy ; Benford, Steve ; Greenhalgh, Chris ; Tennent, Paul ; Chalmers, Matthew ; Brown, Barry: Supporting ethnographic studies of ubiquitous computing in the wild. In: *Proceedings of DIS*, 2006

[CDM+00] Cheverst, Keith ; Davies, Nigel ; Mitchell, Keith ; Friday, Adrian ; Efstratiou, Christos: Developing a context-aware electronic tourist guide: some issues and experiences. In: *Proceedings of CHI*, 2000

[CL87] Csikszentmihalyi, Mihaly ; Larson, Reed: Validity and Reliability of the Experience-Sampling Method. In: *Journal of Nervous and Mental Disease* (1987)

[CM05] Carter, Scott ; Mankoff, Jennifer: When participants do the capturing: the role of media in diary studies. In: *Proceedings of CHI*, 2005

[CMH07] Carter, Scott ; Mankoff, Jennifer ; Heer, Jeffrey: Momento: support for situated ubicomp experimentation. In: *Proceedings of CHI*, 2007

[CMT+08] Consolvo, Sunny ; McDonald, David W. ; Toscos, Tammy ; Chen, Mike Y. ; Froehlich, Jon ; Harrison, Beverly ; Klasnja, Predrag ; LaMarca, Anthony ; LeGrand, Louis ; Libby, Ryan ; Smith, Ian ; Landay, James A.: Activity sensing in the wild: a field trial of ubifit garden. In: *Proceedings of CHI*, 2008

[CRB+10] Cramer, Henriette ; Rost, Mattias ; Belloni, Nicolas ; Bentley, Frank ; Chincholle, Didier: Research in the Large. Using App Stores, Markets, and Other Wide Distribution Channels in Ubicomp Research. In: *Adjunct Proceedings of UbiComp*, 2010

[CSM+05] Consolvo, Sunny ; Smith, Ian E. ; Matthews, Tara ; LaMarca, Anthony ; Tabert, Jason ; Powledge, Pauline: Location disclosure to social relations: why, when, & what people want to share. In: *Proceedings of CHI*, 2005

[CW03] Consolvo, Sunny ; Walker, Miriam: Using the Experience Sampling Method to Evaluate Ubicomp Applications. In: *IEEE Pervasive Computing* 2 (2003), Nr. 2, S. 24–31

[DAS01] Dey, Anind K. ; Abowd, Gregory D. ; Salber, Daniel: A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-aware Applications. In: *Hum.-Comput. Interact.* 16 (2001), Nr. 2, S. 97–166

[DK03] Duckham, Matt ; Kulik, Lars: "Simplest" Paths: Automated Route Selection for Navigation. In: *Proceedings of COSIT*, 2003

[ELF93] Eldridge, Margery ; Lamming, Michael ; Flynn, Mike: Does a video diary help recall? In: *Proceedings of HCI*, 1993

[EML+10] Eisenman, Shane B. ; Miluzzo, Emiliano ; Lane, Nicholas D. ; Peterson, Ronald A. ; Ahn, Gahng-Seop ; Campbell, Andrew T.: BikeNet: A mobile sensing system for cyclist experience mapping. In: *ACM Trans. Sen. Netw.* 6 (2010), S. 1–39

[FCC+07] Froehlich, Jon ; Chen, Mike Y. ; Consolvo, Sunny ; Harrison, Beverly ; Landay, James A.: MyExperience: a system for in situ tracing and capturing of user feedback on mobile phones. In: *Proceedings of MobiSys*, 2007

[FGB11] Fischer, Joel E. ; Greenhalgh, Chris ; Benford, Steve: Investigating episodes of mobile phone activity as indicators of opportune moments to deliver notifications. In: *Proceedings of MobileHCI*, 2011

[FHA+05] Fogarty, James ; Hudson, Scott E. ; Atkeson, Christopher G. ; Avrahami, Daniel ; Forlizzi, Jodi ; Kiesler, Sara ; Lee, Johnny C. ; Yang, Jie: Predicting human interruptibility with sensors. In: *ACM Trans. Comput.-Hum. Interact.* 12 (2005), Nr. 1

[FHL04] Fogarty, James ; Hudson, Scott E. ; Lai, Jennifer: Examining the robustness of sensor-based statistical models of human interruptibility. In: *Proceedings of CHI*, 2004

[FMZ11] Ferrari, Laura ; Mamei, Marco ; Zambonelli, Franco: "All-about" Diaries: Concepts and Experiences. In: *Proceedings of COMSWARE*, 2011

[FWPM10] Feige, Sebastian ; Wenig, Dirk ; Pantel, Christoph ; Malaka, Rainer: Image-based cycle route generation on mobile devices. In: *Proceedings of MobileHCI*, 2010

[GBL06] Gemmell, Jim ; Bell, Gordon ; Lueder, Roger: MyLifeBits: A Personal Database for Everything. In: *Commun. ACM* 49 (2006), Nr. 1, S. 88–95

[GBPW04] Gaver, William W. ; Boucher, Andrew ; Pennington, Sarah ; Walker, Brendan: Cultural probes and the value of uncertainty. In: *interactions* 11 (2004), S. 53–56

[GDP99] Gaver, Bill ; Dunne, Tony ; Pacenti, Elena: Design: Cultural probes. In: *interactions* 6 (1999), S. 21–29

[GDS+10] Gerken, Jens ; Dierdorf, Stefan ; Schmid, Patric ; Sautner, Alexandra ; Reiterer, Harald: Pocket Bee: A Multi-modal Diary for Field Research. In: *Proceedings of NordiCHI*, 2010

[GK13] Gouveia, Rúben ; Karapanos, Evangelos: Footprint tracker: supporting diary studies with lifelogging. In: *Proceedings of CHI*, 2013

[HCKN96] Hartson, H. R. ; Castillo, José C. ; Kelso, John ; Neale, Wayne C.: Remote Evaluation: The Network As an Extension of the Usability Laboratory. In: *Proceedings of CHI*, 1996

[HCN10] Hinze, Annika M. ; Chang, Carole ; Nichols, David M.: Contextual queries express mobile information needs. In: *Proceedings of MobileHCI*, 2010

[HFG+98] Harrison, Beverly L. ; Fishkin, Kenneth P. ; Gujar, Anuj ; Mochon, Carlos ; Want, Roy: Squeeze me, hold me, tilt me! An exploration of manipulative user interfaces. In: *Proceedings of CHI*, 1998

[HHS09] Helmes, John ; Hummels, Caroline ; Sellen, Abigail: The Other Brother: Re-experiencing Spontaneous Moments from Domestic Life. In: *Proceedings of TEI*, 2009

[HHWL01] Hong, Jason ; Heer, Jeffrey ; Waterson, Sarah ; Landay, James A.: WebQuilt: A Proxy-based Approach to Remote Web Usability Testing. In: *ACM Trans. Inf. Syst.* 19 (2001), Nr. 3, S. 263–285

[HI05] Ho, Joyce ; Intille, Stephen S.: Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. In: *Proceedings of CHI*, 2005

[HMG02] Holland, Simon ; Morse, David R. ; Gedenryd, Henrik: AudioGPS: Spatial Audio Navigation with a Minimal Attention Interface. In: *Personal Ubiquitous Comput.* 6 (2002), Nr. 4

[HNS+10] Healey, Jennifer ; Nachman, Lama ; Subramanian, Sushmita ; Shahabdeen, Junaith ; Morris, Margaret: Out of the Lab and into the Fray: Towards Modeling Emotion in Everyday Life. In: *Proceedings of Pervasive*, 2010

[HP13] Henze, Niels ; Pielot, Martin: App Stores: External Validity for Mobile HCI. In: *interactions* 20 (2013), Nr. 2, S. 33–38

[HPB10] Henze, Niels ; Poppinga, Benjamin ; Boll, Susanne: Experiments in the Wild: Public Evaluation of Off-Screen Visualizations in the Android Market. In: *Proceedings of NordiCHI*, 2010

[HPP+11] Henze, Niels ; Pielot, Martin ; Poppinga, Benjamin ; Schinke, Torben ; Boll, Susanne: My App is an Experiment: Experience from User Studies in Mobile App Stores. In: *International Journal of Mobile Human Computer Interaction* (2011)

[HPSH00] Hinckley, Ken ; Pierce, Jeff ; Sinclair, Mike ; Horvitz, Eric: Sensing techniques for mobile interaction. In: *Proceedings of UIST*, 2000

[HR00] Hilbert, David M. ; Redmiles, David F.: Extracting Usability Information from User Interface Events. In: *ACM Comput. Surv.* 32 (2000), Nr. 4, S. 384–421

[HRB11] Henze, Niels ; Rukzio, Enrico ; Boll, Susanne: 100,000,000 Taps: Analysis and Improvement of Touch Performance in the Large. In: *Proceedings of MobileHCI*, 2011

[HRB12] Henze, Niels ; Rukzio, Enrico ; Boll, Susanne: Observational and experimental investigation of typing behaviour using virtual keyboards for mobile devices. In: *Proceedings of CHI*, 2012

[HRKS05] Hagen, Penny ; Robertson, Toni ; Kan, Melanie ; Sadler, Kirsten: Emerging research methods for understanding mobile technology use. In: *Proceedings of OZCHI*, 2005, S. 1–10

[HS88] Hart, S. G. ; Stavenland, L. E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock, P. A. (Hrsg.) ; Meshkati, N. (Hrsg.): *Human Mental Workload*. Elsevier, 1988, Kapitel 7, S. 139–183

[HWB+06] Hodges, Steve ; Williams, Lyndsay ; Berry, Emma ; Izadi, Shahram ; Srinivasan, James ; Butler, Alex ; Smyth, Gavin ; Kapur, Narinder ; Wood, Ken: SenseCam: A retrospective memory aid. In: *Proceedings of Ubicomp*, 2006

[Hyl06] Hyldegård, Jette: Using diaries in group based information behavior research: a methodological study. In: *Proceedings of IIiX*, 2006

[IB08] Iqbal, Shamsi T. ; Bailey, Brian P.: Effects of intelligent notification management on users and their tasks. In: *Proceedings of CHI*, 2008

[IKM02] Intille, Stephen ; Kukla, Charles ; Ma, Xiaoyi: Eliciting user preferences using image-based experience sampling and reflection. In: *Proceedings of CHI EA*, 2002

[IRK+03] Intille, Stephen S. ; Rondoni, John ; Kukla, Charles ; Ancona, Isabel ; Bao, Ling: A context-aware experience sampling tool. In: *Proceedings of CHI EA*, 2003

[ISO10] ISO 9241-210:2010: Human-centred design for interactive systems / International Organization for Standardization. 2010. – Forschungsbericht

[JL07] Jensen, Kasper L. ; Larsen, Lars B.: Evaluating the usefulness of mobile services based on captured usage data from longitudinal field trials. In: *Proceedings of Mobility*, 2007

[KCS+12] Kjeldskov, Jesper ; Cheverst, Keith ; Sá, Marco de ; Jones, Matt ; Murray-Smith, Roderick: Research methods in Mobile HCI: trends and opportunities. In: *Proceedings of MobileHCI*, 2012

[KDB+11] Kelly, Paul ; Doherty, Aiden ; Berry, Emma ; Hodges, Steve ; Batterham, Alan M. ; Foster, Charlie: Can we use digital life-log images to investigate active and sedentary travel behaviour? Results from a pilot study. In: *International Journal of Behavioral Nutrition and Physical Acticity* 8 (2011)

[KG03] Kjeldskov, Jesper ; Graham, Connor: A review of mobile HCI research methods. In: *Proceedings of MobileHCI*, 2003

[KH08] Kapoor, Ashish ; Horvitz, Eric: Experience sampling for building predictive user models: a comparative study. In: *Proceedings of CHI*, 2008

[KHL+08] Klasnja, Predrag ; Harrison, Beverly L. ; LeGrand, Louis ; LaMarca, Anthony ; Froehlich, Jon ; Hudson, Scott E.: Using wearable sensors and real time inference to understand human recall of routine activities. In: *Proceedings of UbiComp*, 2008

[KKO12] Kuhna, Mikko ; Kivelä, Ida-Maria ; Oittinen, Pirkko: Semi-automated Magazine Layout Using Content-based Image Features. In: *Proceedings of MM*, 2012

[KKS+04] Kahneman, Daniel ; Krueger, Alan B. ; Schkade, David A. ; Schwarz, Norbert ; Stone, Arthur A.: A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. In: *Science* 306 (2004), Nr. 5702, S. 1776–1780

[KME+08] Khan, Vassilis-Javed ; Markopoulos, Panos ; Eggen, Berry ; IJsselsteijn, Wijnand ; de Ruyter, Boris: Reconexp: A Way to Reduce the Data Loss of the Experiencing Sampling Method. In: *Proceedings of MobileHCI*, 2008

[Koe10] Koenig, Wolfgang: *Konzipierung und Entwicklung eines Aktivitätsmonitors zur Unterstützung einer Evaluation im mobilen Anwendungsszenario*. Diploma thesis, 2010

[KP12] Kjeldskov, Jesper ; Paay, Jeni: A Longitudinal Review of Mobile HCI Research Methods. In: *Proceedings of MobileHCI*, 2012

[Kru09] Krumm, John: *Ubiquitous Computing Fundamentals*. 1st. Chapman & Hall/CRC, 2009

[KS03] Kern, Nicky ; Schiele, Bernt: Context-Aware Notification for Wearable Computing. In: *Proceedings of ISWC*, 2003

[KSAH04] Kjeldskov, Jesper ; Skov, Mikael B. ; Als, Benedikte S. ; Høegh, Rune T.: Is It Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field. In: *Proceedings of MobileHCI*, 2004

[KSWK10] Kalnikaite, Vaiva ; Sellen, Abigail ; Whittaker, Steve ; Kirk, David: Now let me see where i was: understanding how lifelogs mediate memory. In: *Proceedings of CHI*, 2010

[KVVVM10] Kärkkäinen, Tuula ; Vaittinen, Tuomas ; Väänänen-Vainio-Mattila, Kaisa: I Don't Mind Being Logged, but Want to Remain in Control: A Field Study of Mobile Activity and Context Logging. In: *Proceedings of CHI*, 2010

[Lan58] Landsberger, H.A.: *Hawthorne revisited: Management and the worker: its critics, and developments in human relations in industry.* Cornell University, 1958 (Cornell Studies Series)

[LC83] Larson, Reed ; Csikszentmihalyi, Mihaly: The experience sampling method. In: *New Directions for Methodology of Social and Behavioral Science* 15 (1983), S. 41–56

[LDF11] Li, Ian ; Dey, Anind K. ; Forlizzi, Jodi: Understanding My Data, Myself: Supporting Self-reflection with Ubicomp Technologies. In: *Proceedings of UbiComp*, 2011

[LG06] Landry, Brian M. ; Guzdial, Mark: iTell: supporting retrospective storytelling with digital photos. In: *Proceedings of DIS*, 2006

[LOHI05] Lin, Tao ; Omata, Masaki ; Hu, Wanhua ; Imamiya, Atsumi: Do Physiological Data Relate to Traditional Usability Indexes? In: *Proceedings of OZCHI*, 2005

[LSH+06] Lindström, Madelene ; Ståhl, Anna ; Höök, Kristina ; Sundström, Petra ; Laaksolathi, Jarmo ; Combetto, Marco ; Taylor, Alex ; Bresin, Roberto: Affective diary: designing for bodily expressiveness and self-reflection. In: *Proceedings of CHI EA*, 2006

[Mac07] Macefield, Ritch: Usability Studies and the Hawthorne Effect. In: *Journal of Usability Studies* 2 (2007), Nr. 3, S. 145–154

[May33] Mayo, Elton: *The human problems of an industrial civilization.* Macmillan Co, NY, 1933

[MBIC13] Martín, Henar ; Bernardos, Ana M. ; Iglesias, Josué ; Casar, José R.: Activity Logging Using Lightweight Classification Techniques in Mobile Devices. In: *Personal Ubiquitous Comput.* 17 (2013), Nr. 4, S. 675–695

[MBP+12] Mark, Gloria ; Bagdouri, Mossaab ; Palen, Leysia ; Martin, James ; Al-Ani, Ban ; Anderson, Kenneth: Blogs As a Collective War Diary. In: *Proceedings of CSCW*, 2012

[McD05] McDonald, Seonaidh: Studying actions in context: a qualitative shadowing method for organizational research. In: *Qualitative Research* 5 (2005), Nr. 4, S. 455–473

[McF02] McFarlane, Daniel: Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. In: *Hum.-Comput. Interact.* 17 (2002), Nr. 1

[MG11] Millonig, A. ; Gartner, G.: Identifying motion and interest patterns of shoppers for developing personalised wayfinding tools. In: *J. Locat. Based Serv.* 5 (2011), Nr. 1

[MGK08] Mark, Gloria ; Gudith, Daniela ; Klocke, Ulrich: The cost of interrupted work: more speed and stress. In: *Proceedings of CHI*, 2008

[MHK05] Mulder, I. ; Hofte, H. ter ; Kort, J.: SocioXensor: Measuring user behaviour and user eXperience in ConteXt with mobile devices. In: *Proceedings of MB (Measuring Behaviour)*, 2005

[ML02] McFarlane, Daniel C. ; Latorella, Kara A.: The scope and importance of human interruption in human-computer interaction design. In: *Hum.-Comput. Interact.* 17 (2002), Nr. 1

[MM10] Millonig, Alexandra ; Maierbrugger, Gudrun: Using semi-automated shadowing for analysing stress-induced spatio-temporal behaviour patterns of passengers in public transport infrastructures. In: *Proceedings of MB (Measuring Behaviour)*, 2010

[MMB+10] McMillan, Donald ; Morrison, Alistair ; Brown, Owain ; Hall, Malcolm ; Chalmers, Matthew: Further into the Wild: Running Worldwide Trials of Mobile Systems. In: *Proceedings of Pervasive*, 2010

[MP11] Meyer, Jochen ; Poppinga, Benjamin: Distributed Personal Archives: Deriving Health Profiles from the Web. In: *Proceedings of 'Workshop on Information Access for Personal Media Archives' in conjunction with ECIR*, 2011

[MRGS10a] Magnusson, Charlotte ; Rassmus-Gröhn, K. ; Szymczak, Delphine: Scanning angles for directional pointing. In: *Proceedings of MobileHCI*, 2010

[MRGS10b] Magnusson, Charlotte ; Rassmus-Gröhn, K. ; Szymczak, Delphine: The influence of angle size in navigation applications using pointing gestures. In: *Proceedings of Haptic and Audio Interaction Design (HAID)*, 2010

[MRT10] Meschtscherjakov, Alexander ; Reitberger, Wolfgang ; Tscheligi, Manfred: MAESTRO: orchestrating user behavior driven and context triggered experience sampling. In: *Proceedings of MB (Measuring Behaviour)*, 2010

[MTC06] MORRISON, Allistair ; TENNENT, Paul ; CHALMERS, Matthew: Coordinated Visualisation of Video and System Log Data. In: *Proceedings of CMV*, 2006

[Nie12] NIEHUES, Tobias: *PhysioLog: Betrachtung von Fußgängernavigation mit Hilfe von Sensordaten*. Bachelor thesis, 2012

[NKLM11] NACKE, Lennart E. ; KALYN, Michael ; LOUGH, Calvin ; MANDRYK, Regan L.: Biofeedback Game Design: Using Direct and Indirect Physiological Control to Enhance Game Interaction. In: *Proceedings of CHI*, 2011

[NOP+06] NIELSEN, Christian M. ; OVERGAARD, Michael ; PEDERSEN, Michael B. ; STAGE, Jan ; STENILD, Sigge: It's worth the hassle!: the added value of evaluating the usability of mobile systems in the field. In: *Proceedings of NordiCHI*, 2006

[Oeh12] OEHMCKE, Stefan: *Storyteller: Ein unaufdringliches Werkzeug zum Erfassen von subjektiven Eindrücken*. Bachelor thesis, 2012

[OTRK05] OULASVIRTA, Antti ; TAMMINEN, Sakari ; ROTO, Virpi ; KUORELAHTI, Jaana: Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In: *Proceedings of CHI*, 2005

[PB10a] POPPINGA, Benjamin ; BOLL, Susanne: Activity Recognition Research is More Than Finding the Ultimate Algorithms or Parameters. In: *Proceedings of 'How to do good activity recognition research?' workshop in conjunction with Pervasive*, 2010

[PB10b] POPPINGA, Benjamin ; BOLL, Susanne: An Experimenter's Third Eye: Using the SenseCam as Ground Truth for Unsupervised Evaluations. In: *Proceedings of SenseCam Symposium*, 2010

[PCB+12] POPPINGA, Benjamin ; CRAMER, Henriette ; BÖHMER, Matthias ; MORRISON, Alistair ; BENTLEY, Frank ; HENZE, Niels ; ROST, Mattias ; MICHAHELLES, Florian: Research in the Large 3.0 – App Stores, Wide Distribution, and Big Data in MobileHCI Research. In: *Proceedings of MobileHCI*, 2012

[PHB09] PIELOT, Martin ; HENZE, Niels ; BOLL, Susanne: Supporting Map-based Wayfinding with Tactile Cues. In: *Proceedings of MobileHCI*, 2009

[PHB14a] POPPINGA, Benjamin ; HEUTEN, Wilko ; BOLL, Susanne: NavMem Explorer: An Orientation Aid for People with Mild Cognitive Impairments. In: *Proceedings of PervasiveHealth*, 2014

[PHB14b] Poppinga, Benjamin ; Heuten, Wilko ; Boll, Susanne: Sensor-based Identification of Opportune Moments to Trigger Unobtrusive Notifications. In: *IEEE Pervasive Computing* 13 (2014), Nr. 1, S. 22–29

[PHF+12] Poppinga, Benjamin ; Henze, Niels ; Fortmann, Jutta ; Heuten, Wilko ; Boll, Susanne: AmbiGlasses – Information in the Periphery of the Visual Field. In: *Proceedings of Mensch & Computer (Human & Computer)*, 2012

[PMH+10] Poppinga, Benjamin ; Magnusson, Charlotte ; Heuten, Wilko ; McGookin, David ; Henze, Niels ; Claassen, Ginger ; Pielot, Martin ; Eftring, Håkan ; Peters, Jörn: Observing the Mobile User Experience. In: *Proceedings of NordiCHI*, 2010

[PMPRG11] Poppinga, Benjamin ; Magnusson, Charlotte ; Pielot, Martin ; Rassmus-Gröhn, Kirsten: TouchOver Map: Audio-Tactile Exploration of Interactive Maps. In: *Proceedings of MobileHCI*, 2011

[POHB13] Poppinga, Benjamin ; Oehmcke, Stefan ; Heuten, Wilko ; Boll, Susanne: Storyteller: In-Situ Reflection on Study Experiences. In: *Proceedings of MobileHCI*, 2013

[PPB09a] Pielot, Martin ; Poppinga, Benjamin ; Boll, Susanne: Understanding tourists on a bicycle trip “in the wild”. In: *Proceedings of ‘Mobile Living Labs’ workshop in conjunction with MobileHCI*, 2009

[PPB09b] Poppinga, Benjamin ; Pielot, Martin ; Boll, Susanne: Tacticycle: A Tactile Display for Supporting Tourists on a Bicycle Trip. In: *Proceedings of MobileHCI*, 2009

[PPB10] Pielot, Martin ; Poppinga, Benjamin ; Boll, Susanne: Pocket-Navigator: Vibro-Tactile Waypoint Navigation for Everyday Mobile Devices. In: *Proceedings of MobileHCI*, 2010

[PPB11] Pielot, Martin ; Poppinga, Benjamin ; Boll, Susanne: A Tactile Friend Sense for Keeping Groups Together. In: *Proceedings of CHI EA*, 2011

[PPH+11] Pielot, Martin ; Poppinga, Benjamin ; Heuten, Wilko ; Schang, Jeschua ; Boll, Susanne: A Tactile Compass for Eyes-free Pedestrian Navigation. In: *Proceedings of INTERACT*, 2011

[PPHB11a] Pielot, Martin ; Poppinga, Benjamin ; Heuten, Wilko ; Boll, Susanne: 6th Senses for Everyone! The Value of Multimodal Feedback in Handheld Navigation Aids. In: *Proceedings of ICMI*, 2011

[PPHB11b] Poppinga, Benjamin ; Pielot, Martin ; Henze, Niels ; Boll, Susanne: Unsupervised User Observation in the App Store: Experiences with the Sensor-based Evaluation of a Mobile Pedestrian Navigation Application. In: *Wi: Journal of Mobile Media* (2011)

[PPHB11c] Poppinga, Benjamin ; Pielot, Martin ; Heuten, Wilko ; Boll, Susanne: Evaluating Mobile Accessible Applications is a Challenge: Can the Virtual Observer be a Proper Solution? In: *Proceedings of 'Mobile Accessibility' workshop in conjunction with INTERACT*, 2011

[PPHB12a] Pielot, Martin ; Poppinga, Benjamin ; Heuten, Wilko ; Boll, Susanne: PocketNavigator: Studying Tactile Navigation Systems In-Situ. In: *Proceedings of CHI*, 2012

[PPHB12b] Pielot, Martin ; Poppinga, Benjamin ; Heuten, Wilko ; Boll, Susanne: Tacticycle: Supporting Exploratory Bicycle Trips. In: *Proceedings of MobileHCI*, 2012

[PPHB12c] Poppinga, Benjamin ; Pielot, Martin ; Heuten, Wilko ; Boll, Susanne: Towards an Objective Comparison of Scanning-Based Interaction Techniques. In: *Proceedings of HAID*, 2012

[PPHB14] Poppinga, Benjamin ; Pielot, Martin ; Heuten, Wilko ; Boll, Susanne: Unobtrusive Observation of Cycling Tourists in the Wild. In: *International Journal of Mobile Human Computer Interaction* 6 (2014), Nr. 4, S. 22–42

[PTHB11] Poppinga, Benjamin ; Timmermann, Janko ; Heuten, Wilko ; Boll, Susanne: Stress Aware: An Investigation of Everyday Real-Time Stress Feedback. In: *Proceedings of 'Mobile Wellness' workshop in conjunction with MobileHCI*, 2011

[PVH01] Picard, Rosalind W. ; Vyzas, Elias ; Healey, Jennifer: Toward Machine Emotional Intelligence: Analysis of Affective Physiological State. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (2001), Nr. 10, S. 1175–1191

[PVK+10] Pielot, Martin ; Vester, Björn ; Kazakova, Anastasia ; Brammer, Lennard ; Poppinga, Benjamin ; Boll, Susanne: Natch: A Watch-like Display for Less Distracting Pedestrian Navigation. In: *Proceedings of Mensch & Computer (Human & Computer)*, 2010

[RCT+07] Rogers, Yvonne ; Connelly, Kay ; Tedesco, Lenore ; Hazlewood, William ; Kurtz, Andrew ; Hall, Robert E. ; Hursey, Josh ; Toscos, Tammy: Why it's worth the hassle: the value of in-situ studies when designing Ubicomp. In: *Proceedings of UbiComp*, 2007

[RD39] Roethlisberger, Fritz J. ; Dickson, William J.: *Management and the Worker*. Harvard University Press, Cambridge, MA, 1939

[REJ09] Robinson, Simon ; Eslambolchilar, Parisa ; Jones, Matt: Sweep-Shake: finding digital resources in physical environments. In: *Proceedings of MobileHCI*, 2009

[RFO+09] Rowland, Duncan ; Flintham, Martin ; Oppermann, Leif ; Marshall, Joe ; Chamberlain, Alan ; Koleva, Boriana ; Benford, Steve ; Perez, Citlali: Ubikequitous computing: designing interactive experiences for cyclists. In: *Proceedings of MobileHCI*, 2009

[Rie93] Rieman, John: The diary study: a workplace-oriented research tool to guide laboratory efforts. In: *Proceedings of INTERACT and CHI*, 1993

[RMCB13] Rost, Mattias ; Morrison, Alistair ; Cramer, Henriette ; Bentley, Frank: Informing Future Design via Large-scale Research Methods and Big Data. In: *Proceedings of MobileHCI*, 2013

[Rob08] Robertson, Shanthi K.: Cultural Probes in Transmigrant Research: A Case Study. In: *InterActions: UCLA Journal of Education and Information Studies* 4 (2008), Nr. 2

[Rob10] Robinson, Simon: Heads-up engagement with the real world: multimodal techniques for bridging the physical-digital divide. In: *Proceedings of CHI*, 2010

[Rob11] Robson, Colin: *Real World Research*. Wiley, 2011

[ROPT05] Raento, Mika ; Oulasvirta, Antti ; Petit, Renaud ; Toivonen, Hannu: ContextPhone: A Prototyping Platform for Context-Aware Mobile Applications. In: *IEEE Pervasive Computing* 4 (2005), Nr. 2, S. 51–59

[RRH11] Rümelin, Sonja ; Rukzio, Enrico ; Hardy, Robert: NaviRadar: A Novel Tactile Information Display for Pedestrian Navigation. In: *Proceedings of UIST 2011*, 2011

[RS99] Roddick, John F. ; Spiliopoulou, Myra: A bibliography of temporal, spatial and spatio-temporal data mining research. In: *SIGKDD Explor. Newsl.* 1 (1999), Nr. 1

[RSD+10] Reddy, Sasank ; Shilton, Katie ; Denisov, Gleb ; Cenizal, Christian ; Estrin, Deborah ; Srivastava, Mani: Biketastic: sensing and mapping for better biking. In: *Proceedings of CHI*, 2010

[SDA99] Salber, Daniel ; Dey, Anind K. ; Abowd, Gregory D.: The Context Toolkit: Aiding the Development of Context-enabled Applications. In: *Proceedings of CHI*, 1999

[SFA+07] Sellen, Abigail J. ; Fogg, Andrew ; Aitken, Mike ; Hodges, Steve ; Rother, Carsten ; Wood, Ken: Do life-logging technologies support memory for the past?: an experimental study using sensecam. In: *Proceedings of CHI*, 2007

[SFR+13] Sas, Corina ; Fratczak, Tomasz ; Rees, Matthew ; Gellersen, Hans ; Kalnikaite, Vaiva ; Coman, Alina ; Höök, Kristina: AffectCam: Arousal- Augmented Sensecam for Richer Recall of Episodic Memories. In: *Proceedings of CHI EA*, 2013

[SHS+09] Ståhl, Anna ; Höök, Kristina ; Svensson, Martin ; Taylor, AlexS. ; Combetto, Marco: Experiencing the Affective Diary. In: *Personal and Ubiquitous Computing* 13 (2009), Nr. 5, S. 365–378

[SLGH08] Sohn, Timothy ; Li, Kevin A. ; Griswold, William G. ; Hollan, James D.: A diary study of mobile information needs. In: *Proceedings of CHI*, 2008

[SM13] Singh, Anjeli ; Malhotra, Sareeka: A Researcher's Guide to Running Diary Studies. In: *Proceedings of APCHI (Asia Pacific Conference on Computer Human Interaction)*, 2013

[SRK07] Stark, A. ; Riebeck, M. ; Kawalek, J.: How to Design an Advanced Pedestrian Navigation System: Field Trial Results. In: *Proceedings of IDAACS (Workshop on Intelligent Data Acquisition and Advanced Computing Systems)*, 2007

[SRP07] Sharp, Helen ; Rogers, Yvonne ; Preece, Jenny: *Interaction Design: Beyond Human-Computer Interaction*. 2nd. Wiley, 2007

[SSR+11] Schleicher, Robert ; Shirazi, Alireza S. ; Rohs, Michael ; Kratz, Sven ; Schmidt, Albrecht: WorldCupinion Experiences with an Android App for Real-Time Opinion Sharing During Soccer World Cup Games. In: *Int. J. Mob. Hum. Comput. Interact.* 3 (2011), Nr. 4, S. 18–35

[Str12] Strunck, Marco: *Situational Awareness als Optimierungsansatz für die Experience Sampling Method*. Bachelor thesis, 2012

[TFA10] Tintarev, Nava ; Flores, Ana ; Amatriain, Xavier: Off the beaten track: a mobile field study exploring the long tail of tourist recommendations. In: *Proceedings of MobileHCI*, 2010

[Var07] Varshney, Upkar: Pervasive Healthcare and Wireless Health Monitoring. In: *Mob. Netw. Appl.* 12 (2007), Nr. 2-3, S. 113–127

[VKO10] Vaittinen, Tuomas ; Kärkkäinen, Tuula ; Olsson, Thomas: A Diary Study on Annotating Locations with Mixed Reality Information. In: *Proceedings of MUM*, 2010

[VLR+10] Vermeeren, Arnold P. O. S. ; Law, Effie Lai-Chong ; Roto, Virpi ; Obrist, Marianna ; Hoonhout, Jettie ; Väänänen-Vainio-Mattila, Kaisa: User experience evaluation methods: current state and development needs. In: *Proceedings of NordiCHI*, 2010

[WBM12] Webber, Emily ; Burnett, Gary ; Morley, Jeremy: Pedestrian navigation with a mobile device: strategy use and environmental learning. In: *Proceedings of BCS-HCI*, 2012

[Wei91] Weiser, Marc: The Computer for the 21st Century. In: *Scientific American Special Issue on Communications, Computers, and Networks* (1991)

[WSK+11] Wirz, Martin ; Schlaepfer, Pablo ; Kjaergaard, Mikkel B. ; Roggen, Daniel ; Feese, Sebastian ; Troester, Gerhard: Towards an online detection of pedestrian flocks in urban canyons by smoothed spatio-temporal clustering of GPS trajectories. In: *Proceedings of International Workshop on Location-Based Social Networks*, 2011

[YJJA07] Youngho, Rhee ; Jennifer, Shim ; Junghun, Kim ; Amy, Chung: Create, Sync, Share: The Role of Mobile Device on Social Entity. In: *Proceedings of MobileHCI*, 2007

[Zer12] Zerhusen, Stephan: *Konzeption und Entwicklung einer optimierten Kartendarstellung für Radfahrer*. Diploma thesis, 2012