Investigating the Perceptibility of Smartphone Notifications and Methods for Context-Aware Data Assessment in Experience Sampling Studies

In Fulfillment of the Requirements for the Degree of

Doctor of Engineering

Faculty of Informatics
Karlsruhe Institute of Technology (KIT)

Accepted

Dissertation

by

Anja Exler

Born in Blankenburg-Harz, Germany

Examination Date:	24 Oct 2018
Primary Examiner:	Prof. Dr.-Ing. Michael Beigl
Secondary Examiner:	Prof. Dr. rer. nat. Niels Henze

"Hey! Listen! Link! Listen! Hey! Link! Listen!"

Navi (The Legend of Zelda - Ocarina of Time)

Abstract

A central task in human computer interaction is to conduct user studies. User studies might serve to gain deeper insights into the behavior of users as well as to collect labels to annotate data. The traditional method to collect subjective feedback is the experience sampling method (ESM). By answering questionnaires, participants provide not only information about themselves and their environment. Their responses can also serve as labels for data that was collected at the same time. By now, the smartphone became the standard platform to conduct ESM studies. It is used to deliver ESM prompts for feedback through notifications, to store the collected labels, and to assign them to sensor measurements that were gathered in the background.

In experience sampling, researchers desire to collect rich sets of high-quality data. Achieving this goal requires cooperation and commitment of study participants. Study participants, in contrast, usually want to invest as least effort as possible, i.e., to receive as few prompts as possible.

Different challenges arise that the researcher has to face when designing an ESM study. On the one hand, these challenges are connected to the experience sampling app and its functionality. On the other hand, they relate to the delivery of feedback prompts and the user's perception of these prompts.

Feedback prompts must be triggered in situations of interest, requiring accurate context recognition systems embedded into the ESM app. The number and frequency of feedback prompts as well as the length of the questionnaire need to be kept to a reasonable minimum – finding a compromise between having enough data to answer a research question while not overstraining the study participant with prompts. Both are challenges that need to be addressed by the ESM app that is used to conduct the user study.

To facilitate the configuration of ESM apps, it is advisable to have one main development tool. In the best case, such a tool is easy-to-use and provides access to a large set of sensors from which further contextual information can be inferred, e.g., to trigger prompts event-based. Within this dissertation, we introduce ESMAC, the ESM App Configurator. ESMAC provides different prompting modalities and settings to restrict the number of prompts per day (inquiry limit) or to define a prompt-free time window between two successive prompts (inter-notification time). In addition, it offers access to a wide range of sensor measurements and information. This information is assessed automatically and does not need to be asked for in the questionnaire, leading to a reduced questionnaire length. To assess information in situation of interest, ESMAC offers a variety of event prompts.

Though applied in several ESM studies in one way or another, the usefulness of event prompts was not explicitly investigated in literature. Two factors of interest are location and activity changes which are of relevance, e.g., for interruptibility detection in computer science and for monitoring state changes of patients suffering from affective disorders in applied psychology. Exemplified on a user study focusing on these factors, we show that event-based prompts are useful, especially if the implemented event-triggers (here: location change) relate to the kind of data that is to be collected (here: feedback about user mobility and activity).

The assessment of data labels does not only require event-triggered prompts, but also timely responses from the participants to allow as accurate assignment of labels to data as possible. This requires participants to recognize incoming prompts. Prompts might not be perceived due to an insufficient notification modality or because they drown in the flood of notifications displayed in the smartphone's notification drawer.

The perceptibility of smartphone notifications is influenced by different contextual factors such as the smartphone position, the participant's location and (social) activity, but also by content-related features such as the notification importance. As a basis for further investigations about notification perception, we examine methods to assess these related factors. First, we present a method to improve the detection of specific smartphone positions by running a position transition correction that is based on the assumption that the hand state is a necessary transition between other positions. Next, we investigate different privacy-sensitive alternatives for location assessment. We present how WiFi and place types can be used to describe a user's location and to detect location changes. Related to the assessment of social activity, we present a location-based method to estimate if a smartphone user is in company or not. Eventually, we investigate smartphone features that relate to the perceived importance of smartphone notifications.

After investigating methods to assess influencing factors, we examine relations between the user's perception of incoming notifications and different notification modalities depending on (a) the smartphone position and (b) the location and location-based activity. We present study results indicating how pleasant and perceptible different notification modalities are depending on the smartphone position. For location and location-based activities, we recommend suitable notification modalities based on feedback gained in an online survey and a laboratory study.

Finally, we investigate and evaluate different designs to highlight important notifications - including feedback prompts – to increase their perceptibility within the notification drawer. These designs are based on feedback from interview participants as well as inferred from literature. We present properties of notification designs that were perceived pleasant and useful by survey participants and we recommend designs combining different characteristics.

In summary, this dissertation presents the following contributions:

- Introduction of a tool to build context-aware ESM apps
- Confirmation of the relevance of event-triggers for an exemplary ESM study focusing on location and activity changes
- Presentation of a position transition correction mechanism to improve the detection of smartphone positions
- Presentation of two privacy-sensitive methods to assess a user's current location
- Presentation of a location-based method to estimate if a smartphone user is in company or not
- Introduction of four kinds of importance and presentation of smartphone features that relate to the perceived importance of smartphone notifications
- Recommendations for the selection of suitable notifications modalities based on (a) the smartphone position or (b) the current location and possible location-based activities
- Recommendations for design adaptions and customization options for smartphone notifications to highlight those of higher importance

Deutsche Zusammenfassung

Eine zentrale Aufgabe in der Mensch-Maschine-Interaktion ist die Durchführung von Nutzerstudien. Diese ermöglichen einen tieferen Einblick in das Verhalten von Nutzern, dienen aber auch dazu, Labels zum Annotieren von Daten zu sammeln. Die traditionelle Methode zum Erfassen von subjektivem Feedback ist die Experience Sampling Method (ESM). Durch das Beantworten von Fragebögen stellen Probanden nicht nur Informationen über sich selbst, sondern auch über ihre Umgebung zur Verfügung. Außerdem können ihre Antworten als Label für Daten, welche zeitgleich erhoben wurden, dienen. Inzwischen sind Smartphones zur Hauptplattform zum Durchführen von ESM Studien geworden. Sie werden genutzt, um ESM-Abfragen in Form von Benachrichtigungen auszusenden, um die gesammelten Labels zu speichern und um sie den Sensordaten zuzuweisen, welche im Hintergrund gesammelt wurden.

In ESM-Studien wird angestrebt, möglichst viele und qualitativ hochwertige Daten zu sammeln. Um dieses Ziel zu erreichen, bedarf es einer großen Menge sorgfältig beantworteter ESM-Abfragen. Die Probanden wiederum wollen in der Regel so wenig Abfragen wie möglich erhalten. Es ist notwendig, einen Kompromiss zwischen Abfragehäufigkeit und Probandenzufriedenheit zu finden.

Beim Erstellen von ESM-Studien ergeben sich verschiedene Herausforderungen. Einerseits sind diese mit der ESM-App und deren Funktionalität verbunden. Andererseits stehen sie aber auch mit dem Ausliefern von ESM-Abfragen und deren Wahrnehmung durch den Nutzer im Zusammenhang.

ESM-Abfragen müssen in Situationen ausgesandt werden, welche für den Studiendesigner von Interesse sind. Dies bedarf eines akkuraten Erkennungssystems, welches in die ESM-App eingebunden werden muss. Sowohl die Anzahl und Häufigkeit der Abfragen als auch die Länge des Feedback-Fragebogens sollten auf ein Minimum reduziert werden. Beides sind Herausforderungen, welche die ESM-App, welche zur Durchführung der Studie genutzt wird, adressieren muss.

Um das Erstellen von ESM-Anwendungen zu erleichtern, ist es empfehlenswert, auf ein primäres Entwicklungswerkzeug zurückzugreifen. Im besten Fall ist solch ein Werkzeug einfach zu nutzen und bietet Zugriff auf eine weitreichende Menge an Sensoren, aus denen kontextuelle Informationen abgeleitet werden können – beispielsweise, um ereignisbasiert Abfragen auszusenden. Im Rahmen dieser Dissertation stellen wir ESMAC vor, den ESM App Configurator. ESMAC stellt verschiedene Abfragetypen zur Verfügen, ebenso wie verschiedene Einstellungen, um die Anzahl an Abfragen pro Tag zu begrenzen (inquiry limit) oder um ein abfragefreies Zeitfenster zwischen zwei aufeinanderfolgenden Abfragen zu definieren (inter-notification time). Zudem bietet es Zugriff auf eine Vielzahl an Sensormesswerten und -Informationen. Diese Werte werden automatisch erfasst und benötigen keine Abfrage vom Nutzer, was zu einer reduzierten Fragebogenlänge führen kann. Um Informationen in Situationen zu sammeln, welche für den Studiendesigner von Interesse sind, bietet ESMAC eine Auswahl an ereignisbasierten Abfragen.

Ereignisbasierte Abfragen fanden bereits in diversen ESM-Studien Anwendung. Dennoch wurde ihre Nützlichkeit bisher nicht explizit untersucht. Zwei Faktoren, welche für verschiedene Forschungsbereiche relevant sind, sind Ortswechsel und Aktivitätsänderungen des Nutzers. Diese können beispielsweise für die Erkennung der Unterbrechbarkeit eines Nutzers genutzt werden oder zum Überwachen von Zustandsänderungen bei Patienten, welche unter affektiven Störungen leiden. Am Beispiel einer Studie, welche auf die Erfassung dieser beiden Faktoren ausgerichtet ist, zeigen wir, dass ereignisbasierte Abfragen nützlich sind, vor allem wenn die ausgewählten ereignisbasierten Abfragen (hier: Ortswechsel) im Zusammenhang mit den zu erfassenden Daten stehen (hier: Feedback über die Mobilität und Aktivität des Nutzers).

Die Erfassung von Datenlabels bedarf nicht nur ereignisbasierter Abfragen, sondern auch zeitnaher Antworten von den Probanden, um die Labels möglichst akkurat den gesammelten Daten zuweisen zu können. Hierzu ist es notwendig, dass die Probanden die eingehenden Abfragen rechtzeitig bemerken. Abfragen werden unter Umständen nicht wahrgenommen, weil eine zu unauffällige Benachrichtigungsmodalität gewählt wurde oder weil die ESM-Abfragen in einem überfüllten Notification Drawer des Smartphones untergehen.

Die Wahrnehmbarkeit von Benachrichtigungen wird durch verschiedene kontextuelle Faktoren beeinflusst, z.B. die Position des Smartphones, den aktuellen Ort oder die (soziale) Aktivität des Nutzers. Aber auch inhaltliche Eigenschaften wie die empfundene Wichtigkeit einer Benachrichtigung können einen Einfluss haben. Als Grundlage für spätere Forschung untersuchen wir Methoden, um diese Einflussfaktoren zu erfassen. Zuerst stellen wir eine Methode zur Position-Transition-Korrektur vor, welche die Erkennung der aktuellen Smartphone-Position verbessert. Diese Methode basiert auf der Annahme, dass jeder Wechsel von einer Position zur nächsten über das Halten des Geräts in der Hand erfolgt. Als nächstes untersuchen wir verschiedene Methoden zur Ortserfassung, unter Achtung der Privatsphäre des Benutzers. Wir stellen vor, wie WLAN-Informationen und Ortstypen genutzt werden können, um den Aufenthaltsort eines Nutzers zu beschreiben und Ortswechsel zu erkennen, ohne den exakten Standort abzuspeichern. Basierend auf dem Ortstypen präsentieren wir eine Methode, um abzuschätzen, ob ein Smartphone-Nutzer in Begleitung ist. Abschließend untersuchen wir noch Smartphone-Features, welche mit der empfundenen Wichtigkeit einer Benachrichtigung in Zusammenhang stehen könnten.

Nachdem wir Methoden zum Erfassen von Einflussfaktoren untersucht haben, betrachten wir Zusammenhänge zwischen der Wahrnehmung von eingehenden Benachrichtigungen und verschiedenen Benachrichtigungsmodalitäten. Diese Betrachtung erfolgt unter Berücksichtigung (a) der aktuellen Position des Smartphones und (b) des aktuellen Ortes des Smartphone-Nutzers und möglicher ortsbasierter Aktivitäten. Wir stellen eine Studie vor, welche Aufschluss darüber gibt, wie angenehm und wahrnehmbar verschiedene Benachrichtigungsmodalitäten sind – abhängig davon, wo das Smartphone vom Nutzer aufbewahrt wird. Für den aktuellen Ort und ortsbezogene Aktivitäten stellen wir passende Benachrichtigungsmodalitäten vor, über welche wir im Rahmen einer Onlineumfrage und einer Laborstudie Rückmeldung erhalten haben.

Abschließend erstellen und evaluieren wir verschiedene Designs, um wichtige Benachrichtigungen – welche ESM-Abfragen einschließen – hervorzuheben, indem ihre Sichtbarkeit im Notification Drawer erhöht wird. Diese Designs basieren auf Feedback von Interviewprobanden als auch auf Erkenntnissen aus der Literatur. Wir stellen Eigenschaften von Benachrichtigungsdesigns vor, welche von Probanden einer Onlineumfrage als angenehm und nützlich empfunden wurden. Zudem empfehlen wir auch Kombinationen verschiedener Designeigenschaften.

Die Beiträge dieser Dissertation können wie folgt zusammengefasst werden:

- Vorstellung eines Tools, um kontextsensitive ESM-Apps zu erstellen
- Bestätigung der Relevanz von ereignisbasierten Abfragen am Beispiel einer ESM-Studie mit Fokus auf Ortswechsel und Aktivitätsänderungen
- Vorstellung eines Position-Transition-Korrekturmechanismus zum Verbessern der Erkennung der Smartphone-Position
- Vorstellung zweier Methoden zur Ortserfassung ohne konkrete Offenlegung und Speicherung des konkreten Aufenthaltsortes
- Vorstellung einer ortsbasierten Methode zum Abschätzen, ob sich ein Smartphone-Nutzer in Begleitung befindet oder nicht
- Vorstellen von vier Typen von Wichtigkeit und von Smartphone-Features, welche mit der empfundenen Wichtigkeit von Benachrichtigungen in Zusammenhang stehen
- Empfehlungen für die Auswahl von Benachrichtigungsmodalitäten abhängig von der (a) Smartphone-Position als auch (b) des aktuellen Ortes und möglicher ortsbasierter Aktivitäten
- Empfehlungen für Designanpassungen von Smartphone-Benachrichtigungen, um solche von höherer Wichtigkeit hervorzuheben

Preface

This dissertation originated from the research that I conducted at the Karlsruhe Institute of Technology (KIT). My work and decisions were influenced by many conversations and discussions with colleagues, students, and external researchers working on the topics of smartphone-based experience sampling, notification management, context-aware systems, and human activity recognition, among others. As a research associate at the Karlsruhe Institute of Technology (KIT), I supervised several student projects including Bachelor and Master theses. These theses were mostly related to my research topic and supported me in investigating different PhD-related research questions. Large parts of this thesis originate from scientific papers that were co-authored by me and that originate from collaborations with students, colleagues, and external researchers. The respective chapters contain references to these publications in the introductory part of the chapter. All tables, figures, and diagrams in this paper were either made by myself, originated from theses supervised by me or scientific papers co-authored by me, or contain a reference to their sources.

Due to the collaborative background of my research, I decided to write this dissertation using the scientific plural instead of the singular.

Contents

Part I.

Introduction

1 Introduction

A central task in human computer interaction (HCI) are user studies. They can be used to collect datasets to build descriptive or predictive models, to run correlation analyses, or to gain a deeper understanding of the user. Due to their proliferation, omnipresence, and sensing capabilities, smartphones are widely considered as an assessment tool [140]. They allow the collection of measurements from physical and virtual sensors, the monitoring of user interaction with the smartphone as well as the assessment of subjective feedback from the user that can be used to label the gathered measurements. The traditional technique to assess subjective feedback is the *Experience Sampling Method* (ESM).

ESM, as described by Csikszentmihalyi and Larson [62], is a means for systematic observation of persons and their experiences in daily life [82, 185]. This includes but is no limited to the assessment of personal experiences, sentiment, thoughts, feelings, or behavior change [50, 71, 186] via self-reports. Such self-reports provide insights about daily lives of the participants in real-time without facing the retrospective nature of traditional approaches that might introduce memory errors or recall bias [87, 176, 186, 185]. This is crucial for the in-situ assessment of annotated data as it requires a correct and timely labeling. Besides the timeliness of the responses, factors that influence a user's willingness to answer self-report questionnaires – and thereby to provide data labels – are the number of prompts per day and the time interval between two prompts [162] as well as the length of the ESM questionnaire [50, 181]. Addressing these aspects might be supported by the context-awareness of current smartphones and corresponding ESM apps.

In smartphone-based ESM studies, self-reports are delivered to the smartphone user through notifications. Such notifications prompt the user to reply to a questionnaire whose items serve as labels for the data that is gathered in the background. Literature differentiates three different prompting types: random (i.e., randomly throughout the day), time-based (i.e., interval-based at specific points in time), and event-based (i.e., in case of the occurrence of a certain event) [180]. Prompting types should always be selected in accordance with the study objective. For a general overview of a person's daily life, it is reasonable to have random or time-based triggers. To assess labels to annotate data it is necessary to rely on event-based triggers to actually capture the situation of interest and to gain appropriate labels. Event-based triggers require an ESM app that is able to collect user information and to apply context recognition [87].

Usually, incoming notifications trigger an auditory, haptic, or visual cue [41, 116] to inform the user about the notification. The kind of cue is defined by the selected notification modality of the smartphone. Once perceived, the user can shift their attention towards the notification and either react to it or dismiss it [8]. In studies focusing on collecting data labels, the participant plays a central role. It is essential that they reply to as many prompts as possible which requires them to actually notice as many prompts as possible. Hence, it is crucial to consider the participant's perception of notifications, including their receptivity and interruptibility. According to Consolvo et al., the perception of notifications depends on multiple contextual factors such as the smartphone position, the user's location, or their activity [60, 88]. They also found a relation between the notification modality and the user's interruptibility [60]. In general, smartphone users have different preferences for their default notification modality [77, 88]. Similar to a user's interruptibility, this preference might be influenced by contextual factors such as the user's location [155], their activity [109], the fact of being in company or not [155], the user engagement with the smartphone [86], or the content of the notification [136] such as the notification importance [170].

Admittedly, the perceptibility of ESM prompts is not only limited to the user's recognition of incoming notifications in general, but also includes the user's review of notifications in the notification drawer. Increasing the visibility of ESM prompts within the drawer might lead to an increased response rate and a higher number of collected data labels. Usually, notifications have a unified presentation [170]. However, with the newest Android updates programmers get the chance to adapt the notification design and create a custom notification [3]. For example, this would allow to highlight important notifications by adapting their design while keeping the default design for notifications with low or neutral importance. Due to its restriction to the newest Android version, this customization is not yet applicable in large-scale assessment of smartphone-based data, but still bears potential for investigations.

1.1 CHALLENGES

The objective of smartphone-based assessment of annotated data is to collect a large set of representative, qualitatively high data, i.e., data that contains enough labels to fulfill the study objective such as building a robust predictive model or be able to run correlation analysis. To realize this objective, it is necessary to receive a high number of answered feedback questionnaires and to prompt for feedback in situations of interest. Moreover, the questionnaires must be answered in a reasonable amount of time after the prompt to ensure that the user-provided labels are assigned to the correct set of measurements.

Factors that influence the response behavior include but are not limited to:

- The moment in which feedback prompts arrive, e.g., if they fit to a situation of interest in which labels shall be assessed.
- The number and frequency of feedback prompts, e.g., the number of prompts received per day and in a short period of time.
- The length of the questionnaire, e.g., the time required to complete it.
- The perceptibility of self-report prompts based on the notification modality, e.g., whether the user actually perceived the auditory, haptic, or visual cue indicating incoming notifications prompting for feedback.
- The visibility of self-report prompts within the notification drawer of the smartphone, e.g., the discovery of notifications related to the ESM study within a set of general notifications stored in the notification drawer.

From these factors, we derived challenges which are addressed in this dissertation and described in the following.

Challenge 1 **Prompt in Situations of Interest**

In data annotation studies the objective is to gain labels to annotate a certain dataset. Hence, it is essential to assess the momentary label while a situation of interest takes place and is still ongoing or recent. One example for this is activity recognition. It is relevant to assess the label describing the current activity while this activity is still on-going to ensure that the dataset is labeled correctly. It is necessary to be able to automatically detect an event of interest and deliver event-based prompts to gain appropriate data labels. Several researchers argue that events of interest, especially if occurring less frequent, will not be sampled by random or time-based assessment [124, 176] but instead require event-based assessment [162]. This requires context-awareness of the ESM app and its ability to trigger prompts event-based [87]. Crucial for addressing this challenge is the access to a broad range of sensor sources which provide information to feed the event classifiers. *How to create such powerful and context-aware ESM apps?*

Challenge 2 **Reduce the Burden of Labeling Tasks to the User**
To gather a large amount of high-quality data labels, it is inevitable to rely on the user and to keep them motivated to provide labels. Hence, the burden of providing these labels must be kept to a minimum. This can be achieved by restrictions in terms of prompting frequency and facilitated questionnaire design [59].

Two sub-challenges emerge: the demand to reduce the number of prompts (Challenge 2a) and to reduce the length of the self-report questionnaire (Challenge 2b).

Challenge 2a **Restrict the Number of Prompts**
A user's response rate is related to the number of prompts they receive [59]. A high number of prompts introduces a higher burden to the user [70] and might inflict negative consequences such as a lower response rate or study exit [54, 147, 157]. For event-triggers, it is not possible to predict the number of daily prompts in advance as they might vary per day depending on the participants' actions and the actual number of occurring events. Berkel et al. recommend to introduce an inter-notification time and a maximum number of prompts [50]. The *inter-notification time* describes a prompt-free time interval between two successive prompts to avoid several event-triggered prompts shortly after each other. An *inquiry limit* can be interpreted as a maximum number of prompts per day. Addressing this challenge requires tools for the creation of ESM apps that are flexible enough to allow the setting of these properties, as emphasized by Consolvo and Walker [60]. *How to provide such functionalities to ESM apps?*

Challenge 2b **Reduce the Length of the Self-Report Questionnaire**
Another factor that influences the burden to a user is the length of the questionnaire [59]. The length of the questionnaire should be chosen in relation to the number of prompts per day [176], i.e., less questionnaire items for user studies with a higher number of prompts per day. In general, the length should be reduced to a reasonable minimum that allows a fast response but still captures all items of interest. This can be achieved by reducing the number of questions due to automatic assessment of information by using context recognition [50]. For example, the current activity or location can be assessed automatically by the smartphone and, usually, do not require an assessment via self-report questionnaires. To address this challenge, it is necessary to identify relevant contextual data and to use an ESM app that is able to assess such a variety of measurements automatically. *How to support the reduction of questionnaire complexity by embedding context-awareness into ESM apps?*

Challenge 3 **Support the Perceptibility of Smartphone Notifications**
ESM responses are of highest quality if the user replies to a prompt immediately [176]. The challenge is to inform the user about an incoming notification in a way that they actually recognize the arrival of the notification, e.g., by hearing their ringtone or by feeling the vibration of the smartphone. A solution to address this challenge is to automatically select a suitable notification modality based on the user's context. To our best knowledge, an automatic selection is not yet embedded in any mobile OS nor was it investigated. Consolvo and Walker argue that the notification modality to inform about a prompt should consider the user's context which includes the position the smartphone is stored at, the location, and the location-based activity of the user [60]. Mehrotra et al. identified the content of a notification as another relevant factor [136] which includes the importance of a notification.

Two sub-challenges emerge which we will investigate separately. They focus on the assessment of relevant factors (Challenge 3a) and the investigation of their relation to the perception of smartphone notifications and the user's preference for a specific notification modality (Challenge 3b).

Challenge 3a **Find Methods to Assess Factors that Influence the Perceptibility of Smartphone Notifications**
To be able to investigate relations between context and content-related factors, it is essential to be able to infer those from internal smartphone sensors. Within this thesis, we investigate the assessment of location, social activity and notification importance – three factors influencing the perception of a smartphone notification [60, 88, 155, 170]. For context-related factors, it is necessary to assess them automatically and efficiently, but also in a way that does not interfere with the user's privacy. For the content-related feature of perceived importance it will be necessary to find a definition and related smartphone features that allow an inference. *How to assess these factors of interest?*

Challenge 3b **Examine the Relation Between Notification Perception and Notification Modalities and Identify User Preferences**
The perception of a notification heavily depends on the notification modality. On the one hand, the selection of a modality depends on the perceptibility of the modality with respect to the current position of the smartphone [60], e.g., a less obtrusive modality while the smartphone is in the hand of the user or a more obtrusive modality while the device is stored in the backpack. On the other hand, users tend to have a general preference for notification modalities [66], e.g., silent mode at the cinema or vibration at home (i.e., depending on the location and location-based activity) or ringtone for incoming calls from

family members (i.e., depending on the notification importance). Relations between these perceptibility of incoming notifications and these factors require further investigations to be able to provide recommendations for automatic modality selection. *What are suitable notification modalities depending on the smartphone position and the location of the user? Which preferences do users have?*

Challenge 4 **Increase the Visibility of Important Smartphone Notifications**
The perceptibility of ESM prompts is not only limited to the perception of notifications based on different alerting cues, but also includes the user's visual processing of notifications stored in the notification drawer. Berkel et al. notice that it is difficult to anticipate incoming notifications from other applications than the ESM app [50]. It is possible that they arrive around the same time and that ESM prompts are replaced by common notifications or visually drown among the flood of notifications. A possible solution is to highlight important notifications (including ESM prompts) visually within the notification drawer to support their recognition. A need for different notification designs was already detected by Android which offers ways to customize notifications [3] which can serve as a basis to create different notification designs. Addressing this challenge requires the creation and evaluation of different designs to highlight notifications. *How to customize the design of important notifications such as ESM prompts to increase their visibility?*

1.2 CONTRIBUTIONS

This dissertation makes a contribution to the field of HCI, especially to context-aware ESM studies in terms of collecting labeled datasets. It focuses on the perception of smartphone notifications which are used to deliver self-report prompts to gather data labels through subjective feedback. The contributions address challenges mentioned above and are explained shortly in the following.

Contribution 1 **Introduction of a Tool to Build Context-Aware ESM Apps**
An ESM app has to fulfill different requirements. First of all, it should provide event-triggers and allow to prompt in situations of interest (cf. **Challenge 1**). The set of available events should be manifold and support contexts such as the location or the user activity. Such a tool is also required to offer properties such as the inter-notification time to define a prompt-free time window between two successive prompts and an inquiry limit to manage the frequency of prompts and to restrict the total number of prompts per day (cf. **Challenge 2a**). In addition, the tool has to provide access to a rich set of sensor sources from which further contextual information can be inferred. Depending on the objective of the ESM study, this might allow to

reduce the length of the questionnaire (cf. **Challenge 2b**) as information can be assessed automatically instead of being asked from the user. We introduce ESMAC, a tool that supports the creation of context-aware ESM apps and that offers not only access to a broad set of sensors and events, but also provides several properties such as inter-notification time and inquiry limit which can be set by the study designer.

Contribution 2 **Confirmation of the Relevance of Event-Triggers for an Exemplary ESM Study Focusing on Location and Activity Changes**
ESM tools usually offer three ways to trigger prompts: random, time-based, and event-based. Event-based assessment appears suitable for the assessment of data labels and the assessment of information with relation to an event. However, the usefulness of event-triggers for such cases was not investigated explicitly before. We compare all three types in an ESM study that aims at investigating the events "location change" and "activity change". We show that event-prompts are a good means to assess information in situations of interest (cf. **Challenge 1**).

One aspect of this dissertation is to investigate the perception of smartphone notifications in different contexts and depending on the notification content. However, to be able to carry out these investigations, it is required to have mechanisms to assess the factors of interest (cf. **Challenge 3a**). In our case, these are the smartphone position, the location of the user, an indicator for their social activity, and the perceived notification importance.

Contribution 3 **Presentation of a Position Transition Correction Mechanism to Improve the Detection of Smartphone Positions**
Concerning the *smartphone position*, we identify common smartphone positions and build a recognition model. The model is based on measurements from the accelerometer of the smartphone. Based on the assumption that each position transition needs to be made by a user who holds the smartphone in their hand to move it from one spot to the next one, we implement a correction mechanism. We show that this mechanism can be used to increase the recognition accuracy.

Contribution 4 **Presentation of Two Privacy-Sensitive Methods to Assess a User's Current Location**
Location assessment is facilitated due to the equipment of smartphones with GPS. However, raw GPS coordinates are a very sensitive property of a user as it reveals the exact location what might worry smartphone users. We present alternative, privacy-sensitive approaches based on WiFi and Place Types. We examine their precision and provide a recommendation of when to use which assessment method.

Contribution 5 **Presentation of a Location-Based Method to Estimate if a Smartphone User is in Company**
We assume that locations provide a semantic meaning from which location-based activities can be inferred. We present a method to estimate if a smartphone user is *in company* (i.e., a social activity indicator) based on the type of place the user is currently at, the time, and basic physical activity.

Contribution 6 **Introduction of Four Kinds of Importance and Presentation of Features that Relate to the Perceived Importance of Notifications**
The notification content influences the perception as well. One relevant aspect is the *perceived importance* of a notification which is not yet clearly defined in literature. We introduce four kind of importance that allow to investigate the perceived importance more specifically. We also present smartphone features that showed correlations with the perceived importance as reported by participants of a user study.

Contribution 7 **Recommendations for the Selection of Suitable Notifications Modalities Based on (a) the Smartphone Position and (b) the Current Location and Possible Location-Based Activities**
The perceptibility of incoming notifications depends heavily on the notification modality, but might also be influenced by other contextual factors. We examine the perception of incoming notifications while different notification modalities are selected and (a) while the smartphone is stored at different *positions* (cf. **Challenge 3b**), including the trouser pocket, backpack, and on the table or (b) depending on the user *location* and *location-based activities* (cf. **Challenge 3b**). We present recommendations for suitable notification modalities which should be selected automatically based on the smartphone position or the user location.

Contribution 8 **Recommendations for Design Adaptions and Customization Options for Smartphone Notifications to Highlight Important Ones**
Perception is not only influenced by contextual factors, but also by the content of a notification. A user's willingness to tend to an incoming notification might vary based on the *perceived importance* of a notification. Important notifications could be announced with an auditory alert while the default notification modality is set to vibration. Another possible adaption is related to the actual presentation of the notification. Especially for users who receive a large amount of notifications, important ones might be overseen which causes frustration once noticed. We examine different designs that present notifications in a new way. We identify and recommend pleasant and useful design properties that facilitate the identification of important notifications compared to common ones (cf. **Challenge 4**).

1.3 DISSERTATION STRUCTURE

The composition and content of this dissertation is visualized in Figure 1.

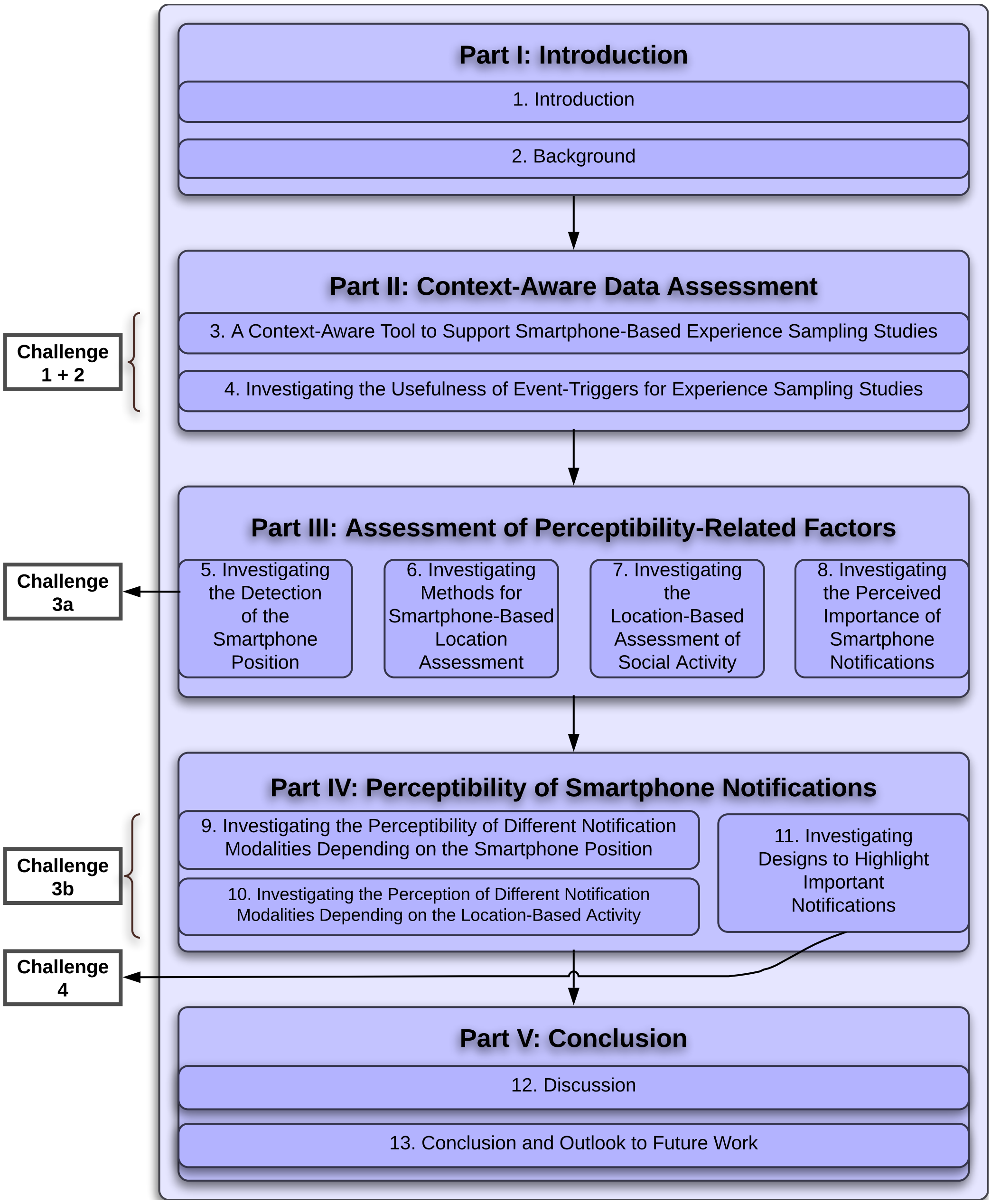


Figure 1.: Overview of the composition of this dissertation.

The structure of the dissertation can be summarized as follows.

Part I *"Introduction"* covers the introduction to the research topic and the contents of this dissertation. In addition, Chapter 2 "Background" introduces relevant concepts and provides information to facilitate the understanding of the remainder of this dissertation.

Part II *"Context-Aware Data Assessment"* focuses on context-aware experience sampling studies. Chapter 3 introduces a tool to build context-aware ESM apps. First of all, we present a survey to assess relevant sensor sources and events. Based on related work and the survey results, we build ESMAC, the ESM app configurator. This tool is explained in detail, including a system architecture and evaluation of its web interface and Android app. Lastly, we present investigations about the usefulness of event-triggers for the assessment of information related to location and activity changes in Chapter 4.

Part III *"Assessment of Perceptibility-Related Factors"* is a cornerstone of the investigation of notification perception: it examines the assessment of factors that influence the perception. As identified by related work, relevant factors include the location of the user, the activity of the user, and the perceived importance of a notification, among others. We investigate these three factors. First of all, we investigate the detection of the smartphone position and a possible correction mechanism in Chapter 5. Next, in Chapter 6, we consider different ways to assess location in a privacy-sensitive manner without storing actual GPS coordinates. Based on the location, we investigate how to assess the sociality of location-based activities, i.e., whether a user tends to be in company or not at a certain place type, in Chapter 7. Last, in Chapter 8, we examine different contextual and content-related smartphone features and their relation to a user's perceived importance of a smartphone notification.

Part IV *"Perceptibility of Smartphone Notifications"* focuses on the perception of smartphone notifications. First, this includes investigations about the influence of different notification modalities and the smartphone position on the perception and reception of smartphone notifications, covered by Chapter 9. Next, we consider the location of the user and location-based activities – influenced by the probability of being in company. We investigate preferences for notification modalities and influence of the location, the notification modality and the task engagement on the perception of smartphone notifications and present our findings in Chapter 10. Lastly, in Chapter 11, we move from contextual factors to content-related factors and evaluate different designs to highlight important smartphone notifications with respect to their perceptibility.

Part V *"Conclusion"* contains Chapter 12 "Discussion" in which we discuss our findings in terms of limitations and generalizability and ESM-related issues. Finally, Chapter 13 concludes this dissertation and provides a short outlook to future work.

2 Background

In this chapter, we define and explain basic concepts and principles to facilitate further understanding of this dissertation's contents. Foremost, this includes an introduction of the experience sampling method and main characteristics of this method. It is followed by a description of smartphone notifications, possible presentation styles, and notification modalities used to inform about incoming notifications. Next, we consider concepts that relate to the perception of smartphone notifications: the interruptibility, receptivity, and task engagement of smartphone users. We also present a differentiation between distraction, disruption, and disturbance as well as a review on negative effects of smartphone notifications on the user and possible solutions. The chapter is concluded by a short summary.

2.1 Experience Sampling Method

The main method to assess subjective feedback in HCI is ESM. ESM was described by Csikszentmihalyi and Larson [62] as a means for systematic observation of persons and their experiences in daily life [82, 185]. This includes the assessment of personal experiences, sentiment, thoughts, feelings, or behavior change [50, 71, 186]. The term *experience sampling* is often used interchangeably with *Ecological Momentary Analysis* (EMA) and *Ambulatory Assessment* [176, 186, 202]. They have different origins [185], but all share the central element of asking participants to answer self-report questionnaires to provide information about themselves and their environment, ideally using electronic devices such as smartphones [186]. Self-reports provide real-time insights about daily lives of the participants [186] without facing the retrospective nature of traditional approaches [185] that might introduce memory errors or recall bias [87, 176].

Traditionally, ESM studies were conducted using paper-and-pencil diaries and a beeper to prompt the users for self-reports that are provided analogue [185]. Due to the proliferation of smartphones and their sensing capabilities, they are considered as a perfect tool for conducting ESM studies [140]. By now, it is common to use an ESM app installed on a smartphone that prompts the user for self-reports by sending a notification that links to an ESM questionnaire within the app.

SCHEDULING OF PROMPTS There are different ways to trigger self-report prompts: randomly, time-based, and event-based [180], also known as signal contingent, interval contingent, and event contingent, respectively. Fisher and To [87] define them as follows:

- **Random**: prompts are sent out at random points in time over the day but often with stratified schedules; e.g., ten randomly scheduled prompts per day but not more than one per hour
- **Time-based**: prompts are triggered at fix points in time that follow an interval schedule; e.g., a prompt at every full hour between 8 a.m. and 10 p.m.
- **Event-based**: prompts every time a certain event takes place; e.g., a prompt at every location or activity change

As a fourth option, many studies encourage the participants to provide self-reports on a voluntary basis, referred to as "free input" [50]. It is possible and, depending on the study objective, might also be recommendable to combine different trigger types [87].

Each of the mentioned trigger types has its benefits and drawbacks. *Random* triggers give a representative sample of impressions and experiences throughout the day of a participant [87]. However, they might burden the participants due to their unpredictable nature [87]. *Time-based* triggers rely on time and appear on a regular basis, providing the participants with predictable timings [87]. Though, they likely skew the data towards events that happen more often [124]. *Event-based* triggers are suitable if the items to be observed relate to events that are selected as triggers [78]. Unfortunately, they might narrow down the overall daily observations by focusing on certain events only [124] and they require a well-functioning event recognition system [87].

Trigger types should always be selected in accordance with the study objective. For a general overview of a person's daily life, it makes sense to have random or time-based triggers. For the assessment of data labels that relate to a specific event, event-based triggers appear most suitable.

PROMPTING PROPERTIES A topic that is widely discussed among ESM experts is the number of prompts per day. To keep the user compliance high, it seems natural to reduce the number of prompts per day to a minimum. The *inquiry limit* (i.e., the maximum number of prompts per day) [50] should be set when creating or initially configuring an ESM app. For random and time-triggered prompts, either the number per day needs to be set or the exact points in time. For event-triggers, it is not possible to predict the number of daily prompts in advance. This number might vary per day depending on the participant's actions and the actual number of occurring events.

It is advisable to introduce an *inter-notification time* (i.e., a prompt-free time window between two prompts) [50] to avoid multiple event-triggered prompts shortly after each other. Otherwise, it might happen that an inquiry limit of prompts per day is reached within minutes, e.g., if the defined event is the reception of a WhatsApp message and the user receives several messages in a row via a group chat. Both properties, inquiry limit and inter-notification time, should be adjustable properties offered by a tool to create or configure ESM apps.

In general, researchers agree that the number of prompts should be aligned with the study duration and the complexity of the questionnaire [50, 181]. To further reduce the burden to the user, it is advisable to keep the complexity of the ESM questionnaire as low as possible by reducing the number of questions [50] or to minimize the number of open-ended questions [60]. The number of questions can be reduced by applying context recognition for automatic assessment of information [50]. Hence, access to a variety of sensors and recognition mechanisms is a desired functionality of an ESM app configurator.

Considering the perception of ESM prompts, such an ESM app configurator should further offer different modalities to inform about ESM prompts, i.e., about incoming smartphone notifications.

2.2 SMARTPHONE NOTIFICATIONS

Notifications are the main interaction features between a smartphone and its user to inform about newly available information [170]. They inform users about a variety of events [170] such as reminders for calendar events, about incoming or missed calls or text messages, about updates in social networks, new messages in instant messengers, or system updates. ESM prompts are nowadays delivered through smartphone notifications which ask the user to respond to a self-report questionnaire. Usually, these are push notifications that are send to the user and visualized via the GUI once triggered by the source app. In the context of this dissertation, we focus on the mobile operating system Android as it is the most widespread operating system [25] and gives programmers the highest freedom when developing apps and collecting data. This mobile operating system is a suitable platform for ESM apps. Android specific information is mainly based on the Android documentation [4].

NOTIFICATION PRESENTATION AND APPEARANCE Notifications can be presented in different forms and at different positions on the smartphone screen:

- Status bar notification: the notification appears in the status bar on top of the screen by showing the app icon (see Figure 2a)
- Notification drawer (standard): notifications are listed in the notification drawer that is available by swiping down the status bar (see Figure 2b)
- Heads-up notification*: only when the screen is on and the phone is unlocked, a notification appears on top of the screen in the foreground (see Figure 2c)
- Lock-screen notification*: notifications may appear when the screen is locked and be displayed to the user within the lock screen (see Figure 2d)
- App icon badge**: a small circle placed in the upper right corner of the app icon indicates new notifications of this app (see Figure 2e)

  *available on devices running Android 5.0 and higher

  **available on devices running Android 8.0 and higher

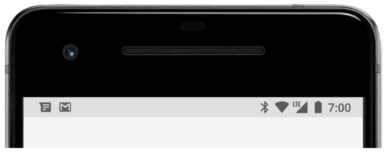
Commonly, smartphone users use the notification drawer to review notifications. There are two, or occasionally three, ways to interact with a standard notification in the notification drawer:

- React (acceptance): tapping a notification opens the app behind the notification
- Remove (decline): swiping sideways removes the notification, so that it does not appear in the notification drawer anymore
- Quick response*: swiping downwards reveals quick respond options if available; e.g., "call back" or "message" for a missed call (see Figure 3)

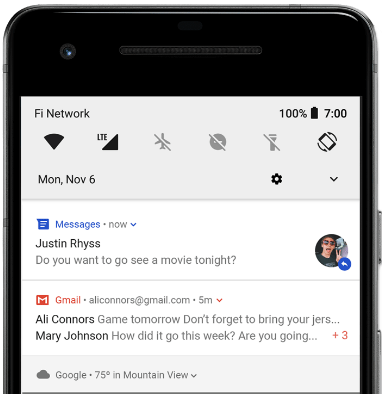
  *available on devices running Android 7.0 and higher

To answer an ESM prompt, it is required to tap the notification to be forwarded to the ESM app containing a questionnaire. For user studies with a small number of labels, it might be possible to provide the labels using quick response options.

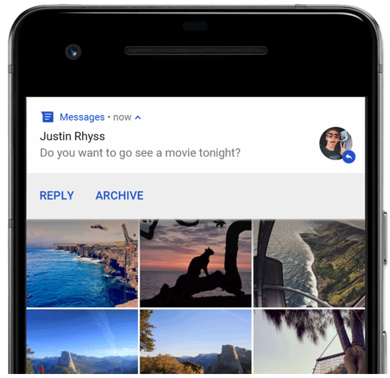
So far, notifications have a unified presentation [170]. Currently, there is no possibility to highlight ESM prompts. However, this option might be available for future Android versions which allow custom notification designs [3]. So far, custom designs are limited to a small percentage of devices only, but it seems promising that they will be available to a larger range in a few years. Such designs are worth being investigated regarding the perception of ESM prompts or other important notifications.



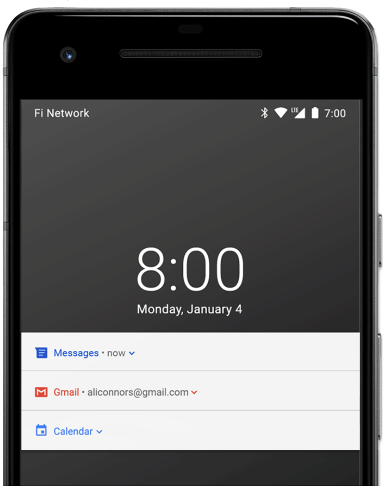
(a) Status bar notification



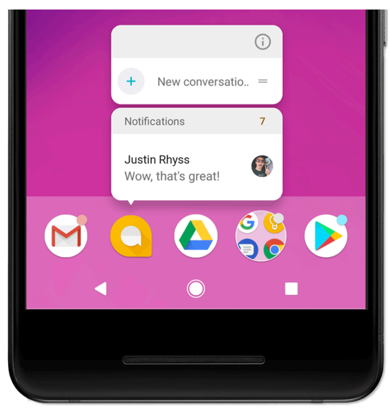
(b) Notification drawer (standard notification)



(c) Heads-up notification



(d) Lock-screen notification



(e) App icon badge

Figure 2.: Different presentations and spots for notifications to appear. Figures are taken from the Android notification documentation [5].

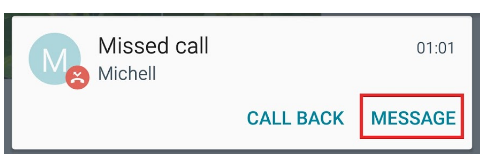


Figure 3.: Example of quick response options for a missed call notification [18].

NOTIFICATION MODALITIES The perception of smartphone notifications is influenced by the modality used to inform about incoming notifications. Common alerting modalities are:

- Auditory cues (e.g., ringtone): the phone sends out an audible cue whose volume and sound can be pre-defined by the user. Auditory cues are perceived as very obtrusive, but also very perceptible [77, 116].
- Haptic cues (e.g., vibration): the phone sends out haptic cues and vibrates in a pre-defined pattern and duration. Haptic cues are perceived less obtrusive while still being perceptible [77, 116].
- Visual cues (e.g., illuminated display, notification LED, or flashlight): the primary visual cue is the illumination of the display, showing notifications on the lock screen. If supported by the smartphone and enabled by the user, a notification LED on the front side of the smartphone sends out visual cues by blinking. This is the least obtrusive modality, but it is also more likely to be missed [77, 116]. Some applications such as *Flash Notification* [12] use the flashlight of the camera to indicate new notifications.

These alerts might occur alone or in combination. It is also possible to set the smartphone to silent mode so that the phone does not send out any cue at all and does not inform the user about incoming notifications in any way. In general, this is not an option in ESM studies as it prevents users from perceiving incoming ESM prompts. However, if compatible with the study objective, it might be a good option to provide some prompt-free time to the user to keep the compliance high.

More modalities exist that might be used to transfer information to a human being, e.g., smell, taste, or heat. However, they are not considered in this dissertation since they are not (yet) available on smartphones.

2.3 CONCEPTS RELATED TO THE PERCEPTION OF SMARTPHONE NOTIFICATIONS

Smartphone users have different preferences for their default notification modality [60, 77]. A suitable selection of a notification modality is crucial for the perception of an incoming notification by the user. The perception of a notification relates to the reception of the user that a new notification has arrived. Usually, this includes the perception of an alert caused by the selected notification modality which informs about an incoming notification. According to Consolvo et al., the perception of notifications depends on multiple contextual factors such as the smartphone position, the user's location, or their activity [60]. After recognizing the arrival of a notification, a smartphone user has to decide if they tend to the smartphone to handle the notification or if they decline an attention shift towards their device. This decision is influenced by properties such as the user's interruptibility and receptivity, but also their current task engagement.

INTERRUPTIBILITY OF SMARTPHONE USERS Interruptions cause users to shift their attention from the primary task to a secondary task [46] – e.g., towards an incoming notification requesting the user to check the currently provided information. Interruptions in unfortunate moments may cause frustration and annoyance of smartphone users [123, 139]. Many researchers already investigated interruptibility, especially in terms of finding opportune moments to deliver smartphone notifications [183, 155]. They found that interruptibility depends on different features such as the user activity [109], being in company [155], location [75], and engagement with the smartphone [86], but also on the *content* of notification [136] such as its importance [95, 170]. Mehrotra et al. found that notifications are least disruptive at the beginning of a new task and most disruptive in the middle or short before finishing a task [139]. They also found that the complexity of the primary task correlates with the disruption caused by the interruption [139]. In summary, interruptibility depends on the context of the user and the current task.

RECEPTIVITY OF SMARTPHONE USERS A concept closely related to interruptibility is receptivity. Begole et al. defines receptivity as "one's willingness to be interrupted" [48]. Fischer et al. also relate receptivity with interruption and define it as a user's overall reaction to an interruption, including the user's interruptibility and the experience towards the interruption [85]. They mention that a smartphone user might be receptive of a smartphone notification even though it is interruptive – if the content matters to the user [85]. The same applies vice versa: a user might be interruptible due to an idle activity, but still annoyed by the content of a notification [139].

Mehrotra et al. define receptivity depending on [139]:

1. How interesting, entertaining, relevant and actionable the content of a notification is to the user [85]
2. The type of app that triggers the notification [170]
3. The time criticality and social pressure [159]

The terms receptivity and interruptibility are interacting concepts that depend on similar features. However, there are slight differences that should not be neglected. As pointed out by Fischer et al., interruptibility is rather interesting from a *sender side* as it deals with opportune moments for notification delivery (rather system-oriented) while receptivity anticipates the sentiment and reaction of the *recipient* (rather user-centered) [85]. Though, they agree that both concepts have the same objective of reducing the burden of interruptions – which is also an objective of our research.

TASK ENGAGEMENT Urh and Pejovic argue that "task engagement directly impacts a user's sentiment and reaction towards an incoming notification", i.e., the receptivity [188]. Several researchers found out that suitable moments to interrupt are those at breakpoints [144, 148]. These are points in time in which one task ends and another starts, i.e., the task engagement decreases before it increases again. Okoshi et al. found that the user's willingness to engage with the smartphone and attend notifications is higher if notifications are provided in an interruptibility-aware manner [151].

DISTRACTION, DISRUPTION, AND DISTURBANCE Three concepts that appear related to interruptibility, receptivity, and task engagement are distraction, disruption, and disturbance. Distraction is defined as "the process of being distracted" [30], i.e., the act of "having one's attention diverted" [29]. Disruption is defined as "an interruption to the regular flow or sequence of something" [28]. Disturbance is defined as "the act of disturbing, being disturbed" or "an interruption of that which is normal or regular" [31]. All definitions refer to interruptions and a shift of attention from the current task towards the interruption. While distraction and disruption rather neutrally state that the attention is drawn and the flow or sequence is intermittent, disturbance signifies a negative effect: it refers to the cause of "distress or worry; upsetting or unsettling" [32]. Not all notifications are disruptive or disturbing [135]. Among others, the level of disruption or disturbance depends on the relevance of the notification's content [136]. It seems necessary to avoid disrupting and disturbing notifications but to ensure the delivery of relevant notifications [136].

NEGATIVE EFFECTS OF SMARTPHONE NOTIFICATIONS Smartphones introduce permanent availability. With growing number of installed apps the number of notifications also increases [197] – it "virtually explodes" [149]. If notifications are not handled in accordance to the user's interruptibility and receptivity they might induce negative effects. This includes loss in productivity [45, 118], a decrease in task completion time [64, 63, 142], changes in emotional, social and psycho-physiological behavior [34, 45, 204], but might also lead to technostress or digital burnout [126]. Users need to be supported in finding what they are looking for or by hiding what is not of interest or currently not of relevance, i.e., delivering "the right type of information at the right time" [85, 109]. Solutions might include systems that filter unimportant notifications [161], deliver notifications interruptibility-aware at opportune moments [155], or learn about the perceived importance of notifications and adapt to it [170]. In summary, there is a need for a notification management system that handles notifications with respect to the user's receptivity and interruptibility and the perceived importance of a notification. This thesis and its contributions are a first step towards such a system.

2.4 SUMMARY

In this section, we introduced ESM, the traditional method for collecting subjective feedback in the wild. Due to the omnipresence of smartphones and their sensing capabilities, they became a popular tool to conduct ESM studies. The main way to deliver ESM prompts are smartphone notifications. They can be presented in different ways and be announced by different alerts. In addition to the selected presentation style and alert, the interruptibility and receptivity of the user as well as their task engagement influence whether and how incoming notifications are perceived. Perceived notifications are distracting, might disrupt and, thereby, bear potential to disturb users. Especially if they are badly-timed or occur with a high frequency, notifications possibly inflict negative effects, e.g., on health or productivity. However, they might contain information of personal importance or that is useful and relevant for the current task of the user. There is a growing need for a notification management system that delivers notifications intelligently. Such a system has to consider the current context and state of the user as well as it has to rate the relevance and importance of the notification's content. This requires knowledge about the user, their behavior and preferences – either embedded into a smartphone app, inferred from physical and virtual sensors during runtime, or provided as input by the user through app settings. Within the remainder of this dissertation, we examine methods to assess context and content-related features, investigate how perceptibility is influenced by such factors, and identify user preferences for notification modalities and presentation designs.

Part II.

Context-Aware Data Assessment

3 A Context-Aware Tool to Support Smartphone-Based Experience Sampling Studies

Collecting subjective feedback in everyday life during everyday activities is a fundamental task in different disciplines [106]. Usually, ESM is used to gather this information. We focus on user studies that apply ESM to collect data labels to annotate smartphone measurements since this is a crucial but challenging task in computer science, especially for supervised machine learning. Such an assessment requires an ESM app that is able to gather data from internal physical and virtual sensors and that prompt the user for feedback in form of self-reports in situations of interest. To avoid the need to create a new ESM app for each user study from scratch, it is advisable to rely on tools that support the app creation. Such tools have to fulfill different requirements including the assessment of information of interest (i.e., sensor measurements) in situations of interest (i.e., relevant events) [90]. To do so, it is necessary that the tool allows to access a wide range of sensors and that it offers multiple event-triggers. It is important that the study participants respond to as many self-report prompts as possible. To reduce the burden of answering questionnaires, it is advisable that ESM tools allow to set an inquiry limit and an inter-notification time [50] – properties that are not yet supported by all ESM tools. Common smartphones are powerful and offer the capabilities for a context-aware assessment of data within ESM studies. However, many existing ESM tools are limited in their functionalities and do not exploit the full potential of the sensing device.

We address the need for a tool which provides a broad access to sensors and event-triggers and which offers various notification settings. As a solution, we present ESMAC: a context-aware tool to support the creation of ESM apps. As a first step, we review related work to identify properties and requirements for such a tool. Next, relevant sensors and events are identified within a survey among ESM experts. Based on the gathered information, we design and prototype a platform to create context-aware ESM apps – *ESMAC*. The prototype of this platform is evaluated through a comparison with a state of the art system. Each of these steps is explained in more detail in the following sections.

The identification of relevant sensor sources presented in this chapter was published and presented as a poster at UbiComp'15 [43]. The *ESMAC* system and the corresponding results were published and presented in an oral presentation at MobiHealth'15 [44].

3.1 RELATED WORK

The creation of an app to conduct an ESM study using smartphones might require specific programming knowledge [44, 50]. Several platforms are available that support study designers when creating ESM apps. They vary in format, required programming knowledge, available sensors for logging in the background, and trigger types – especially in terms of available event-triggers. A fairly broad overview can be found online on the website of the Society for Ambulatory Assessment [24].

In this chapter, we focus on context-aware ESM software to create mobile apps. Platforms such as *ESm Capture* [10], *LifeData* [16], or *Ilumivu's mEMA* [14] that do not offer event-triggers are thereby excluded.

The first notable platform for app creation was *MyExperience* [90], "a context-aware data collection platform for capturing objective and subjective data as it's experienced" [21]. Study designers can design ESM studies by choosing from a set of question types and by selecting sensors to be accessed, e.g., GPS, GSM or keystroke dynamics. In addition, experts can define event-triggers based on additional, external sensors such as a heart rate sensor. The studies are then executable on Windows Mobile devices – which are, by now, rather unpopular in contrast to iOS or Android [25]. In addition, *MyExperience* requires knowledge about a specific XML schema to configure the sensor logging and the event-triggers.

An Android equivalent to MyExperience is *movisensXS*, "the next generation research tool for ambulatory assessment" [19]. It provides a similar functionality as *MyExperience*, but offers additional wearable sensors for rent with built-in data assessment within the corresponding mobile app: Move 3, LightMove 3 and EcgMove 3 [20]. The free basic version of *movisensXS* has very restricted functionality while the beta version offers a wider range of options. The app allows to log measurements from a variety of internal sensors such as accelerometer, ambient light sensor, battery status, nearby Bluetooth devices, connectivity status, or location. The offered external sensors allow to log further sensor information, for example, the heart rate using the EcgMove 3 sensor. *movisensXS* supports random and time-triggered prompts as well as event-triggers. The basic version of *movisensXS* only considers level of activity and location as event-triggers. Within the last years this platform enhanced their line-up of event-triggers, now including app usage, SMS reception, iBeacon in range, sensor triggers, and intent triggers for 3rd party apps. However, most functionalities are restricted to paying customers only. Providing a platform with great functionality free of cost is an important aspect in research why we aim at constructing a free-to-use platform.

A platform suitable for creating both Android and iOS apps is *Metric Wire* [17]. It allows to capture accelerometer data, location data, and communication information. The app offers random and time-triggered prompts. The only supported event-triggers are location-based. The functionalities offered by this platform are rather limited.

A similar tool to *Metric Wire*, again with a focus on Android and iOS apps, is *Paco* [23]. *Paco* offers to log the user location and app information and allows to trigger based such measurements in addition to traditional random and time-based prompts. The possibilities for context-aware experience sampling with *Paco* are rather restricted again.

A system that support the programming of apps for multiple mobile operating systems is *ohmage* [22]. It provides random, time and location-triggered prompts. In addition, it offers access to accelerometer data, WiFi, mobile radio cell information and GPS. To enhance compliance, *ohmage* allows study participants to adjust event-triggers. However, adjustment of event-triggers might interfere with the study objective and should only be managed by study designers. In addition, the platform's support for context-awareness is rather low as it only supports time and location-triggers.

A young platform is *Jeeves*, a "visual programming environment for mobile experience sampling" [168]. This platform allows researchers to create Android apps for ESM studies in a visual manner, supported by a drag and drop functionality. It allows to log location information, accelerometer data and communication information. Prompts can be triggered randomly, on a timely basis, or based on sensor measurements. However, the creators of *Jeeves* did not reveal details about which sensors are accessible that can be used to trigger event-based prompts. Unfortunately, the developers did not publish their code, inhibiting to actually run this platform or to investigate the supported event-triggers.

A fairly young tool is *AWARE*, "an Android instrumentation framework for logging, sharing and reusing mobile context" [7]. This tool offers a wide range of sensors, e.g., location, accelerometer, call, messages, screen activity, or app usage, and provides all three prompting types. Unfortunately, the AWARE website does not offer detailed information about the kinds of event that can are supported. Hence, no final statement about the functionalities and capabilities of this tool can be provided.

In summary, there are many different tools available that have their own benefits and drawbacks (see Table 1). If a platform is free-to-use, it usually only offers a limited functionality in terms of sensor logging and event-triggers. Many systems already grant sensor access for logging and to enable event-triggers, but they do not use the full potential provided by current mobile operating systems. If event-triggers are supported then they usually only cover events related to the current location. However, this does not cover many situations of interest.

Table 1.: Overview of considered ESM tools.

Tool	**Trigger Types**	**Supported Sensors**	**Supported Events**	**Mobile OS**
AWARE [7]	random, time, event	accelerometer, app usage, communication, location, phone usage	no details revealed	Android, iOS
Jeeves [168]	random, time, event	accelerometer, communication, location	no details revealed	Android
MetricWire [17]	random, time, event	accelerometer, communication, location	location	Android, iOS
movisensXS [19]	random, time, event	accelerometer, ambient light sensor, battery status, connectivity type, location, nearby Bluetooth devices + more	activity, location	Android
MyExperience [21]	random, time, event	communication, GSM, location, phone usage + more	communication, location, phone usage + more	Windows Phone
ohmage [22]	random, time, event	accelerometer, GMS, location, WiFi	location	Android, iOS
Paco [23]	random, time, event	app usage, location	app usage, location	Android, iOS

There is an emerging need for a free, easy-to-use, but powerful platform to create and configure Android apps for conducting ESM studies. As a first step, we investigated suitable sensor sources and contexts that are of interest for ESM study designers. These findings served as a basis for a platform that enables building such context-aware ESM apps.

3.2 IDENTIFICATION OF RELEVANT SENSOR SOURCES

Several internal smartphone sensors were already considered in related work, e.g., accelerometer [68] or GPS [74]. However, a wide range of available sensors [6] was neglected so far, as visible in an overview of ESM studies and their properties by Berkel et al. in their survey paper [50]. In addition, researchers or tool designers tend to not reveal details about which events can be inferred from sensors measurements and how relevant they are for ESM practitioners. For clarification and as a basis for our ESM platform, we created an online survey on Google Forms and distributed it among members of the *Society of Ambulatory Assessment* [1]. Overall, 29 of these ESM experts answered the survey.

The survey consisted of 20 questions about the relevance of sensor sources and the desired format or related event format in which this data shall be gathered. We applied the MoSCoW priorization [52], i.e., asked all participants to state their desire using "must", "should", "could" or "won't" statements. In addition, we added a "don't" option to allow participants to express a desired exclusion of a sensor (similar to a "must not" [51]). The selectable items were chosen based on their usage in related work or their availability through Android APIs.

We ranked the responses per question from 0 ("don't") to 4 ("must") and averaged the values. Table 2 visualizes the results and highlights them by priority.

The results indicate that the most relevant sensor sources are *time*, *date*, *user activity* and *location*. This is not surprising as these are the classical items that are usually assessed in ESM studies to grasp an idea of the participants' daily activities and their biorhythm. The items *notifications* and *accelerometer* are also of interest. This is just reasonable, since the accelerometer reveals a certain level of activity. Notifications are considered relevant as they are the means for communication between the smartphone and its user. Furthermore, they often refer to social interaction, e.g., updates from social networks or instant messages [160, 170], similar to calls and SMS which were also identified as relevant.

Table 2.: Prioritization of sensors with an indication of a relevancy category of either "must" (≥ 3.0), "should" (≥ 2.0), or "could" (≥ 1.0).

Sensor	Priority
Time	3.79 (± 0.77)
Date	3.76 (± 0.79)
User Activity	3.03 (± 0.94)
Location	3.00 (± 1.00)
Notifications	2.48 (± 1.12)
Accelerometer	2.41 (± 1.05)
Calls	2.21 (± 0.98)
SMS	2.21 (± 1.05)
Weather	2.14 (± 1.03)
Bluetooth Devices	2.14 (± 1.19)
Ambient Light	1.97 (± 0.87)
Current App	1.97 (± 0.82)
Social Networks	1.97 (± 1.12)
Screen Activity	1.97 (± 0.94)
Ambient Noise	1.93 (± 0.96)
Touch Activity	1.90 (± 1.21)
Calendar	1.83 (± 0.85)
App Crashes	1.76 (± 1.18)
WiFi	1.69 (± 1.20)
Connectivity Type	1.55 (± 1.09)

We also asked for the desired format of information derived from specific sensors so that it can serve to trigger prompts event-based. The participants were free to pick multiple options so that the overall number of responses does not match the number of survey participants. Table 3 presents the responses and how often an option was selected.

Many participants are interested in having either specific knowledge about the participant's context (e.g., *specific time* or *certain day*) or an abstraction of it (e.g., *abstract location* or *movement yes/no*). It is evident that for many sensor sources certain formats are desired. If we consider each answer that was chosen by at least one third of all participants (i.e., received about 10 picks), we can see that there is at least one specific format of interest for almost every sensor. For all of these formats and events, study designers have to consider to which extent they want to trigger prompts. Events such as "receiving an SMS" or "receiving a notification from WhatsApp" might occur very often and require an inquiry limit or an inter-notification time to reduce the burden to the user.

Table 3.: Overview of desired formats of sensor measurements and related events with indication of number of times an option was picked.

Sensor Source	**Format / Event**
Time	Specific time (21); range (4); daytime (3)
Date	Certain day (21); repeating day (6); higher contexts (5); range (4)
Location	Abstract location (17); certain location (14); certain area (5)
Accelerometer	Movement yes/no (15); averaged movement (14); movement in certain axis (11)
Notifications	Any notification (10); notifications from a certain app (8); number of notifications (7)
Ambient light	Light level range (12); Specific light level (7)
Ambient noise	Speech recognition (10); noise level range (10); specific noise level (8)
Calendar	Calendar status (10); number of events (8); priority of events (3); calendar type (1)
Weather	Weather context (16); temperature (15)
Bluetooth	Device identification (10); number of devices nearby (8)
App Activity	Certain app (9); certain category of apps (7); number of active apps (2)
SMS and Telephone	Any SMS/call (12); SMS/call from a certain number (7); number of missed calls/SMS (7)
Social networks	Number of posts (8); activity (8); number of friends (6)

To conclude the sensor analysis, we investigated the accessibility of all sensors and their formats for smartphones running at least Android 4.4.4 (KitKat). We were able to implement most of the sensors shown in Table 2. Unfortunately, *current app*, *app crashes* and *touch activity* are only accessible with root access or via accessibility services and cannot be accessed on the user's own phone in the context of our platform. For now, these sources are only available on test devices that are handed out to the users by the study designer.

3.3 A TOOL TO BUILD CONTEXT-AWARE APPS FOR EXPERIENCE SAMPLING STUDIES

Based on the properties, sensors and events identified in the previous sections, we built a platform to create context-aware ESM apps. We introduce *ESMAC*: a context-aware ESM app configurator for Android apps.

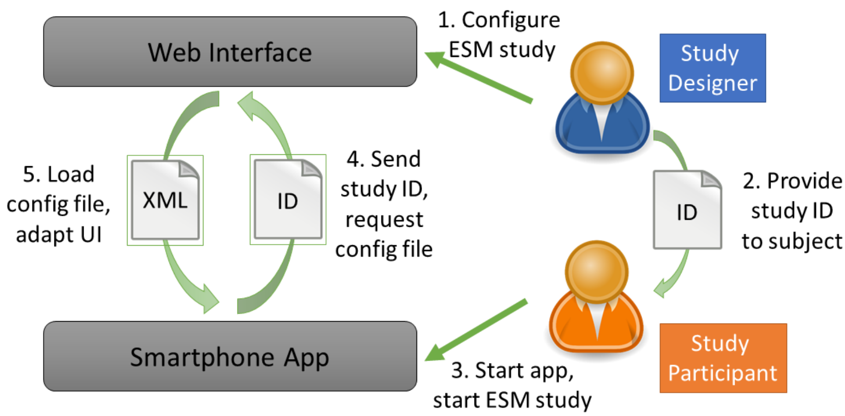


Figure 4.: Core components of the *ESMAC* system.

3.3.1 *System Architecture*

ESMAC [11] is an open-source, extensible *ESM App Configurator*. It combines the benefits of *movisensXS* (web interface for non-programmers to configure individual ESM apps) and *MyExperience* (XML-based configuration of event-triggers). *ESMAC* consists of two components: a *web interface* to configure ESM studies and an *Android app* to conduct the ESM study (see Figure 4). In addition, there is an XML file that serves as an exchange format to communicate the configuration from the web interface to the Android app.

The *web interface* allows the study designer to configure an ESM study (step 1) by selecting question types, sampling strategies, sensors for continuous assessment in the background, notification modalities, and additional prompting settings. This configuration is stored in an XML-based *data exchange format*, defined by a specific XML schema definition (XSD), and assigned to a unique ID. The study designer shares this ID with their study participants (step 2). On the first start of the app (step 3), the participant is asked to insert the ID (step 4) causing the app to download the corresponding configuration file. Once transferred to the smartphone, the configuration file is interpreted by the *Android app* which adapts its GUI, prompting and sampling strategies dynamically to this configuration (step 5).

There are four concepts related to the configuration of an ESM study with *ESMAC*: forms (which questions to ask), rules (when to prompt), sensors (what information to log in the background), and notification type (modality, further prompting settings). There is one view for the definition of forms, one view for the specification of prompting rules, and one joint view to specify sensors and notification type. Each view will be described in more detail in the following.

THE FORM VIEW The *Form View* (see Figure 5) offers a wide range of question types to choose from:

- Open-ended questions
- Multiple choice questions with single or multiple answers
- Slider questions
- Likert scales
- Conditional multiple choice questions; follow-up questions are only showed in case pre-defined option was selected

The study designer may select as many questions as desired. In addition, the study designer can create Android views ("questionnaire pages") and define which questions are shown together in one view.

THE SAMPLING VIEW The functionality provided by the *Sampling View* (see Figure 6) is based on Boolean algebra and allows to specify rules to trigger self-report prompts. Prompts can be set to be triggered randomly, time-based, or event-based. Each prompting rule is represented by a sensor expression. A sensor expression is a concatenation of a sensor type, a value type, an operator, and a value. Available sensors, as identified in the online survey, include: accelerometer, ambient light, Bluetooth, call log, display state, GPS, notifications, time, user activity, and weather.

Sensor expressions are combined using *and* or *or*. The resulting expression can either be true or false. A prompt will be triggered in case that the configured conditions yield true.

THE SENSOR AND NOTIFICATION VIEW In the *Sensor and Notification View* (see Figure 7), the study designer selects sensors from which available measurements shall be logged continuously in the background. These sensors are, again, based on the selection made in Section 3.2. The study designer selects one or more notification modalities that shall be used to inform the participant of the ESM study about incoming prompts. Available modalities are ringtone, vibration, and notification LED, if a notification LED is available on the device. The study designer might set further prompting properties such as an inquiry limit or an inter-notification time. Moreover, there is an option to allow or decline voluntary self-reports, depending on the study objective.

FINISH All configurations are transformed into an XML representation and available for transference to the smartphone of user.

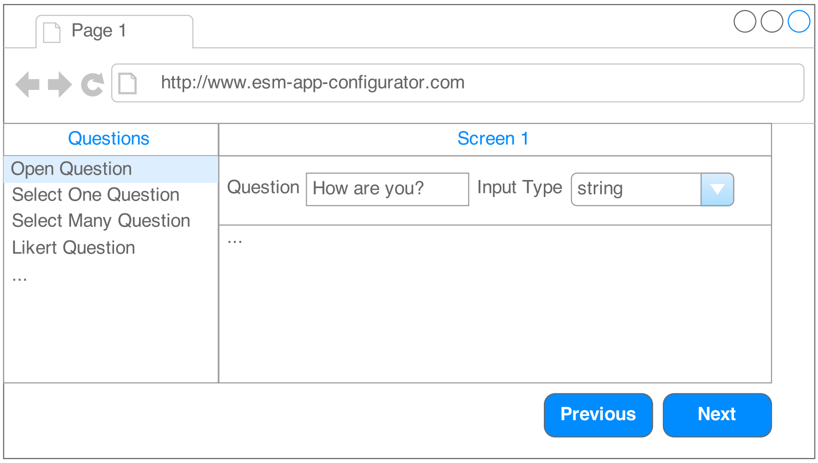


Figure 5.: Example of *ESMAC*'s form view.

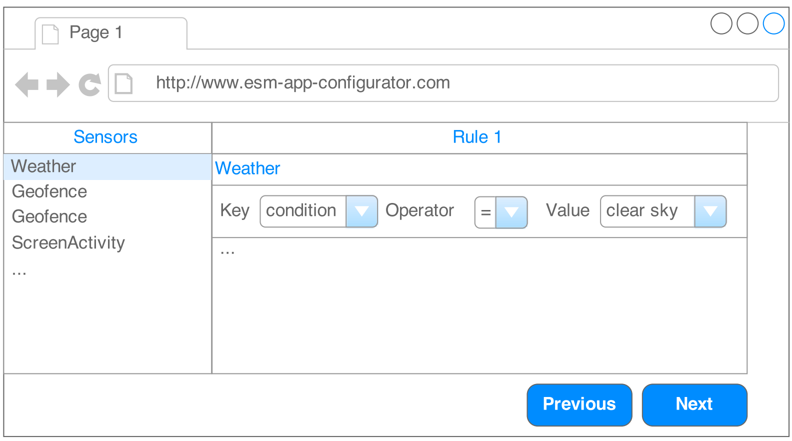


Figure 6.: Example of *ESMAC*'s sampling view.

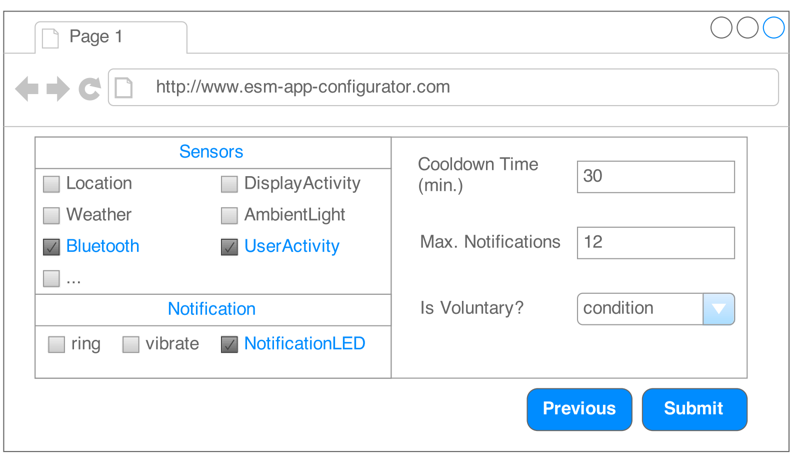
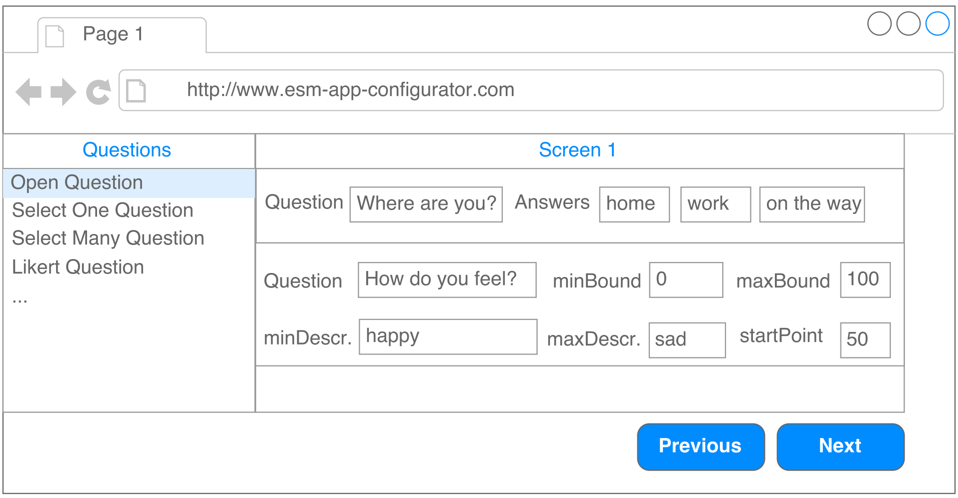


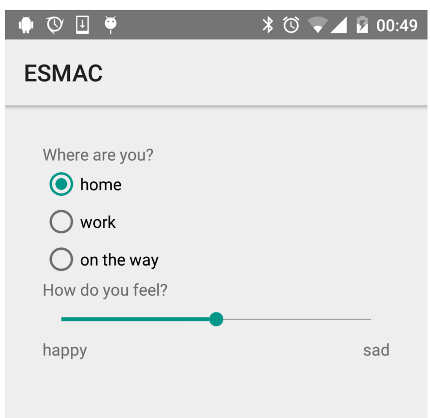
Figure 7.: Example of *ESMAC*'s sensor and notification view.

3.3.2 *Android App*

On start, the Android app validates and parses the XML representation to generate the specific GUI. The app adapts to the defined prompting rules and continuously gathers measurements from the selected sensors. The app runs on smartphones equipped with Android 4.4 and above. Figure 8 shows an example GUI of an ESM study: Figure 8a shows the question in the *Form View* while Figure 8b visualizes the representation of the same question within the Android app.



(a) Web Interface



(b) Android App

Figure 8.: Visualization of an exemplary configuration. The question is displayed within the webinterface during the configuration process and within the Android app during runtime.

To get access to different sensors we implemented an open-source, extensible *sensor library* which applies an observer principle to access sensors in an energy-efficient manner. After change of a relevant value the evaluation mechanism is triggered and checks each prompting rule. A prompt is triggered in case that an expression turns true, the inquiry limit per day is not yet reached, and the time difference to the last prompt exceeds the inter-notification time. A prompt is delivered by an Android notification.

To guarantee the ongoing evaluation of all rules an Android service was implemented which is forced to stay in memory or, if closed, forced to restart in time. The app is also programmed to notice if the system is rebooting and programmed to re-start automatically to ensure a permanent run of the app during the study.

In case of a valid rule, a notification is sent out with the specified parameters and based on the default ringtone, vibration pattern, and LED patterns. The user can interact with the ESM notification in the usual manner: a tap opens the ESM app and reveals the questionnaire.

3.4 EVALUATION

The evaluation of the system was divided into two parts according to the two main components of the system: the web interface and the Android app. In both cases, *ESMAC* was tested against *movisensXS* [19], a state of the art platform with similar architecture and functionalities.

3.4.1 *Evaluation of the Web Interface*

STUDY DESIGN The web interface was evaluated in a laboratory setting by two ESM experts from the field of applied psychology. They are members of the Society for Ambulatory Assessment and already participated in the sensor identification survey. Both of them are experienced with conducting ESM studies using smartphones and familiar with the premium version of the *movisensXS* platform.

The participants were asked to design a short ESM study that aims at assessing the emotion of students during lecture time for three days. The conditions were counterbalanced [36], i.e., one participant started using *movisensXS* and *ESMAC* afterwards, the second participant vice versa. To measure the usability, user experience, and mental workload of the participants we used three standardized questionnaires: System Usability Scale (SUS) [53], User Experience Questionnaire (UEQ) [125], and NASA Task Load Index (NASA-TLX) [105]. In addition, we handed out a free text questionnaire to receive qualitative feedback.

RESULTS Both systems achieved an average SUS score of over 90 points: *ESMAC* scored 95 and *movisensXS* 93.7. Thus, both qualify for an A+ grading according to Sauro and Lewis [171].

Concerning user experience, we reviewed the results in each of the six dimensions of the UEQ as summarized in Figure 9. Schrepp et. al provide a benchmark to classify UEQ average scores [174]. Based on this benchmark, both *movisensXS* and *ESMAC* show *excellent* results in the dimensions attractiveness, perspicuity, efficiency, dependability, and stimulation.

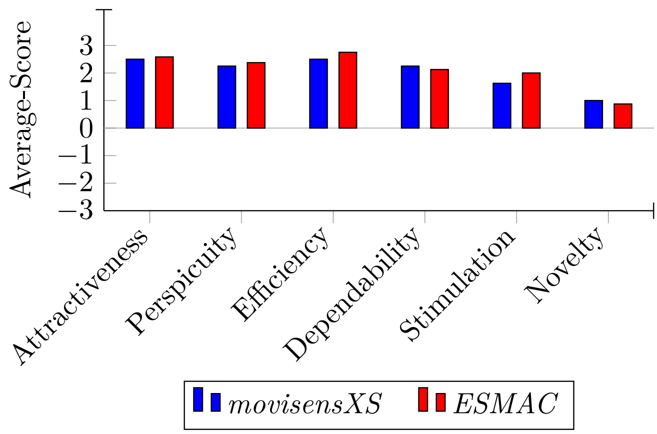


Figure 9.: UEQ results for both web interfaces.

The score for mental workload measured by the NASA-TLX was 24.58 for *ESMAC* and 25.83 for *movisensXS*. Neither of them tends to underload or overload [104].

Both apps show a great performance in usability, user experience, and mental workload. Due to the small sample size, the results need to be treated carefully. Though, they can be seen as an indicator for *ESMAC*'s performance compared to *movisensXS*. The results suggests that *ESMAC* is able to keep up with a state of the art platform – while, in addition, offering a wide range of event-triggers for prompting.

In the free text questionnaire one expert noted that *ESMAC* increases the possibilities of performing ESM studies due to its combinations of sensor sources and event-triggers. This emphasizes the relevance of event-triggers and usefulness of *ESMAC*. The other expert missed a few *movisensXS* features he was used to. Though, these are only available for customers in a pay-to-use version and were thereby not available in the context of our evaluation.

3.4.2 *Evaluation of the Android App*

STUDY DESIGN To evaluate the Android app, we chose one configuration that was created in the evaluation of the web interface. The ESM configuration was in German, but can be translated as follows:

- **Questions: 6-Point Likert Scales**
  - How do you feel? energetic - tired
  - How do you feel? tense - relaxed
  - How do you feel? good mood - bad mood
- **Rules**
  - 17 randomized time triggers from 08:00 to 23:59
  - Number of Bluetooth devices >= 2 (*)
  - Call status = answered (*)
  - User Activity = walking, running, on bicycle, in vehicle (*)
  - Notifications of WhatsApp, Facebook (*)

  (*only for *ESMAC* due to its event-triggers)

We had two study conditions: usage of *ESMAC* and usage of *movisensXS*. Both received configurations as similar as possible, even though many event-triggers are only available for *ESMAC* as indicated by a star in the listing.

The evaluation of the Android app was carried out as a field study with a duration of six days. We recruited 10 participants (7 male, 3 female) with an average age of 24 years. The study was conducted within-subject with counterbalanced conditions. Group 1 (participants 1-5) experienced the *ESMAC* app for the first three days and the *movisensXS* app for the last three days, group 2 vice versa. After finishing each study condition (i.e., after using one app for three days), we asked all participants to answer a SUS questionnaire, an UEQ questionnaire and a free text questionnaire. At the end of the study we handed out another free text questionnaire to gain additional feedback on both apps, especially in comparison to each other.

RESULTS Both SUS and UEQ scores differ between the two apps.

ESMAC reaches an average SUS score of 83.5 whereas *movisensXS* only reaches an average score of 74.25.

The UEQ scores are visualized in Figure 10. *ESMAC* received high scores on all scales with the highest difference in the attractiveness scale.

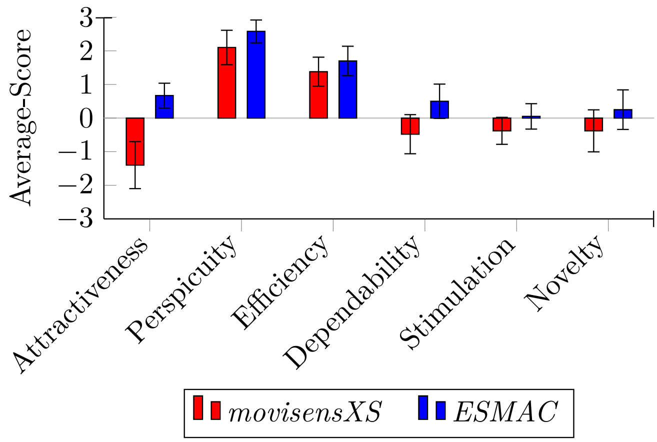


Figure 10.: UEQ results for both web interfaces.

We investigated the average *Response Time* (i.e., the time between prompt and answer), *Response Rate* (i.e., how many prompts did the participants react to) and *Task Completion Time* (i.e., the time between reacting to a notification and submitting the self-report). The response time of *movisensXS* is very low (10.21s) in comparison to *ESMAC* (783.87s). Both lie in the required time limit of 20 to 30 minutes [176]. The high response time of *ESMAC* is caused by outliers that originate from the event prompts that might even occur at night. The study configuration did not include any prompt-free night time. For *movisensXS* prompts, this barely made a difference as the defined events only took place during daytime. This is different for *ESMAC*: a participant might receive a Whats-App message at night, but notices the prompt only in the morning. If we focus on daytime-only prompts, *ESMAC*'s response rate is reduced to 420.58s which is still much higher than the one for *movisensXS* prompts. We assume that this difference is caused by ESMAC's less obtrusive prompting modality which leads to less perceptible notifications.

The average response rate of prompts for *movisensXS* is 76.72%, *ESMAC* only reaches 46.12%. Low values are again often caused by nightly prompts which were ignored by a sleeping participant and deleted from the notification view once a new prompt came in. Considering day-time prompts only, *ESMAC*'s response rate is 64.07%.

For *ESMAC* we were able to measure the task completion time. The mean is 42.68s with a median of 12.18s – a fair answering time for 3 questions. Some outliers, e.g., one response time of 1.15h, distort the results. High task completion times are usually caused by participants who open the app, get distracted, and finish the self-report questionnaire later. Future versions of the *ESMAC* system might include an expiration time for prompt notifications as well as for finishing the self-report questionnaire.

We also investigated the share of trigger types to investigate the power of event-triggers (see Table 4). 84.47% of all prompts were event-triggered which shows the effectiveness and importance of this trigger type.

Table 4.: Share of all prompt triggers.

	Notifications	**Bluetooth**	**User Activity**	**Call**	**Time**
Share	42.92%	29.22%	11.57%	0.76%	15.53%

QUALITATIVE FEEDBACK At the end of the user study, we assessed qualitative feedback. 8 participants liked the ease-of-use of *ESMAC*. For the *movisensXS* app, only 4 participants mentioned ease-of-use. 4 participants reported that their usage behavior changed during the course of the study – applying to the usage of both ESM apps. This was mainly due to an increased battery drain that led to a higher frequency of battery charging. In addition, these participants mentioned that they turned off the smartphone more frequently during the *movisensXS* part of the study due to the obtrusive notification modality to avoid undesired interruptions.

3.5 DISCUSSION

The number of participants in both evaluations was rather small. Especially for the evaluation of the web interface, we only had two experts who evaluated the system. This is not enough for a generalized interpretation, but provides first impressions *ESMAC* system. Results of this evaluation indicate that *ESMAC*, in its current state, is usable and comparable to the state of the art platform *movisensXS*. It also affirms further consideration and development of *ESMAC*.

Concerning the evaluation of the Android app, the number of participants was higher but still rather low with only 10 participants. Though, this number is considered to be sufficient to find large usability defects [145] and allowed us to infer principle issues that need to be addressed in future *ESMAC* versions. The evaluation also helped to reveal differences between *movisensXS* and *ESMAC* that might cause changes in the *ESMAC* systems or that influence a study designer's selection of one or the other system.

Both apps were considered comparable in perceived ease-of-use. However, they differed in terms of pleasantness as well as response time and response rate. *ESMAC* only uses a single notification whereas *movisensXS* uses the ringtone to inform about every prompt and applies a repeated alarm clock to remind users about unanswered prompts. It is comprehensible that a rather obtrusive notification alert such as a repeated auditory cue causes a higher perception of a notification which leads to a faster response time and, possibly, to a higher response rate as more notifications are perceived. However, obtrusive notifications might cause unpleasant disruptions and a low user experience. This might have less effect in a short-term study similar to our app evaluation (with a duration of 6 days) than in a long-term user study (e.g., with a duration of 4 weeks). Study designers need to find a compromise between high perceptibility of prompts and pleasantness to the user.

3.6 SUMMARY

In this chapter, we identified sensors, sensor formats, and events that are relevant for ESM experts. Based on these findings, we built *ESMAC*, a platform for creation of Android apps to conduct ESM studies. This platform consists of a web interface to configure the study, an XML-based format to transfer the configuration to a smartphone, and a smartphone app that prompts the participant for self-reports through smartphone notifications.

Both the web interface and the smartphone app were evaluated and compared to a state of the art platform, *movisensXS*, in terms of usability and user experience. Both apps achieved similar values which suggests that *ESMAC* can keep up with the state of the art.

During the evaluation of the app, we noticed that study designers have to find a trade-off between user experience and response time when configuring the prompting settings, especially the notification modality. An analysis of prompt triggers showed that 84.47% of all *ESMAC* prompts were event-triggered. This emphasizes the relevance of event-triggers and suggests further investigations.

4 Investigating the Usefulness of Event-Triggers for Experience Sampling Studies

The evaluation of ESMAC's app component in the previous chapter revealed that, if applied, a high share of prompts originate from event-triggers. This could be caused by coincidence or already indicate that event-triggers should not be neglected, but applied and investigated further. We hypothesize that event-triggers are useful for ESM studies, especially when the objective of the study is to assess information that is related to the events that trigger the prompts – e.g., when the objective of the study is to collect labels to annotate data. To investigate this hypothesis, we designed an ESM study to test the suitability of event-triggered prompts. We decided in favor of an ESM study focusing on location changes and location-based activities as these are two concepts that are of great interest, e.g., for interruptibility detection in human computer interaction [155] or for monitoring state changes in patients suffering from depression in applied psychology [65]. In addition, these two aspects relate to the perception of notifications [60] which is a central issue of this dissertation.

The results presented in this chapter were published and presented in an oral presentation at MobiHealth'17 [78].

4.1 Related Work

Several researchers acknowledged the potential of event-triggered prompts for ESM studies [140, 185]. However, this trigger type is still neglected or underestimated in related work.

In their survey paper, van Berkel et al. reviewed 110 ESM-related scientific works and observed that only 21 of them actually used event-triggered prompts and another 10 of them a combination of random and event-triggered prompts [50], i.e., only 28% of the reviewed papers considered event-triggers. However, a considerable amount of related work involved the assessment of location or communication data which offers the opportunity to trigger prompts event-based. Berkel et al. also reviewed tools for the creation of ESM apps [50]. Their report reveal that only half of the tools offered any event-triggers at all, even though 92% of them access sensors for data logging that would, in theory, allow to trigger event-based prompts.

A large data assessment project that used ESM for data assessment is Crowdsignals.io [37]. They collected survey responses to assess ground truth about user demographics, place labels, contact labels, activity intervals, and situational information such as well-being [38]. These responses were assessed using EMA, interval labels, and lock-screen surveys. However, only a fraction of the event-related features was collected through event-triggered prompts. Unfortunately, the project owners withhold details about the event-triggers they actually applied and only name geofence-based event-triggers as an example. Information about the current place or sedentary activity could have been assessed using event-triggers that detect location changes or changes in the physical activity.

Apparently, the full potential of event-triggered prompts is not yet used – possibly due to inexperience or missing research on the usefulness of event-triggers. We take a first step towards bridging this gap by providing insights about the usefulness of event-triggers for ESM studies focusing on location and activity changes.

4.2 USER STUDY

4.2.1 *Study Design*

To investigate the usefulness of event-triggered prompts in contrast to random and time-triggered prompts, we decided to conduct a field study. This allows to get feedback under realistic conditions. We used three study conditions, one for each trigger type. To keep the required number of participants to a minimum and to allow comparable circumstances for each condition, we decided to run the experiment as a within-subject study. We randomized the order of the conditions to avoid carry-over and learning effects.

We decided to restrict the time frame for prompts from 8 a.m. to 10 p.m. to allow the participants to rest over night without being disturbed. 14 prompts were sent out randomly over the day for the *random* condition. *Time*-triggered prompts appeared at each full hour, i.e., also 14 times. *Event*-triggered prompts appeared for each detected location change, i.e., the number varied per day and per participant.

To rate each trigger-type, we assessed the number of prompts, the response rate (number of prompts the participant reacted to), and the percentage of prompts that were triggered after an actual location or activity change.

4.2.2 *Location Change-Aware Experience Sampling App*

To assess location changes and activities we required an experience sampling app. We enhanced the ESMAC system which was introduced in Section 3.3.

First, we added a location change detection mechanism as a new event-trigger. We defined location changes as a situation in which a user showed movement behavior six times in a row. Movement behavior was defined as moving at least 60 meters in one minute, i.e., moving with at least 1m/s.

Next, we configured the ESM questionnaire. It consisted of questions about the current and last location and about the current and last activity.

Last, we had to configure the trigger type for each study condition. In the end, we had three different configuration files: each one for random, time-triggered, and event-triggered, respectively.

4.2.3 *Procedure*

At the beginning of the study, we met with the participants, explained the study and asked them to sign a consent form. Afterwards, we installed the app with the first configuration and assessed demographic information. The study lasted three weeks, i.e., with one week per trigger type. It took place during lecture time to guarantee fairly similar circumstances for each week. Data was collected from Monday to Friday. On the weekend, we exported and pseudonymized all log files and questionnaire responses from the smartphone, handed out feedback questionnaires about the experience with our app during the week and installed the new configuration file. At the end of the study, we assessed the general experience with our app over all three weeks.

4.2.4 *Participants*

Initially, 23 participants joined the study. However, 4 of them quit during the study and for 2 participants no data was collected due to technical issues. Three of the remaining 17 participants were female, 14 were male. All participants stated to be between 18 and 29 years old. We focused on students as participants as they are digital natives and accustomed to the usage of smartphones in everyday life. In addition, they have a regular week structure which guarantees comparable circumstances for all experimental conditions.

4.3 RESULTS

To rate the usefulness of event prompts in comparison to random or time-triggered prompts, we considered the overall number of prompts, the response rate, and the accurate detection of actual location and activity changes. The latter is represented by the relation between the number of questionnaires prompted after an actual location or activity change relative to the total number of prompts. Table 5 gives an overview of the results.

Table 5.: Overview of number of prompts, response rates, and the accurate detection of actual location and activity changes for each trigger type.

	Time	**Event**	**Random**
Number of prompts	62.80 (± 39.94)	19.50 (± 9.55)	62.70 (± 35.11)
Response rate	37% (± 10%)	43% (± 9%)	31% (± 18%)
Percentage of detected prompts after an actual location change	28% (± 16%)	71% (± 23%)	0.29 (± 19%)
Percentage of detected prompts after an actual activity change	41% (± 19%)	69% (± 27%)	0.37 8 (± 19%)

It might be surprising that some participants received less than the expected 70 prompts for random and time-triggered (5 days, 14 prompts per day). Apparently, some participants turned their phone off during the study which caused less prompts. What is visible is that event-triggered prompts were triggered less frequently but more accurately in terms of prompting after actual location or activity changes. Event-triggered prompts also show a higher response rate that might be caused by a higher user compliance due to fewer prompts and well-timed prompts.

To determine if the differences between the three trigger types are statistically significant or rather caused by coincidence, we ran correlation analyses. As the data is not normally distributed, we decided to perform parameter-free Friedman tests [89]. The results are listed in Table 6 to 9.

For all three aspects, the differences between *event-triggered and time-triggered prompts* and between *event-triggered and random prompts* showed p values below 0.05 and, thereby, statistical significance. This emphasizes that location-aware event triggers are useful for experience sampling studies focusing on location and activity changes.

Table 6.: Results of the pairwise comparison of all trigger types for the variable "number of prompts". Statistically significant results are marked: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Trigger Type 1	**Trigger Type 2**	**Mean Difference**	**p Value**
Time	Event	43.3	0.014*
Time	Random	0.1	1
Event	Random	43.2	0.005**

Table 7.: Results of the pairwise comparison of all trigger types for the variable "response rate". Statistically significant results are marked: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Trigger Type 1	**Trigger Type 2**	**z Value**	**p Value**
Time	Event	-1.988	0.047*
Time	Random	-1.682	0.093
Event	Random	-2.497	0.013*

Table 8.: Results of the pairwise comparison of all trigger types for the variable "percentage of prompts after detected location change". Statistically significant results are marked: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Trigger Type 1	**Trigger Type 2**	**z Value**	**p Value**
Time	Event	-2.805	0.005*
Time	Random	-0.459	0.646
Event	Random	-2.701	0.007**

Table 9.: Results of the pairwise comparison of all trigger types for the variable "percentage of prompts after detected activity change"'. Statistically significant results are marked: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Trigger Type 1	**Trigger Type 2**	**Mean Difference**	**p Value**
Time	Event	0.285	0.001**
Time	Random	0.04	1
Event	Random	0.326	0.048*

4.4 DISCUSSION

In our survey, they covered situations of interest more accurately and triggered less prompts overall. There results confirm the usefulness of event-triggers. However, we only ran one ESM study with the special focus on location and activity changes. For this kind of scenario, the usage of event-triggers is reasonable. There are other scenarios, e.g., gaining a broad overview of a participants daily activities, for which other trigger types are more suitable. It is always recommendable to select trigger types in accordance to the study objective.

In addition, we had a rather small sample and homogeneous sample. However, we wanted to focus on evaluating the trigger types and not inferring information from the participants themselves. Hence, the influence of the sample's homogeneity on the results might be marginal. Concerning the sample size, it was still big enough to yield significant results for several aspects of our research questions.

There were some technical issues that hindered the assessment of data for some participants. It is possible that these issues were caused by a disabled GPS connection or by the device being turned off. We are confident that such errors can be avoided in future user studies, e.g., by a regular check for GPS and a reminder to keep both device and GPS turned on.

Overall, we see potential for using event-triggers in ESM studies, especially if researchers are interested in investigating specific events.

4.5 SUMMARY

In this chapter, we investigated the usefulness of event prompts triggered by location changes in an ESM study focusing on location and activity changes.

Within a three-week field study we collected location change and activity information from 17 participants using three different trigger types for self-report prompts. We compared all three trigger types in terms of number of prompts, response rate, and accurate detection of actual location and activity changes. We found that the event-trigger scored best in all categories: fewest number of prompts, highest response rate, and most accurate detection. Statistical tests proof that the differences are statistically significant between *event-triggered and time-triggered prompts* and between *event-triggered and random prompts*. For event-triggered prompts, a low number of prompts co-occurs with a high response rate. We assume that this is due to a higher user experience: fewer prompts and prompts that relate to the current user context (location and activity change) result in a higher user experience and compliance. As a consequence, we suggest to use event triggers whenever an event-trigger is available that relates to items in the questionnaire. When gathering data labels, it seems unavoidable to rely on event-trigger to actually collect labels for situations of interest.

Part III.

Assessment of Perceptibility-Related Factors

5 Investigating the Detection of the Smartphone Position

Smartphones are an essential part of our everyday lives and accompany us almost everywhere as ubiquitous, personal wearables. The position of the smartphone, i.e., the location users store their device, is a basic property that influences the perception of notifications, e.g., in terms of selecting an appropriate notification modality [77, 91]. Different researchers already investigated the automatic inference of the smartphone position [35, 91, 121, 122, 177, 189, 198] with satisfyingly accuracy. To further improve the accuracy, we introduce a position transition correction (PTC). We assume that each position transition has to involve the "hand" state: to take the smartphone out of the trouser pocket and to put it into the backpack, it is necessary to pick up the phone, hold it in the hand and move it by hand from one position to the next one. Hence, we further assume that an apparent transition that did not include a hand state might be an error and not an actual position transition. Our correction mechanism builds up upon these assumptions.

Parts of this chapter have been published and presented as a poster presentation at MobiCASE'18 [80].

5.1 Related Work

Smartphone position detection was investigated in different ways before. Some researchers approached the task by recognizing the user's activity. Kunze et al. [122] identified a walking activity first and the device position next. They assumed that, while walking, certain movement patterns manifest themselves which help to classify the positions head, breast, and wrist. They applied a majority voting on the walking sequence and achieved a recognition accuracy of up to 100%.

Vahdatpour et al. [189] first identified walking sequences using unsupervised activity discovery. Next, they used support vector machines (SVM) to classify the on-body regions lower arm, upper arm, and head. Using a model trained on 500 randomly drawn samples from a dataset with 2500 entries, they achieved an accuracy of 89%.
Alanezi et al. [35] followed a more complex approach. They also based their investigations on an activity recognition. However, they did not limit themselves to the walking activity, but differentiated between idle, walking, and running. They presented a design for a recognition system and a first prototype.

Further related work classifies the smartphone position directly without prior activity recognition. Kunze et al. [121] classified positions during different everyday activities. Using a Hidden Markov Model (HMM) and a window size of 6 minutes, they achieved an accuracy of 82%. After merging front and back trouser pocket into one class, the accuracy rose up to 92%.
Shi et al. [177] combined measurements from accelerometer and gyroscope to estimate the rotation radius. Subsequently, they calculated features based on the rotation radius and the angular velocity. They considered the positions chest pocket, trouser pocket, belt bag, and hand. Using an SVM and five-fold cross-validation, they achieved an accuracy of 91.69%.
Wiese et al. [198] utilized accelerometer data to detect smartphone positions and investigated the usefulness of other sensors. The accelerometer data alone yielded an accuracy of 79%. Including further sensors such as proximity sensor and ambient light sensor increased the accuracy to 85%.
Fujinami [91] investigated smartphone position detection exclusively based on the accelerometer. They reached an accuracy of up to 80.1% for nine different position classes (around the neck (hanging), chest pocket, jacket pocket (side), front pocket of trousers, back pocket of trousers, backpack, handbag, messenger bag, and shoulder bag) and 85.9% for five different position classes (merging the four types of bags into one class and the two trouser pockets into one class).

It seems promising to rely on smartphone features, especially accelerometer data, to detect the device position automatically. For approaches that implement a position detection without prior activity recognition, the detection accuracies have room for improvement. Some researchers already included the user's hand as a potential position, but without special consideration regarding state transitions. To our best knowledge, only Antos et al. mentioned the meaningfulness of a hand state as a transition between different positions [40]. We will combine these ideas and present a position recognition approach based on smartphone features and a position transition correction based on the hand state.

5.2 COMMON SMARTPHONE POSITIONS

To assess where users commonly store their smartphones and which positions we should consider in our investigations, we ran a short online survey. Overall, 76 persons participated, aged between 17 and 36 with an average of 23 years (± 3). We asked our participants to imagine to perform common activities [13]: sit, stand, walk, jog, and ride a bicycle. For each activity, we assessed the estimated frequency with which the participants would store their phone in the following positions: trouser pocket, backpack, jacket pocket, purse, shirt pocket, wristband, belt bag, back pocket, on the table, or in the hand – positions considered in related work, e.g., [91, 177]. The results are depicted in Table 10. We decided to include all positions with at least 5% average usage frequency in our investigations, namely: trouser pocket, hand, backpack, jacket pocket, purse, and on the table. However, we excluded jacket pocket from our investigations as it we missed to specify in the online survey if the jacket is closed or opened. This fact might affect the way the smartphone is stored and, hence, hinder a clear detection of this position.

Table 10.: Common smartphone positions with their usage frequency.

Position	**Activity**					
	Sitting	*Standing*	*Walking*	*Jogging*	*Bicycling*	***Average***
Trouser Pocket	25.23%	30.42%	34.83%	45.75%	39.29%	**35.10%**
Hand	21.78%	23.16%	20.90%	19.84%	7.66%	**18.67%**
Backpack	11.52%	13.62%	14.83%	1.62%	26.13%	**13.54%**
Jacket Pocket	6.99%	10.74%	11.69%	13.77%	14.73%	**11.58%**
Table	25.59%	10.74%	4.27%	0.40%	0.20%	**8.24%**
Purse	6.26%	7.55%	8.54%	0.00%	6.68%	**5.81%**
Shirt Pocket	2.18%	3.18%	3.82%	3.24%	3.14%	**3.11%**
Wrist	0.36%	0.50%	0.67%	10.53%	0.98%	**2.61%**
Belt Bag	0.09%	0.10%	0.45%	4.05%	0.20%	**0.98%**
Back Pocket	0.00%	0.00%	0.00%	0.81%	0.98%	**0.36%**

5.3 PREDICTING THE SMARTPHONE POSITION

5.3.1 *Study Design and Sample Description*

We wrote an Android app to gather measurements from accelerometer, gyroscope, proximity sensor, ambient light sensor, and screen activity. Within the study, participants had to sit, stand, and walk, and optionally to jog or ride a bicycle. During each activity, the phone was stored at each considered smartphone position – excluding the combination hand and bicycle due to security concerns. For each combination of participant, activity, and position we collected one minute of data. Within an in-field study, we collected data from 20 participants. 6 of them were female and 14 were male.

5.3.2 *Data Preprocessing and Feature Selection*

As preparation for the data analysis, the collected data was preprocessed. First, the data was down-sampled to a common frequency of 30Hz. Next, it was transformed using Fast Fourier Transformation (FFT), if applicable. For each feature, we considered the following values: average per frame, average of the FFT bin, FFT max bin index, FFT sum of each quarter, highest / lowest / last value of the frame, first / third quantile, root mean square, standard deviation, sum of all values, squared sum, variance, and number of zero crossings. This lead to a total number of 198 features (11 sensor measurements * 18 values). The calculation of different features required the definition of a window size and step size for a sliding window approach. We evaluated different combinations and, eventually, chose a window size of 120 and a stepsize of 60 since they yielded the highest accuracy in a pretest (see Figure 11).

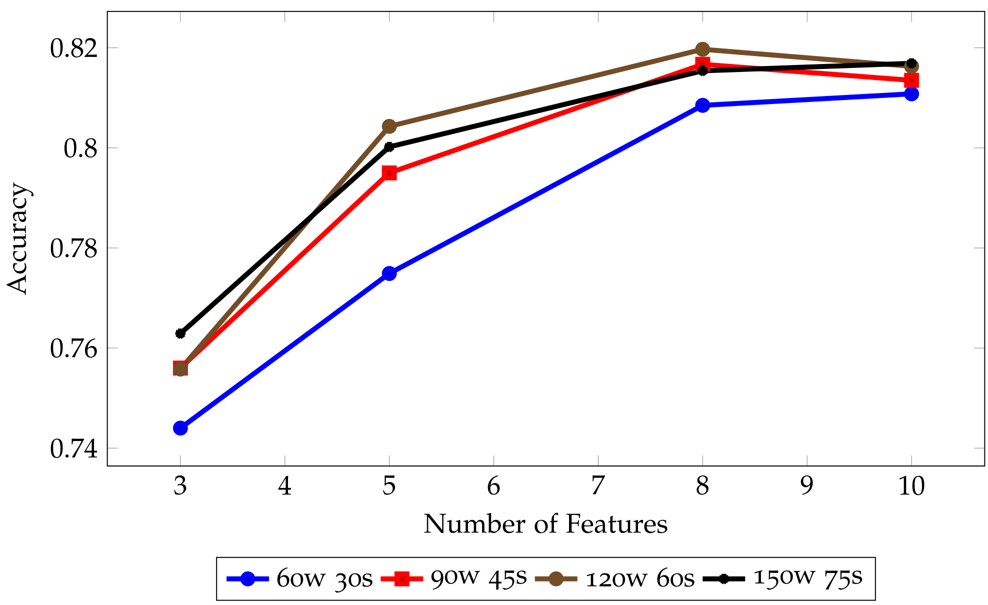


Figure 11.: Maximum accuracy for classification results based on datasets with different window and step sizes. w = window size in number of samples; s = step size in number of samples.

To reduce the high number of features for the classification process, we ran different feature evaluation mechanisms provided by Weka [103], namely: InfoGainAttributeEval, OneRAttributeEval, and SymmetricalUncertAttributeEval. The results are visualized in Table 11. In each case, the features derived from the accelerometer yielded the best results (top 10 per evaluator). Hence, we decided to rely exclusively on accelerometer data during the classification process.

Table 11.: Top 10 features selected by *InfoGainAttributeEval*, *OneRAttributeEval*, and *SymmetricalUncertAttributeEval*. Apparently, all top features were derived from the accelerometer sensor.

InfoGainAttributeEval	**OneRAttributeEval**	**SymmetricalUncert AttributeEval**
accelerometer_ Y_average	accelerometer_ Y_average	accelerometer_ Z_average
accelerometer_ Y_sum	accelerometer_ Y_sum	accelerometer_ Z_sum
accelerometer_ Z_sum	accelerometer_ Y_25%percentile	accelerometer_ Y_average
accelerometer_ Z_average	accelerometer_ Y_75%percentile	accelerometer_ Y_sum
accelerometer_ Y_25%percentile	accelerometer_ Z_sum	accelerometer_ Z_25%percentile
accelerometer_ Y_75%percentile	accelerometer_ Z_average	accelerometer_ Y_75%percentile
accelerometer_ Z_25%percentile	accelerometer_ Z_25%percentile	accelerometer_ Z_minimum
accelerometer_ Z_75%percentile	accelerometer_ Y_squaredSum	accelerometer_ Y_25%percentile
accelerometer_ Y_squaredSum	accelerometer_ Y_rootMeanSquare	accelerometer_ Y_FFT_sum_ 1st_quarter
accelerometer_ Y_rootMeanSquare	accelerometer_ Z_75%percentile	accelerometer_ Y_latestValue

5.3.3 *Classification*

Using the features identified in the feature selection process, we trained different classifiers, again provided by Weka [103], namely: a Support Vector Machine (LibSVM), two tree-based methods (RandomForest and RandomTree) and two instance-based approaches (KStar and IBk). We decided to use leave-one-person-out cross-validation. The accuracy value for each classifier is depicted in Table 12. The highest accuracy of 81.97% was achieved by the KStar classifier. A confusion matrix of the classification result is depicted in Table 13.

Table 12.: Maximum accuracy for recognizing smartphone positions per classifier.

Classifier	*LibSVM*	*Random Forest*	*Random Tree*	*KStar*	*IBk*
Accuracy	81.29	81.01	77.24	81.97	81.73

Table 13.: Confusion Matrix of a classification result gained by the *KStar* classifier based on 8 features selected by the *SymmetricalUncertAttributeEval* evaluator.

Actual Position	**Predicted Position**				
	Trouser Pocket	*Oh the Table*	*Hand*	*Backpack*	*Purse*
Trouser Pocket	*74.1*	0.4	8.7	9.3	7.5
Table	0.0	*100*	0.0	0.0	0.0
Hand	3.6	0.6	*84.3*	10.2	1.2
Backpack	7.3	0.3	11.3	*75.8*	5.3
Purse	11.5	0.0	12.1	15.3	*61.0*

5.4 POSITION TRANSITION CORRECTION (PTC)

5.4.1 *PTC Theory*

Antos et al. [40] already labeled the state during a position transition as *hand*, because their participants used their hands to change the device's position. Based on this finding, we assume that every significant position transition is realized using the hand. This assumption can be illustrated by the following example: a user takes the smartphone out of their trouser pocket (p_0) using their hand (h) and places it in their shirt pocket (p_1):

$$TrouserPocket\ (p_0) \rightarrow Hand(h) \rightarrow ShirtPocket(p_1)$$

Consider the following, exemplary classification result:

$$TrouserPocket\ (p_0) \rightarrow ShirtPocket\ (p_1) \rightarrow TrouserPocket\ (p_0)$$

If we assume that a hand position has to appear in between any other two positions then this example must contain a recognition error. Either, the hand state was missed, it was misinterpreted as a shirt pocket, or the device stayed in the trouser pocket the whole time and was incorrectly recognized as being in the shirt pocket.

Our PTC mechanism would inspect every window of data within the sequence. First, we look for each hand transition in the sequence. Next, we perform a majority voting on the transitions in between to decide in which position the smartphone is during that sub-sequence. An example for a successful PTC is visualized in Figure 12.

		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
Ground Truth	···	h	h	h	p_0	p_0	p_0	h	p_1	h	h	p_2	p_2	p_2	p_2	p_2	h	···
Classification	···	h	h	h	p_0	p_0	p_1	h	p_1	p_2	h	p_2	p_1	p_2	p_0	p_2	h	···
						p_0			p_1					p_2				
Correction	···	h	h	h	p_0	p_0	p_0	h	p_1	p_1	h	p_2	p_2	p_2	p_2	p_2	h	···

Figure 12.: A sequence correction that successfully reduced the number of errors.

5.4.2 *PTC Evaluation*

As input for the PTC evaluation we used a simulated sequence. The sequence was created from ground truth data and transformed using probabilities taken from the confusion matrix of the classifier results we gained from the leave-one-person-out cross-validation (cf. Table 13). This manipulation was necessary to create a sequence that is error-prone similar to an actual result of a smartphone position classifier.

To rate the PTC, we compare the ground truth information with the PTC-corrected version of the simulated sequence. Thanks to the PTC almost 50% of all errors could be reduced and the accuracy was increased to about 90%. However, we have to note that a good detection of the hand position is essential for the correct functioning of the PTC. If the hand state is not detected correctly, the PTC might worsen the results, as exemplified by Figure 13.

		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
Ground Truth	···	h	h	p_0	p_0	p_0	p_0	p_0	p_0	h	p_2	p_2	p_2	h	p_2	p_2	h	···
Classification	···	h	h	p_0	p_0	p_0	p_0	p_0	p_1	p_1	p_2	p_2	p_2	h	p_1	p_2	h	···
									p_0						p_1			
Correction	···	h	h	p_0	p_0	p_0	p_0	p_0	p_0	p_0	p_0	p_0	p_0	h	p_1	p_1	h	···

Figure 13.: A sequence correction that increased the number of errors.

5.5 DISCUSSION

There are different existing approaches that differ from our approach and which were able to detect the smartphone position accurately. One difference is the methodology: other researchers started with an activity recognition instead of detecting the position directly like we did. Direct classification seemed reasonable since the results of the online survey showed that users usually have preferred smartphone storage positions. Though, there were differences in the probability of a position per activity. It might make sense to consider this probability as an additional feature in future position detection approaches.

Another difference is the sensor selection: other approaches relied on a variety of smartphone sensors while we based our investigations of accelerometer data only, similar to some other approaches. This decision was based on the results of the feature selection we performed. However, the increasing number of embedded smartphone sensors and wearable such as smartwatches also lead to a higher number of possible features used for the classification. Future investigations might re-evaluate features and information gain to evaluate if other features proof useful. Classification results we gained were acceptable but with room for improvement.

The PTC algorithm we proposed was able to improve the accuracy for a generated sequence. However, it was not applied and evaluated in a real-world scenario. In addition, the accuracy of the PTC algorithm heavily depends on a correct detection of the hand state. This detection might be improved by considering new features as mentioned above: especially screen status and the existence of typing or touch interactions with the smartphone screen might indicate a hand state.

Eventually, we only considered a certain set of smartphone positions and users might store there device elsewhere, e.g., at a smartphone mount for bicycles. This is another restriction of our research that should be loosened in future studies.

5.6 SUMMARY

This chapter focused on predicting the smartphone position based on smartphone features while the phone is stored at different positions during different everyday activities.

First, we ran an online survey to assess positions at which the smartphone is commonly placed during activities such as sit, stand, walk, jog, and ride a bicycle. We identified hand, trouser pocket, backpack, purse, and on the table as common smartphone positions.

We collected data from 20 participants while they performed different everyday activities and stored the smartphone at different positions. We considered the smartphone sensors accelerometer, gyroscope, proximity sensor, ambient light sensor, and screen activity. After running different feature selection algorithms provided by Weka, we decided to focus on accelerometer data only. We predicted the smartphone position directly and did not perform an activity recognition first. Using common classifiers, again provided by Weka, we achieved recognition accuracies of up to 81.97%.

We also proposed a position transition correction (PTC). The PTC mechanism assumes that each position change has to include a hand transition. Applied to a simulated sequence of position changes, the PTC reduced the errors by about 50% and improved the recognition accuracy to about 90%.

Based on these accuracies, it seems promising to detect smartphone positions even based on accelerometer data only.

6 Investigating Methods for Smartphone-Based Location Assessment

Information about a user's location changes and activities – usually linked to a location – are useful for a variety of research fields, e.g., for investigating notification perception [60], for interruptibility detection in computer science [155], or for detection of states and state changes in patients suffering from affective disorders in clinical psychology [184]. The current location covers information about the place, related activities and might indicate a social activity [79]. Such information relates to the current level of interruptibility of a smartphone user [75, 155] and can be used to select an appropriate notification modality [77]. In addition, information about location and activity changes can provide insights about physical activity (lethargically staying at home vs. moving from one place to another) or avoidance of other people (staying at home vs. changing location). These are symptoms of depression [65] which relate to states and state changes in affective disorders such as the bipolar personality disorder [101]. Traditionally, experience sampling questionnaires are used for the assessment, i.e., users are asked for their current location and on-going (social) activities on a regular basis. However, questionnaire prompts cause a disruption in the user's daily routine [66, 170]. In addition, ESM questionnaires intended for assessment of location-related information should be triggered by location changes which requires a method to detect location and location changes automatically. However, an automatic assessment of location information is a sensitive issue as many users are concerned about revealing their exact position, in fear of being tracked. We investigated two alternatives: a WiFi-based approach and the usage of abstract locations in form of place types provided by the Google Places API.

Contents of this chapter have been published and were presented in oral presentations at different occasions: MobiHealth'16 [76], the UbiTtention workshop of UbiComp'16 [75], and MobiCASE'18 [79].

6.1 RELATED WORK

Smartphones offer different sensors to assess location and location changes. The most common example is GPS [33, 69, 155, 193]. However, this is delicate in terms of data protection as it reveals the actual position in fine granularity. In addition, it is fairly expensive in terms of energy consumption. Therefore, we seek alternative assessment methods. Low-cost alternatives for location detection are GSM, Bluetooth or WiFi [93, 152].

GSM is a standard for mobile communication and digital cellular networks. Via GSM it is possible to identify the cell tower a mobile device is connected to or to create a fingerprint of nearby cell towers. This allows a location detection with coarse granularity [69]. However, location tracking via GSM is too inaccurate for our setting as a person can be connected to two different towers while being at the same location (urban areas) or be connected to the same tower but changing location significantly (rural areas).

An alternative used in [155] is *Bluetooth*, a wireless technology that allows data exchange over short distances. It is possible to create location fingerprints based on nearby devices, even though this is not always accurate. Alternatively, it is possible to equip locations with Bluetooth beacons and identify locations by their unique beacons. However, the process of labeling locations with Beacons is inefficient and not suitable for real-world scenarios in which the user destination is not known beforehand. In addition, smartphone manufacturers restrict the visibility of devices via Bluetooth: the device is only visible to other devices while the Bluetooth settings menu is the app running in the foreground [2].

WiFi is a technology that allows devices to connect to wireless LAN (WLAN). Nowadays, WiFi is often used as a synonym for WLAN as most WLAN rely on this standard. Within the last years, WiFi access became omnipresent [26], especially in urban environments. To our best knowledge, WiFi was so far only considered for location detection in form of WiFi fingerprints which were labeled in advance [55, 141]. Again, the pre-labeling is expensive and not suitable for real-world assessment.

It is apparent that relying on WiFi data is a promising approach, since it is suitable for real-world scenarios if no pre-labeling is required. Hence, as a first option for location assessment, we investigated an approach to detect location patterns based on unlabeled WiFi fingerprints.

An alternative location representation are place types. The survey in Subsection 3.2 confirmed that abstract location is an acceptable presentation format of a user's location. Place types were investigated in related work before, especially in the area of interruptibility detection [183, 155]. In addition to the beforementioned assessment methods based on GPS, Bluetooth fingerprint, and WiFi fingerprint changes, Pejovic et al. rely on user-provided location information ("residential", "work", "public") which, however, are only available in case a user responded and are limited to these three options. However, for many applications, it is insufficient to just differentiate between residential, work, and public, e.g., to accurately predict interruptibility location-based only. More precise location types should be considered. Ter Hofte [183] analyzed interruptibility in relation to self-reported location, among others. However, self-reported locations are difficult to generalize and might underlie recall bias. Hence, we propose to use a common basis of place types and to assess them automatically.

The process of place identification was investigated in depth in a PhD thesis by Nurmi [146]. He defined his own place identification process consisting of data preparation, preprocessing, clustering and analysis in combination with a labeling process. However, this is a fairly complex process.

Related work showed the benefits of place types, but also some drawbacks of recent assessment methods. Based on our findings gained from literature screening, we decided to investigate a detailed but generic set of locations provided by the Google Places API as a second option for automatic location assessment.

6.2 Investigating the Suitability of WiFi for Location Assessment

6.2.1 *Assessment App*

We built an Android app to collect WiFi information using the 2.4GHz and the 5GHz frequency band. We decided to assess MAC addresses instead of WiFi SSIDs, because networks may use the same SSID or broadcast their SSID from multiple access points. We only want to consider long-term stays at locations and location changes between them. This shall avoid that passing a location during a transit of the user is counted as a location change. We decided to log the currently available WiFi networks every five minutes which is a very energy-efficient sampling rate. Location changes can be identified by comparing the WiFi networks detected at two or three consecutive points in time. In this context a location change occurs when none of the set of access points recorded in measurement M_1 is present in the set of access points recorded in measurement M_2 (five minutes later) or M_3 (ten minutes later).

6.2.2 *Study Design*

To investigate the suitability of WiFi information for location assessment, we conducted a user study. The study lasted ten days and was taken by 10 participants, 7 of them male and 3 of them female. They were aged between 21 and 68 with an average age of 29. In a first meeting, all participants were informed about the objective of the study and the data that is captured by the app. Afterwards, we asked them to sign a consent form and installed the smartphone app after confirmation of the participant. During the course of the study, participants were supposed to use their own smartphones and to use them as usual. As ground truth, participants recorded their location changes in a chart, providing start and end times for stays at every location they visited. We differentiated between *private*, *business*, and *public* locations, analog to location labels in [155]. We asked our participants to provide the type of location from which they were coming and going to in the format of "A→B".

6.2.3 *Results*

We could only include datasets from nine participants in our analysis. One participant missed to restart the app after a smartphone reboot. During the study, 17,406 different access points were detected overall and between 458 and 3426 per participant. We counted 1,065 location changes in total and between 33 and 208 per participant.

The collected data was compared to the manually recorded ground truth to determine the recall (a.k.a. sensitivity) of our detection. This metric specifies how many true location changes were detected correctly by our approach. When considering only the last two measurements, i.e., the current location in comparison to the location detected five minutes ago, we only achieve an average recall of 80%. The high number of errors is caused by fluctuations in the WiFi data, e.g., due to loss of WiFi connection during the measurement process or caused by participants moving to another room within the same but large building. Hence, we decided to consider the last three measurements, i.e., the set of WiFi MAC addresses seen within the last 10 minutes. Now, we achieve an average recall of 98%. The remaining errors are caused by location changes that were logged by the participant but were shorter than the five or ten minute minimum that our app requires. Alternative reasons for errors are the potential loss of a WiFi connection due to an energy-saving mode of the smartphone.

6.3 investigating the suitability of place types for location assessment

6.3.1 *Reduction of Place Types*

The Google places API offers more than 120 place types [98]. Not all of these places are necessarily visited on a regular basis. We screened the list and, in a first step, reduced it to 89 place types by removing obviously meaningless places (e.g., "synthetic geocode"), by combining essentially identical place types (e.g., include "ATM" in "bank" or combine "bus station" and "subway station"), or by removing too generic places (e.g., "point of interest"). In a next step, we handed out a list of these 89 place types to ten participants which were recruited randomly at the city center. Most of them were students, 6 of them were male and 4 female. We handed them a sheet of paper with all place types and ask them to cross out all locations that they visit less than once a month. Next, we collected the remaining place types and counted how many participants stated to visit this place. We kept all places that had a count of 4 or higher. This is a reasonable threshold, because it reduces the list with a 0.95 confidence and 0.31 margin of error. That means that a majority of the represented population would visit this place regularly, with the error leaning towards keeping too many places instead of deleting too many. Eventually, the list was reduced to 20 place types which should be considered in future location-aware user studies:

- Bakery
- Bank
- Bar
- Bus / subway station
- Café
- Clothing store
- Gas station
- Grocery store
- Gym
- Library
- Meal takeaway
- Movie theater
- Night club
- Park
- Parking
- Post office
- Restaurant
- Shopping mall
- Store
- University

6.3.2 *Location Categories*

To have a more abstract taxonomy, we investigated categories for the identified place types. Zheng et al. [203] propose to categorize places as *Food & Drinks*, *Sports & Exercises*, *Movies & Shows*, *Shopping*, and *Tourism & Amusement*. This categorization covers private contexts pretty well but lacks a business category.

Riboni and Bettini [165] rather focus on differentiating between private and business matters and propose *Communication / Meeting*, *Play*, and *Social Business Activity*. Liang et al. [128] propose a similar categorization and suggest *Work*, *Play*, *Develop*, and *Connect*. Liao et al. [131] include *Work* as well as *Sleep*, *Leisure*, *Visiting*, *Pickup*, and *On/Off Car*. However, these three categorizations are rather abstract and raw especially about private contexts.

We base our categorization on the one suggested by Zheng et al. They have a representative selection of categories that covers private contexts very nicely. Since they lack a business context, we propose to include the category *Work and Education* to cover both business matters and education such as being at school or at the university. We further propose to rename the category *Tourism & Amusement* into *Recreation & Amusement* to include recreative activities.

6.3.3 *Precision of the Google Places API*

To evaluate how well the place recognition itself worked, we compared the place types detected by the Google Places API with the place types provided by study participants (ground truth). We ran a three-week user study in which 24 users participated, 10 of them female and on average 24 years old. At the beginning, we explained the objective of the study, asked the participants to sign a consent form and installed a smartphone app. The purpose of the app was first to assess the location by picking the most probable place type identified by the Google Places API and second to ask the participants if this location is correct and, if not, to provide the correct place type. Based on their feedback, we calculated the precision of the Google Places API by counting how often the users rejected the suggested place and picked a different one. If the user labeled the place as "at home", "on the way", "work", or "other" the datum was not counted, because those places were not detectable by the Google Places API. These places might be detectable by WiFi (identification of "home" or "work" due to automatic connection with a WiFi with a specific SSID) or the Google Activity Recognition API (identification of "on the way" through an activity recognition of "in transit", "on foot", "in vehicle", or "on bicycle"). Based on the ground truth, we calculated a recognition precision of 73%. This result is significantly better than guessing.

Results might be enhanced by considering more place types than only the most probable one that is returned by the Google Places API. We only considered the place type that had the highest probability. However, the Google Places API returns a list of suitable place types with probabilities. Elhamshary and Youssef already showed that considering the top 5 venues is advisable: their approach yielded a 99% precision for the actual venue to be in the top 5 candidate list [73]. For future studies, a weighted approach considering the five most probable places should be considered.

6.4 DISCUSSION

There are several options for location assessment. Depending on the user study, it might be reasonable to select a more privacy-aware methods while in other studies it might be necessary to rely on actual fine-granular GPS coordinates. For scenarios such as the selection of suitable notification modalities, interruptibility detection, or the monitoring of state changes in patients suffering from affective disorders, it is sufficient to have information about location changes in general or the kind of location. If the focus is on monitoring location changes only, it is sufficient to rely on WiFi information: they provide insights about movement between different locations without revealing the kind of location or the actual position of smartphone or user. In addition, most smartphone users usually have WiFi enabled most of the time to save precious mobile bandwidth while they might tend to turn off their GPS to conserve battery power. The drawback of WiFi fingerprints is that they do not allow any semantic interpretation unless they are labeled, e.g., as home or work environment. This is a beneficial characteristic of place types: they can reveal location changes but also provide further information about the nature of the location, its category, or location-based activities. However, the identification of a place types requires enabled GPS. Smartphone users might turn off their GPS out of fear of being tracked or to conserve battery power. In addition, GPS might underlie drifts or inaccuracies that cause wrongly detected place types. A location assessment methods should be selected depending on the scenario and the need for semantic information.

6.5 SUMMARY

Assessing information about movement and location changes requires automatic, privacy-aware and energy-efficient approaches.

We presented results of a study conducted to investigate the feasibility of purely WiFi-based detection of location changes using smartphones. A recall of 98% proves a successful detection of location changes by our approach and shows the approach's high potential for application. Apart from the mere number of location changes, WiFi information can reveal regularity, duration, and frequency of location visits. These aspects can give a deeper insight into a user's receptivity and interruptibility, but might also support the monitoring of affective disorder symptoms. However, WiFi-based location detection only allows the detection of a pre-defined place or the detection of a location change. It does not allow a semantic interpretation of the location – or only if the WiFi fingerprint was labeled manually before. There is no possibility to infer further context from a WiFi-based location such as location-based activities or a social activity indicator. In addition, it is not possible to gain abstract information about the location such as its type or category – which might reveal information about user preferences or characteristics.

Place types, in contrast, offer a semantic and allow an interpretation of a location in terms of category, location-based activities or, possibly, even social activity. They can reveal user preferences (e.g., a restaurant to a fast food restaurant), reveal hobbies (e.g., being a cineaste or athlete) or personal traits (e.g., regular visit of crowded places or frequent visit of new places). As shown, the precision of considering only the most probable identified place type is good, but not precise enough. Based on results from related work [73], we propose to consider up to 5 most probable places. In addition, it might be advisable to consider place categories [203]. They are especially useful to infer general findings about a category to new place types whose characteristic is not yet known but it is assigned to a category.

Based on our findings, we recommend to rely on place types to identify the current location of a user and the user's location changes. For places that are visited frequently but not available by the Google Places API, we recommend to apply a WiFi-based assessment to label these places in a privacy-aware manner. This might be applicable to the place types "home" or "work".

Location is often considered in combination with social activity, e.g., for interruptibility detection [155] or monitoring of state changes in affective behavior [83]. Hence, it is a logical next step to investigate correlation between both and the possibility to predict social activity based on the place type.

7 Investigating The Location and Activity-Based Estimation of Being in Company

The current social context, i.e., if a user is *in company* or *alone*, is a useful information for various disciplines such as HCI. For example, it influences our interruptibility and how we respond to smartphone notifications [88, 155, 183]. Also, the social context or a change in social context might be useful to support the detection of states and state changes in bipolar personality disorder or depression to perform an appropriate treatment [83, 184, 186].

Commonly, the social context is provided by the users themselves via self-reports at discrete and sparse points in time. This does not allow a continuous monitoring that captures changes in social activity. In this chapter, we explore a possibility to detect whether a user is *in company* or *alone* based on different place types, temporal features, and the user's activity. First, we propose a relationship between different place types and the probability of being *in company* or *alone*. We investigate this assumption within an online survey. Second, we enrich location features with temporal features and activity, because activities change during the day according to our biorhythm and habits [42, 102]. This approach has been evaluated within a field study. Analyses include identification of feature importance and evaluating predictive models based on their accuracy.

The contents of this chapter have been published and presented in an oral presentation at MobiCASE'18 [79].

7.1 Related Work

There is a vast amount of research investigating the detection of social activity including social sensing [69, 199], group activity recognition [99, 108], or flock detection [117, 127]. Many of these approaches rely on Bluetooth-based recognition of nearby devices [69, 128, 199]. However, due to raising privacy-awareness and security reasons, the visibility of Bluetooth-enabled devices was restricted by various mobile OS vendors during the last years. Smartphones with active Bluetooth are only visible if the user is currently in the Bluetooth settings. Hence, this approach is no longer an option. Alternative approaches for social sensing, e.g., by Pentland et al., include the collection of location data and its storage on a server [69]. Every phone at which this app was installed provides data and allows an online comparison of this data to check if devices are nearby. Due to privacy reasons, we focus on a detection approach that relies on data from one single device only. All processes run on the device itself without relying on an external server.

The usefulness of activity, location, or temporal features for group activity detection has already been confirmed by related work. A common method is to extract information from videos and analyze it with the objective to differentiate activities which can later on be labeled as group or single activities [39, 57, 169]. Some of these approaches focused on the spatio-temporal evolution of crowd behavior, so-called crowd context [57], while others relied on temporal and spatial information [39, 169] – proving that spatio-temporal data is well-fitted for recognizing group activities. However, these approaches have the drawback of using intrusive, non privacy-aware, and high energy-consuming video techniques. It would be less energy-consuming and more privacy-aware to predict being *in company* based on more abstract smartphone data, gathered automatically and harvested in an energy-aware manner. In addition, it would be more privacy-aware to process all data directly on the phone itself.

A relation between self-reported place types and social activity was already found, e.g., to infer interruptibility [155]. In contrast to the mentioned related work, we focus on automatically detected location and place types as they are generalizable and do not require user involvement. In addition, place types are more abstract and hence more privacy-aware than raw GPS values. Our idea is to combine location data with temporal features and activity information. While related work focused on detecting groups and group activity we choose a more abstract approach and focus merely on recognizing being *in company*.

7.2 ONLINE SURVEY

We conducted an online survey to assess if users tend to visit a place rather *in company* or *alone*. We highlighted that being *in company* applies even if the user is only accompanied by one other person. The survey was conducted in German, but will be presented in English within this Chapter for a better understanding.

As mentioned before, locations were based on the place types that are available through the Google Places API [98]. To reduce the number of questions within the survey, the high number of over 120 place types was reduced to 20 places as explained in Subsection 6.3.1. In addition, we considered place categories as introduced in Subsection 6.3.2.

For each place type we asked:

1. "In which category would you assign the currently displayed place type?" and offered categories in form of *select many* checkboxes
2. "Do you visit the displayed place type rather alone or in company?" and offered a rating in form of a 5 point Likert scale ranging from 1 ("always alone") to 5 ("always in company")

The categories were assessed to be able to abstract a probability of social activity for more abstract places. This might proof useful in the future as it allows to include new place types for which only the category but no probability for social activity is known. The answers to the Likert scale can be interpreted numerically as a likelihood of being *in company*, i.e., 1 being "always alone / never in company" and 5 being "never alone / always in company".

PARTICIPANTS The survey was created with Google Forms and performed online. To recruit participants we spread the link to the survey via social media. 68 people answered the survey, 50% male and 50% female. The average age was 33 years ($\pm$12). Almost all participants had a school degree that qualified them for higher education. 63% even had a university degree which is a strong bias. The largest occupational category was information- and communication technology.

RESULTS The results of the survey are summarized in Table 14 and 15. Analyzing the place types (cf. Table 14), it is visible that users are usually *in company* when visiting night clubs, bars, movie theaters, restaurants, and cafés. In contrast, users tend to visit post offices and gyms preferably *alone*. In addition, there are some places which are visited *alone* as well as *in company*. Prominent examples for these are shopping malls, universities and meal takeaways. For these places more information about the users and their activities are required to decide whether they are *in company* or *alone*.

Considering categories (cf. Table 15) we computed a likelihood for being *in company* while being in such places. Attending *Movies & Shows* is usually done *in company*. According to the place categories, *Recreation & Amusement* and *Food & Drinks* locations are visited *in company* in 2 out of 3 cases. *Sports & Exercises* are performed *in company* only in 1 out of 3 cases, probably depending on the kind of sport. *Work & Education* and *Shopping* are not decidable. The decision probably depends on the purpose of the business (e.g., having a meeting vs. writing a paper) or the shopping purpose (e.g., doing the weekly shopping vs. buying new clothes).

Table 14.: Average answer per place type stating if a user visits a place rather *in company* (5) or *alone* (1).

Place Type	**Average**	**Standard Deviation**
Night Club	4.74	0.56
Bar	4.65	0.54
Movie Theater	4.49	0.73
Restaurant	4.37	0.69
Café	4.08	0.71
Park	3.39	0.85
University	3.11	1.10
Shopping Mall	3.03	0.68
Meal Takeaway	2.87	0.75
Clothing Store	2.76	0.82
Parking	2.77	0.76
Store	2.69	0.62
Bus / Subway Station	2.60	0.65
Grocery Store	2.37	0.75
Bakery	2.29	0.55
Gas Station	2.28	0.76
Library	2.11	0.97
Bank	2.04	0.86
Gym	1.89	1.12
Post Office	1.77	0.64

Table 15.: Likelihood of being *in company* per defined place category.

Place Type	**Likelihood**
Movie & Shows	84.8 %
Recreation & Amusement	67.2 %
Food & Drink	61.7 %
Work & Education	40.2 %
Shopping	39.0 %
Sports & Exercise	33.3 %

Overall, it becomes clear that location alone is not a distinct feature to differentiate between being *in company* and *alone*. It seems necessary to investigate its combination with further contextual data – such as activity or time. Thus, we conducted an in-field user study to collect and analyze data.

7.3 USER STUDY

STUDY DESIGN The purpose of the study was to gain insight about the context in which people are *in company* or *alone*. The study lasted three weeks.

There was an initial meeting with the participants in which we described the purpose of the study. Participants were free to ask questions about the study. We informed them that they were free to drop out of the study if they feel uncomfortable at any time. Afterwards, we asked them to sign a consent form to confirm their participation and to allow us using their personal data anonymously and for scientific purposes only. Next, we installed our app on their smartphone. We asked the participants to keep the location service enabled and only switch it off if they need to, for example, to save battery, if they do not want a place to be recorded, or when they are outside the country and needed to prevent network access. We explained to them how to respond to notifications and how to add data later on using the retrospective log functionality of our app. After three weeks, we met again to export the recorded data and to ask the participants for feedback, such as problems or difficulties.

PARTICIPANTS We recruited 30 participants, of which 24 started the study. The others had exclusion criteria, such as not having a cellular connection for large parts of the study or finally decided not to participate in the study because of privacy concerns. The participants were between 19 and 31 years old with an average of 24 years. 10 participants were female and 14 male. There was an equal distribution of students and working population.

ASSESSMENT OF SMARTPHONE FEATURES We developed an Android app to assess the desired features: place types (via Google Places API), user activity (Google Activity Recognition API), and temporal features (via system time).

To assess the location, we sent longitude and latitude values to the Places API which returned a collection of *PlaceLikelihood* objects, one for each probable place the user could currently be at. For simplicity, we visualized the structure of such a result returned by the API in JSON notation (see Listing 7.1). We decided to always consider the most likely place and the first (i.e., most suitable) place type.

The Activity Recognition API relies on data from physical sensors such as accelerometer and gyroscope, but also GPS. It returns the most probable activity and the confidence of the activity classifier.

For temporal features, the app stores the internal system time as a unix timestamp. From the timestamp, we can derive features such as hour of day, day of week, or workday.

```
{
   "likelyPlaces": [
      {
         "likelihood": 0.4,
         "place": {
            "name": "ZKM Karlsruhe",
            ...
            "placeTypes": [66, 5, 1013]
         }
      }, {
         "likelihood": 0.12,
         "place": {
            "name": "Filmpalast",
            ...
            "placeTypes": [64, 34]
         }
      },  ...
   ]
}
```

Listing 7.1: Simplified exemplary result of a Places API call represented in JSON.

ASSESSMENT OF GROUND TRUTH DATA At every location change, i.e., whenever the app detected a new place type, the user was prompted for feedback by a smartphone notification. The user had the choice to respond immediately or to add the information later using the retrospective log functionality. The questionnaire was delivered in German, but will be presented in English for a better understandability. Whenever reacting to the response, promptly or later on, the user was confronted with questions similar to the following example:

1. Are you currently at this place? *University*
2. *If yes:*
   a) Are you *in company* at this place?
      → push either *in company* button or *alone* button
3. *If no:*
   a) At which place are you currently?
      → select place type out of a drop down list
   b) Are you *in company* at this place?
      → push either *in company* button or *alone* button

The retrospective log function of the app presented a list of all places that a user visited that day. Each row showed the time of arrival and departure as well as the detected place type. Only visits of the same day were shown. A longer period would require to show dates as well and would eventually bloat the list with too many entries. In addition, retrospective bias or memory gaps might have occurred. A click on a list entry started the same interface that was used in case a user responds to a feedback prompt. This ensured that the user did not have to learn a new design but was already used to the same feedback interface.

7.4 RESULTS

7.4.1 *Descriptive Data Analysis*

The final dataset consisted of 1745 instances from 24 different participants. There were more instances in the dataset of class *in company* (993) than *alone* (752). A common comment was that the participants sometimes felt it was difficult to decide whether to declare a situation as being *in company* or *alone*, because there were other people present but the degree of social interaction was low.

Table 16.: Number of recorded visits for each place type ordered by the probability of being *in company*.

Place Type	**Occurrences**	**% in Company**
Movie Theater	1	100.00
Bar	22	95.45
Meal Takeaway	12	91.67
Work	72	90.28
Restaurant	41	87.80
Café	27	85.19
Clothing Store	12	83.33
Gym	11	81.82
Other	48	77.08
Shopping Mall	18	61.11
Store	12	58.33
Grocery Store	39	51.28
Home	697	50.65
Bakery	6	50.00
Bank	3	33.33
Bus / Subway Station	77	29.87
Gas Station	7	28.57
Park	8	12.50
Parking	8	12.50
Post Office	1	0.00

Place types were also imbalanced and distributed very unevenly. As Table 16 shows, only 13 places had more than 10 occurrences. Some places, on the other hand, were strongly represented, e.g., "home" with almost 700 instances.

ANALYSIS OF PLACE TYPE, TIME, AND ACTIVITY Figure 14 presents the distribution of being *in company* or *alone* plotted against all considered place types and the hour of day. "Bars" and "restaurants" were frequently visited *in company*. In contrast, "bus / subway stations" were mostly visited *alone*. The distribution of being *in company* or *alone* was rather balanced for places such as "home", "on the way", "other", and "university".

Focusing on the hours of the arrival times (y axis) it can be seen that firstly, place and arrival time were dependent and secondly, that at night many places were visited *in company*. Though, there was not very much data with this pattern.

It is visible that there were more records of activities performed *alone* than *in company*. This phenomenon might be biased by the labeling process. If a participant labeled data only in case of being *alone* and never while being *in company* – for example, because it would be impolite to use the smartphone while being with others – such an imbalance could occur. In addition, the definition of being *in company* was strongly dependent on the participant's interpretation. According to the user feedback it was also hard to judge where being *in company* began and where it ended. One example for this is "home" where it was not easy to tell if the fact of living in a shared apartment or with a partner always counted as being *in company* or only when performing joint activities. For "home" and "university", it happened that participants were *in company* but not actually involved in a common activity. This is obvious for such place types as there might be individual activities performed in the presence of other people. In these cases the place alone seems to be no useful indicator for social activity.

As shown in Figure 15, user activity as a single feature did not yield a clear separation between being *in company* or *alone*. The closest explanation would be that the shares of activities were not sufficient to distinguish between the classes, and time sequences play a larger role.

7.4.2 *Feature Analysis*

To assess the quality of our features we calculated information gain and χ^2 with Cramér's V. Those numbers reveal how the features perform and how they compare to each other. However, they do not reveal how combinations of the features might be correlated to being *in company*, but give a tendency. The combination of features is measured with the performance of the classification.

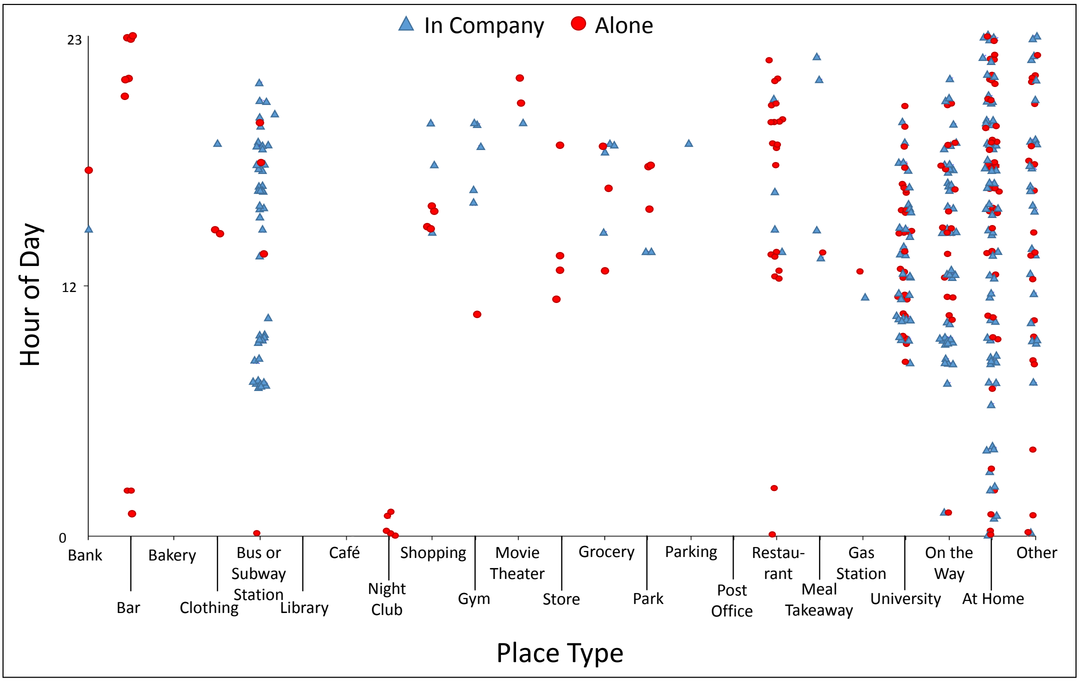


Figure 14.: Distribution of being *in company* (blue triangles) or *alone* (red circles) at a specific place type (x axis) at a specific hour of day (y axis).

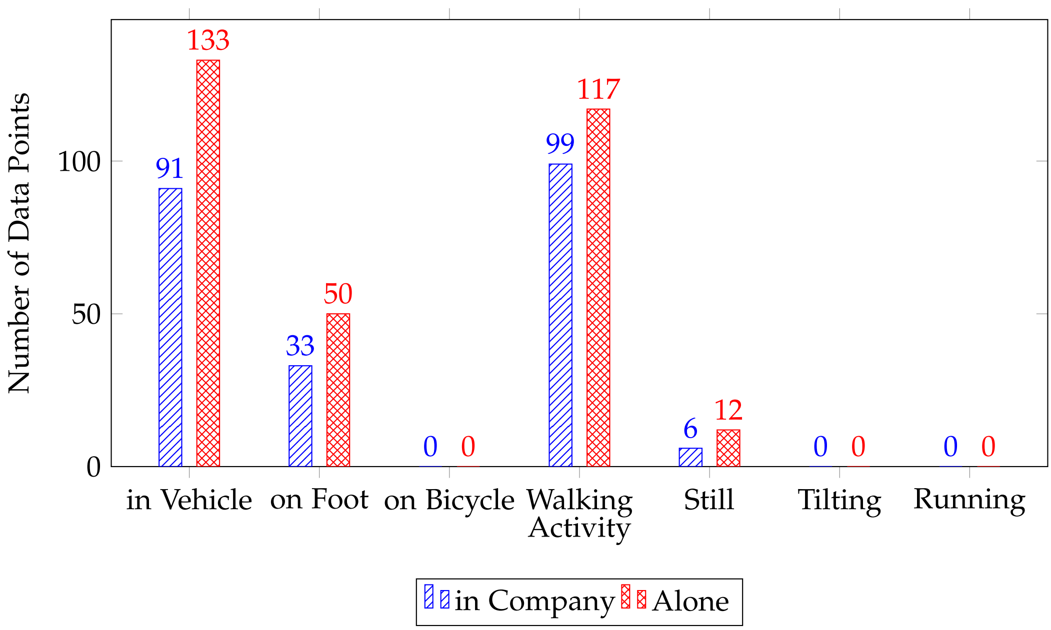


Figure 15.: Distribution of being *in company* (blue) or *alone* (red) per activity.

INFORMATION GAIN Information gain is a measure of how much the entropy of the class distribution is reduced when only considering the different values of a feature. The maximum information gain is 1. An information gain of 0 means that the feature has no additional value for the classification. A reduction of entropy is desirable. For example, if the data is separated by place type, it would be beneficial if within each place the class value would either be mainly *in company* or *alone*. The stronger the social activity indicator leans to one side, the lower the entropy. Information gain cannot be calculated on numeric attributes. Therefore, we binarized numeric attributes, i.e., transformed the attribute into the values zero and non-zero.

Table 17.: Overview of the information gain for each feature of the mixed dataset for predicting being *in company* or *alone*.

Feature	**Information Gain**
Place	0.11481
Weekday	0.02976
Hour of Day	0.02678
Activity	0.00113

Table 17 shows the information gain for each feature. Place is by far the best feature according to this metric. Temporal features perform not as good but still provide some gain. User activity however only yields marginal information gain in the magnitude of statistical error. Based on the information gain the most valuable features are place, weekday, and hour of day.

χ^2 AND CRAMÉR'S V χ^2 is a metric to test distributions of variables for independence. It is calculated by the sum of squared differences between observations. The corresponding p value indicates the likelihood that the difference in the observations is caused by statistical error. The purpose of the test in the present case is to see if the selected features are actually dependent on being *in company* and, most importantly, if the results are statistically significant despite the low amount of data.

Cramér's V is a measure of association between two variables and is related to the χ^2 test. It shows the correlation strength between variables. We used the bias-corrected version of Cramér's V which is utilized to ensure comparability between features that differ in the number of values [49]. This measure helps to judge which features are worth to be investigated further and which are neglectable.

Table 18.: Overview of χ^2 values for each feature and its significance and effect size in form of p-values and Cramér's V, respectively. Statistically significant results are marked: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Feature	**Pearson's χ^2**	**p Value**	**Cramér's V**
Place Type	104.42	< 0.001***	0.245
Hour of Day	38.777	0.02099*	0.149
Weekday	5.708	0.04567*	0.057

Table 18 shows χ^2 values for all features compared to the class attribute, i.e., being *in company*. It also indicates whether there is a significant correlation between the feature and the class attribute, determined by the p value. If this is the case, Cramér's V is provided to indicate the strength of the correlation. The place type certainly seems to be an indicator for being *in company* with a clear correlation expressed by a V of about 24%. Hour of day is also correlated with statistical significance and a V of about 15%. According to Cohen [58] both qualify as a weak effect size. The weekday has no apparent significance which might be caused by the inhomogeneity of the sample as students and working population have different schedules for each day.

Since place type and all temporal features correlate with statistical significance and a small effect size, they are considered useful in classification.

7.4.3 *Prediction of Being in Company*

PRELIMINARY CONSIDERATIONS To evaluate the predictive power of the identified features, we built a classification model to estimate if a user is *in company*. The classification result is binary: a user is either *in company* (1) or *alone* (0). Pure guessing would result in 50% accuracy on average. However, there are more instances in the dataset of class *in company* than *alone*. Always choosing *in company* would result in 57% accuracy on average. This value represents the baseline for the recognition accuracy of our predictive model. The optimal target accuracy heavily depends on the use case of the social activity estimation. For an application in clinical psychology with a socio-psychological component, an accurate indicator for social activity is of high interest and high recognition accuracies are required; misclassification might lead to misdiagnoses and wrong treatment. For context-dependent notifications lower accuracies might be more acceptable as missing a notification or being notified one more time might not have that severe consequences. Still, the accuracy should be as high as possible with only marginal error.

CLASSIFICATION Based on the identified features, we evaluated different classification algorithms from the Weka toolkit [27] and compared them in terms of recognition accuracy, i.e., the ratio of correctly classified instances and total number of instances. In addition, we calculated precision, recall, and F1 measure for a deeper understanding of the model's performance.

For each classifier we ran a 10-fold cross-validation: the dataset was randomly split into 10 buckets of equal size and then tested 10 times, each time with 9 buckets of training data and one bucket of test data. The results are then averaged to provide a final accuracy. All classification algorithms were run with reasonable default parameters. We did not perform any parameter tuning during this evaluation.

The selected classifiers are popular representatives from different types of classification methods. We considered *J48* (a Weka implementation of the C4.5 decision tree) and *Random Forest* as tree-based methods, *IBk* (a k-Nearest-Neighbor implementation) with k=1 as a lazy learning method, *SMO* (Support Vector Machine) with polynomial kernel, *Multilayer Perceptron* as an Artificial Neural Network, *Naive Bayes* and *Bayes Net* as probabilistic methods, *Logit Boost* with Decision Stump to include a method with logistic regression, and *VFI* (Voting Feature Intervals) [67] as an alternative.

Table 19.: Overview of the classification results.

Classifier	**Accuracy**	**Precision**	**Recall**	**F1 Measure**
J48	91.90%	92.00%	91.90%	91.90%
Random Forest	91.90%	92.00%	92.00%	92.00%
IBk	91.50%	91.70%	91.50%	91.60%
SMO	77.60%	76.80%	77.60%	76.70%
Multilayer Perceptron	86.20%	86.20%	86.30%	86.20%
Naive Bayes	76.10%	75.50%	76.10%	75.70%
Bayes Net	76.80%	76.60%	76.90%	76.70%
Logit Boost	77.70%	77.10%	77.70%	76.00%
Vote	68.60%	47.10%	68.70%	55.90%

Table 19 shows the results. All classifiers yield a higher accuracy than the baseline. Tree-based methods, Nearest Neighbor classification, and Multilayer Perceptron perform best on our dataset. All other classifiers yield mediocre results.

We considered the cardinalization of the place type feature as an a priori probability calculated from the results of the survey. However, it was found to be harmful for classification results except for Naive Bayes, a probabilistic approach. Hence, we neglected the a priori probabilities as a feature.

7.5 DISCUSSION

Detecting of social activity is not a trivial task and can be interpreted differently and investigated to different extends. We focused on the simple differentiation between being *in company* or *alone*. For future studies, it is worth considering to extend this to actual detection of specific individual or group activities – possibly including further sensor such as accelerometer.

While we followed a location-based approach, it might be worth investigating a combination with other methods such as conversation or meeting detection. We only considered common places out of a set that is provided by the Google Places API. That is a restriction that should be loosened for future user studies. We assessed categories for the place types we considered but did not yet explicitly use these categories. Place types that we did not included in our investigations yet might be assigned to categories and assigned the probability of being *in company* of the location category.

Concerning place types, accurate detection might be affected by the accuracy of the GPS signal as mentioned in Chapter 6. For our approach, the accuracy of detecting the current location directly affects the accuracy for estimating being *in company*. Hence, researchers have to ensure that the location is detected correctly.

A factor that influences the generalizability of our results is our sample that was rather small and homogeneous. The results we gained might serve as a basis to infer a basic probability for being *in company*, especially for young technophiles. We recommend to assess a broader set of experiences for being *in company* in combination with user characteristics or demographics so that a fundamental probability base can be established.

Overall, we see potential in our approach as it provides a new way to unobtrusively estimate if a smartphone user might be *in company* without the need to analyze video or audio material.

7.6 SUMMARY

Automatically assessed indicators for being *in company* or *alone* are a desired feature in many areas of social science and computer science. Smartphones, as personal wearables and ubiquitous sensor systems, are a suitable platform for an automatic assessment of this feature. Several researchers investigated how to infer group activity based on sensor measurements such as audio data, video, detected Bluetooth devices, or GPS locations. However, to our best knowledge, we are the first who relied on a generic set of place types which allows the assessment in a more privacy-sensitive and energy-efficient manner.

As a first step towards a location-aware detection system we ran an online survey to assess a basic separability of being *in company* or *alone* based on the place type. We identified that place types with a high frequency of being *in company* usually belong to the *Recreation & amusement* category, e.g., "bar", "movie theatres", or "night club", or belong to the *Food & Drink* category, e.g., "café" or "restaurant". In contrast, users tend to visit place types on their own if they are assigned to the place category *Sports & Exercise*, e.g., visiting the "gym". For some place types such as "university", "park", and "shopping mall" a differentiation is not possible without further information.

These results encouraged us to run an in-field user study to gather real world location data in combination with complementary smartphone features: time and activity. The study lasted three weeks and was taken by 24 participants.

The gathered data consisted of place type, temporal features such as day of week and hour of day, and the user activity. We calculated information gain and χ^2 in combination with Cramér's V to rate the feature importance. Both showed a statistical significance for place types, with a small effect V value of 0.245, and temporal features, with a small effect of hour of day indicated by a V value of 0.149.

Based on these features, we built and evaluated different classifiers using the Weka toolkit. Results of up to 91.9% recognition accuracy are above the baseline of 50% (guessing) or 57% (predicting the most frequent class), respectively. This recognition accuracy is pretty high, but still has room for improvement which is required for the classifier to be applicable in delicate use cases such as social science studies where accurate prediction of social activity is important. Though, it is worth to consider to implement and test an adaptive classifier that runs in the background and only asks for user feedback in case its confidence is below a certain threshold or missing feedback for a specific place type.

Plotting place type against social activity indicator revealed that some place types are reliably separable, e.g., "bus / subway station" and "restaurant". For other place types the distribution seems random, e.g., "university", and requires further information.

The imbalance in the dataset and specifically the sparse data for some places impacted the results negatively. Some places might be very well distinguishable in terms of being *in company* or *alone*, but correlations, for example, with hour of day, were visible in a graphical visualization but could not be confirmed by statistical tests.

Within this chapter, we only investigated generalized models and no personalized models. This is mainly caused by the fact that the location sample from the user study was fairly sparse and we would not have had sufficient samples per place type and person. In addition, an analysis on the full dataset allowed for a generalized interpretation of the results and inference of more general findings. However, in future work, personalized models should be investigated further. The online survey already suggested the existence of either interpersonal differences or external factors that influence the decision of being *in company* or *alone* at some place types such as "university". Hence, further context and sensor sources, e.g., enhanced activity classifiers, calendar information, or device usage statistics, should be considered. Presuming that smartwatches become more widespread, more complex activities could be detected without specialized hardware or laboratory setups. Furthermore, there is potential in recognizing long-term patterns and routines of individual persons, such as regular sport events or working hours.

8 Investigating The Perceived Importance of Smartphone Notifications

Due to the omnipresence of smartphones we are exposed to notifications nearly at all times. Notifications can be useful in delivering information. However, the excess of notifications can also cause information overload [150], interruption overload [151], technostress [196], or digital burnout [126]. With the growing number of installed apps the number of incoming notifications also increases [197] – it "virtually explodes" [149]. If notifications are not handled in accordance to the user's receptivity and interruptibility they might inflict negative effects.

There is a growing need to reduce the number of notifications by filtering unimportant notifications [138] or to improve the delivery process of notifications by predicting opportune moments considering the content of a notification [136] or the context of the user [155]. To improve the receptivity and responsiveness of the user, it is advisable to automatically select a suitable notification modality depending on the notification importance [77, 134]. In achieve this, it is crucial to know which notifications are important to the user, since the importance influences the user reaction towards a notification [170]. Related work investigated this issue, but either under the assumption that the importance of a notification is known or while omitting a clear definition of importance [116, 134, 159]. As future work, Shirazi et al. mentioned the need to seek an automatic assessment of perceived importance [170]. Weber et al. share the same opinion and express a need to consider a larger feature base and different granularity to find indicators for user preferences [195]. We took a first step into that direction and investigated what makes a notification important.

This chapter contains a review on previous research on notification importance, a user's interruptibility and receptivity. This allowed us to apprehend which smartphone features could potentially influence the perceived importance of a notification. These features were considered in an in-field user study to collect real-world data for analysis and interpretation to draw conclusions about the perceived importance of notification recipients.

8.1 RELATED WORK

Not all notifications share the same importance [66], but the importance influences the user's reaction towards a notification [170]. The importance of a notification as perceived by a smartphone user is influenced by the content of the notification (e.g., its interestingness), but also by the context of the user (e.g., if the notification content is relevant or useful for the user at the moment of delivery) [138]. Notification importance is related to the user's interruptibility and receptivity – concepts explained in detail in Chapter 2. Even though there is no existing feature base describing what makes a notification important, it is assumable that the perceived importance relates to smartphone features that influence these related concepts.

We conducted a literature review about features that proved to be related to interruptibility and receptivity. An overview of considered features with an indication of the related work is provided by Table 20. Features can be assigned to one of two categories: notification *content* and user *context*. Content-related features can be derived from the notification or phone settings, e.g., the app sending a notification. Context includes information gathered from internal sensors (e.g., the location of the user) or is provided by the user via ESM self-reports (e.g., the current emotion of the user). In the following, we shortly explain features which are not self-explanatory from their respective name.

NOTIFICATION SENTIMENT The notification sentiment refers to the sentiment of the content of the notification. We classified the sentiment as either positive, negative, or neutral based on keywords [112, 120, 163].

OTHER PARTY For communication apps there is an other party, i.e., the sender or receiver of a message or call. Especially the relationship between smartphone user and other party is of interest [136].

SOCIAL CONTEXT The social context describes whether a user is alone or in company [113, 155]. The presence or absence of a conversation as well as the kind of conversation such as its formality and meaningfulness / emotionality relate to the user's interruptibility [175].

FORMALITY OF AN ACTIVITY We transfer the idea of the formality of a conversation [175] to activities and introduce the *formality of an activity*.

MEANINGFULNESS / EMOTIONALITY OF AN ACTIVITY Similarly to the previous feature, we derived *meaningfulness / emotionality of an activity* from Schulze et al.'s work about conversation types [175].

USER INTEREST Interest can be defined as part of a person's personality relating to individual hobbies and goals [178]. Depending on the notification content in relation to the user interest, the receptivity might be high even if the interruptibility is low [85]. We consider the user interest to generate interest-based notifications, i.e., notifications with content related to the user interest.

Table 20.: Overview of features that were investigated before in related work. Entries marked with * were not considered in our study.

Content-Related Features	Related Work Considering this Feature
App Name	[136, 139, 160]
App Category	[136, 159, 170]
Notification Content*	[85, 138, 139]
Notification Frequency	[41, 159, 170, 175]
Notification Sentiment	-
Other Party	[41, 136, 139, 166, 175, 179]
Notification Urgency*	[41, 139]
Notification Relevance*	[41, 64, 114, 132, 139]
Notification Modality*	[41, 116, 132, 134, 136, 139, 160, 166]
Past User Reaction*	[136, 138, 160, 170]
Phone Attendance	[134, 136]
Internet Connectivity	[136, 139]
Battery Level	-
Context-Related Features	**Related Work Considering this Feature**
Time	[34, 64, 85, 88, 113, 134, 136, 139, 138, 155, 160, 159, 166, 175, 179, 201]
Location	[75, 88, 116, 136, 139, 138, 155, 166, 175, 179, 201]
Activity	[34, 109, 113, 114, 116, 136, 139, 138, 147, 155, 166, 175, 201]
Social Context	[88, 113, 116, 136, 139, 155, 166, 175]
Task Engagement*	[139, 155, 156]
Formality	-
Meaningfulness/Emotionality	-
User Personality	[139, 201]
User Emotion	[155, 175]
Emotion Intensity	-
User Interest	[85]
Smartphone Position*	[88, 132]

EXCLUDED FEATURES There are several features that were investigated by related work, but not considered in our study (marked with a * in Table 20). We did not consider the notification content due to anticipated privacy concerns of our participants. Urgency and relevance of a notification are concepts that require separate investigations on their own and would probably interrelate to other contextual factors that we access such as the location or activity. The notification modality influences the perception of a notification, but smartphone users usually have one default setting that they use [60, 88] resulting in barely any merit for our investigations. The past user reaction was excluded due to interrelations with other contextual factors which might have inhibit a clean investigation of the relation between this feature and the perceived importance. We did not consider the task engagement as it would require a distinct assessment of the current activity and an ESM-based assessment of the engagement with this task. We excluded the smartphone position, since its assessment would require a classification mechanism relying on accelerometer values of which the gathering would drain the battery too much over the course of the study.

8.2 USER STUDY

We designed and conducted a four-week user study to collect a wide range of features (cf. Section 8.1) and to investigate how they influence the perceived importance. After an one-week pilot study, we revised the structure of the study and the questionnaires.

8.2.1 *Course of the Study*

The course of the study included three steps: briefing, daily questionnaires, and debriefing. In the following, we will describe each step in more detail.

BRIEFING In an initial meeting we explained the purpose of the study, asked the participant to sign a consent form, and installed our smartphone app. We shortly explained the questionnaires that the participants would have to answer. Afterwards, we assessed topics of interest and disinterest to be able to send interest-based notifications. We also asked for triggering topics to avoid sending any harmful content.

DAILY QUESTIONNAIRES Within the study, we used an ESM app we wrote to collect the features identified in Table 20 and to capture subjective feedback through a questionnaire, depicted in Table 21.

Table 21.: Daily questionnaire triggered after arrival of selected notifications.

Question	Answer
At what kind of place were you when the notification arrived?	Place Type [98]
At what kind of place were you when the notification was handled?	Place Type [98]
How did you primarily feel shortly before the notification arrived?	Emotion Type [172]
How intense did you feel? (emotion at arrival)	Likert Scale from 1 ("not intense") to 7 ("very intense")
How did you primarily feel after reading the notification?	Emotion Type [172]
How intense did you feel? (emotion after reading)	Likert Scale from 1 ("not intense") to 7 ("very intense")
What kind of activity were you engaged in when the notification arrived?	Activity Type [139, 201]
Were you doing this activity alone?	Alone/Not Alone
How meaningful / emotional was the activity?	Likert Scale from 1 ("not meaningful / emotional") to 7 ("very meaningful / emotional")
How formal was the activity?	Likert Scale from 1 (not formal) to 7 (very formal)
Who sent the notification?	Other Party Type [139]
How interesting is the content?	Likert Scale from 1 ("not interesting") to 7 ("very interesting")
How important is the content?	Likert Scale from 1 ("not important") to 7 ("very important")

Questionnaire prompts were triggered either after the reception of a common or an interest-based notifications. For common notifications, we prompted ten times a day between 6 a.m. and 12 midnight in two hour intervals. The number of prompts might fall below 10 if the participant turned off their phone for a longer period of time. To decide when to trigger within each interval, we considered apps that generated notifications frequently during the past time slot. However, if this app already triggered a high number of prompts in the past, we considered less-frequently sending apps to occasionally include these as well.

To guarantee that participants receive notifications of different interestingness, we distributed interest-based notifications. They were sent out twice per day, again between 6 a.m. and 12 midnight. They were mostly related to topics of interest, but might also include topics of disinterest. The share was about 90% interesting and 10% uninteresting. In comparison to common notifications, interest-based notifications always triggered a questionnaire prompt.

Each prompt was triggered once the related notification was removed: either by tapping the notification, opening the corresponding app, or swiping the notification away. In these cases, our ESM app sent out a notification asking the user to respond to our questionnaire. In order to mitigate the workload on our participants, they did not have to answer the questionnaires immediately.

Overall, participants received up to 12 ESM prompts per day which is in line with related work [162].

DEBRIEFING At the end of the four weeks we conducted a debriefing session. It included a final interview, two questionnaires, and the transfer of the collected data. The interview included questions about the perception of our interest-based prompts, subjective ratings of relations between features and perceived importance and receptivity, respectively, and was concluded by the question "What makes a notification important for you?". In the first questionnaire, we assessed the personality of the participants based on the Eysenck Personality Inventory [81]. The final questionnaire was handed out to assess demographic data. While participants were answering the questionnaires, we exported study-related data from their smartphones. After the debriefing session, we thanked them for their participation and they were free to deinstall the study app.

8.2.2 *Notifications of Interest*

For each topic of interest and disinterest that was mentioned during the briefing session, the study lead collected one web article. To avoid to confront the participants with information that they already know, we selected articles that are not among the top 3 Google results. We created a set of notifications of interest for each participant. The participants were to believe that the app itself was retrieving these web articles based on their interest on a regular basis.

We used a combination of Postman and Firebase to distribute interest-based notifications via HTTP requests that were sent to our smartphone app. The app received a JSON string that contained an anticipated delivery time, the notification title, and the notification content including a link to the web article.

8.2.3 *Participants*

Our participants were mainly recruited in Germany and Hong Kong via mailing lists and social networks. 32 participants took part in our study, 13 of them female, 19 male. The age ranged from 19 to 30 years with an average of 23 years ($\pm$ 3). The participants worked or studied in various fields: STEM, economics, and culture. All participants regularly used smartphones and were familiar with Android 5.0 or above. We rewarded participants with different gifts such as food and drinks during the meetings, but we also raffled three vouchers for local stores worth about $23 (two vouchers) or $32 (one voucher), respectively. Furthermore, we rewarded 15 randomly selected participants with $5.

8.3 RESULTS

As mentioned in Section 8.2, we gathered user feedback via interviews and smartphone data via our app. In the following, we will first present the results of the interview and afterwards analyze the data gathered by our app.

8.3.1 *Interview Feedback*

FEEDBACK ON INTEREST-BASED NOTIFICATIONS 17 participants liked the interest-based notifications. Some participants mentioned that interesting notifications might be acceptable, even though not being important, especially in times or boredom. There were participants who stated that they do not want or need interest-based notifications. Reasons for that refusal are mostly privacy-related: users do not want the system to know their interest. Another reason is the evolving nature of interests and the difficulty to cover them correctly. One participant mentioned that they want to use the device for communication purposes only.

In summary, the receptivity for interest-based notifications is highest in situations of boredom such as idling or resting (11 participants), during leisure time (8 participants) or while in transit (4 participants). Though, some participants also mentioned that they would like to receive such notifications anytime (2 participants), while not in company (1 participant), when they feel happy or curious (1 participant each), or if the notification is interesting to the current situation (1 participant). 3 participants stated that they do not want to receive any interest-based notifications at all, again mostly due to privacy-related concerns.

Table 22.: Influence of Features on the Perceived Importance (1="very low", 7="very high")

Feature	**Mean**
Other Party	5.156 (±1.716)
App	5.125 (±1.293)
Category	5.031 (±1.425)
Activity	4.906 (±1.487)
Interests	4.875 (±1.293)
Formality	4.844 (±1.66)
Time	4.594 (±1.885)
Social Context	4.593 (±1.637)
Meaningfulness / Emotionality	4.5 (±1.561)
Internet Connectivity	4.313 (±1.722)
Location	4.188 (±1.648)
Phone Attendance	3.969 (±1.759)
Emotion Intensity	3.906 (±1.548)
Personality	3.875 (±1.691)
Notif. Sentiment	3.813 (±1.509)
Emotion	3.813 (±1.424)
Battery Level	3.719 (±1.972)

Table 23.: Influence of Features on the User's Receptivity (1="very low", 7="very high")

Feature	**Mean**
Activity	5.125 (±1.495)
Interests	5.125 (±1.409)
Time	5.031 (±1.741)
Formality	4.906 (±1.627)
Social Context	4.781 (±1.727)
Other Party	4.75 (±2.077)
Category	4.719 (±1.7)
App	4.656 (±1.613)
Location	4.438 (±1.749)
Internet Connectivity	4.406 (±1.835)
Meaningfulness / Emotionality	4.313 (±1.379)
Emotion Intensity	4.281 (±1.718)
Emotion	4.156 (±1.481)
Battery Level	4.031 (±2.172)
Notif. Sentiment	3.844 (±1.481)
Phone Attendance	3.688 (±1.685)
Personality	3.563 (±1.952)

SUBJECTIVE RATINGS OF THE INFLUENCE OF SMARTPHONE FEATURES ON THE PERCEIVED IMPORTANCE AND RECEPTIVITY We asked our participants to rate the influence of different features on their perceived importance and receptivity on Likert scales from 1 ("very low") to 7 ("very high"). The results are depicted in Table 22 and 23, respectively.

It is visible that the influence on the perceived importance and the receptivity share some similarities. Overall, each feature has at least minor influence on both, as shown by values above 3.5. When we compare Table 22 and 23, we notice that the perceived importance is rather influenced by content-related features while the receptivity depends more strongly on the context of the user.

Features that influence the perceived importance the most are *other party*, *app*, and *app category*. This confirms findings from related work who found that communication apps and messages from close contacts are considered more important [170]. The perceived importance is further influenced by the *interestingness* and *formality* of the notification. Though, some contextual factors such as the current activity or social context also influence the perceived importance – suggesting a relation between interruptibility and receptivity to the perceived importance. Again, this confirms findings from related work that users are rather interruptible and receptive if they are alone or the task engagement is low [155].

Context-related features that influence the receptivity the most are *activity*, *user interests*, and *time*. It is reasonable that context-related features have a stronger influence on the receptivity. Context relates to the current situation of the user which is changing over time and which influences a user's willingness to be interrupted by a notification as found by related work [75, 77, 155]. Even the two content-related features *battery level* and *internet connectivity* change over time and, possibly, depending on context-related features such as the location.

PERCEIVED INFLUENCE OF FEATURES ON THE NOTIFICATION IMPORTANCE For each content and context-related feature there were at least some participants who believed that these could influence their perceived importance of a notification. The interview results were mainly in line with the results of our final questionnaire depicted in Table 22. All participants acknowledged the *other party* as an influential factor. Even though the *battery level* was rated to have a low impact on the perceived importance (cf. Table 22), 22 participants believed it to influence their perceived importance if it reaches a critical threshold of 5% to 20% of remaining battery. It is possible that participants intuitively relate receptivity to perceived importance in this case: if the battery level is low, a participant thinks twice before turning the screen on to check for an incoming notification. This might be related to an anticipated importance of a notification that influences the willingness to tend to the smartphone and attend the notification. A relation between perceived importance, anticipated importance and receptivity should be investigated further in future work.

DIFFERENT TYPES OF IMPORTANCE Based on the qualitative feedback we identified four types of importance: subjective, objective, public, and situational importance.

> **Subjective Importance** *Subjective importance* refers to notifications which a user would personally rate as important. However, a third party might not declare the notification important. The content of the notification might be of trivial nature but it is valuable to the user due to emotional or personal reasons.

Objective Importance A notification which is *objectively important* describes a notification which would be rated important by the user and a third party. For instance, a notification which notifies about a deadline at work or an emergency call from a family member would be regarded as important even by another party.

Public Importance We noticed that some participants put emphasis on notifications which are of *public importance*. These notifications deliver content which might not be related to the user at all. Such notifications could originate from news apps informing about political issues which are of public relevance, even if abroad. They do not necessarily affect the user directly. However, due to empathy some users would acknowledge these issues as important due to their effect on other people's lives.

Situational Importance *Situational importance* is more dynamic and context-related while the other types are more static. One participant gave the example of using Google Maps: notifications of this app are only relevant in situations in which they are outside and either seeking a way from A to B or feeling lost and requiring an information about their current position. Only in specific situations notifications from such apps are perceived as important.

The occurrence of each of these kinds of importance varies among individuals.

SUBJECTIVE DEFINITION OF PERCEIVED IMPORTANCE We asked our participants to define what makes a notification important. The answers were diverse and many participants mentioned multiple aspects. 15 participants mentioned the personal *relevance* of a notification. However, relevance is ambiguous as it can refer to the relevance to the current task (e.g., a notification delivering an information required to complete a task), sentimental relevance (e.g., a message from a close friend), or relevance for work (e.g., a notification informing about an appointment), among others. Relevance is an aspect that cannot be grasped by a smartphone app easily but requires a recognition mechanism and research on its own. We also found relations between smartphone features and perceived importance in the participant responses. Many participants mentioned that the content-related features *other party* (13 participants), *notification content* (8 participants), and *interestingness* (6 participants) influence the perceived importance which aligns with results depicted in Table 22. 5 participants mentioned *urgency* which is, similar to relevance, hard to grasp by the smartphone and also task-related. Relation to the current task was also found to be relevant: several participants reported *relation to work or study* (4 participants) or a relation to the current *activity* (3 participants).

The influencing factors mentioned by our participants were mostly content-related (relevance, other party, notification content, interest, urgency), but also partly context-related (relation to work/study, current activity). It shows that the properties of the notification itself mostly determine its importance for the participants – properties that already proved to be related to receptivity [85]. However, context-related features can also play a role: according to Table 22 the perceived importance relates to the daily routine (time, location, activity) with a focus on the current activity – which refers to *situational importance*.

Overall, we believe that all mentioned features can influence the perceived importance of notifications. There are also user characteristics which we did not take into account such as age, culture, and occupation which can influence the perceived importance as well. Especially the formality of a notification or conventions about smartphone usage while being in company vary among age groups, cultures, and different occupation groups. These factors are worth to be considered in future, large-scale studies.

8.3.2 *Analysis of Smartphone Features*

In the previous subsection, we investigated subjective impressions about relations between smartphone features and the perceived importance. In the following, we examine if such correlations and associations can be found between actual smartphone measurements and reported perceived importance.

DATASET DESCRIPTION Among all collected notifications, 42845 notifications were accepted and 80469 notifications were declined. Our participants handled roughly 127 notifications per day. A total of 3772 questionnaires was answered. 813 of these questionnaires were about dedicated interest-based notifications sent out by our app and 2959 were about common notifications the user received. The number of answered questionnaires varied depending on the participant. Among all 2959 notifications, 53.3% were rated as not important. Also, 51.1% of these 2959 notifications were rated as not interesting. This shows that a large percentage of notifications usually does not require attention from the user.

CORRELATION ANALYSIS We used the Pearson's χ^2 Test of Independence [84] to test categorical features for correlations with the perceived importance. To rate the effect size, we also calculated Cramér's V [61]. For features which did not meet the necessary requirements for Pearson's χ^2 Test of Independence we used the Kruskal-Wallis Test [119] instead. For numerical values, we used Spearman's correlation coefficient [84], also called Spearman's ρ. To avoid an inflation of type I errors, p values were corrected using the Holm-Bonferroni method [110]. Our statistical tests lead us to the results depicted in Table 27, 28, and 29.

Table 24.: Correlation between smartphone features and perceived importance analyzed using Pearson's χ^2 Test of Independence. Statistically significant results are marked: $*p < 0.05$; $**p < 0.01$; $***p < 0.001$

Feature	**Pearson's χ^2**	**Corrected p Value**	**Cramér's V**
Social Context	9.033	0.859	0.049
Phone Attendance	32.555	<0.001***	0.093
Internet Connectivity	33.831	<0.001***	0.095
Notification Sentiment	21.507	0.001**	0.160
Change of Emotion	173.298	<0.001***	0.214

Table 25.: Association between smartphone features and perceived importance analyzed using the Kruskal-Wallis Test. Statistically significant results are marked: $*p < 0.05$; $**p < 0.01$; $***p < 0.001$

Feature	**Kruskal-Wallis Test Statistics H**	**Corrected p Value**
Arrival Time	23.557	0.135
Removal Time	24.340	0.099
Arrival Place Type	115.720	0.001**
Removal Place Type	103.921	0.008**
Activity	88.373	<0.001***
Category	227.379	<0.001***
Other Party	247.494	<0.001***
Emotion	228.492	<0.001***
Reaction Emotion	477.091	<0.001***

Table 26.: Correlation between smartphone features and perceived importance analyzed using Spearman's ρ. Statistically significant results are marked: $*p < 0.05$; $**p < 0.01$; $***p < 0.001$

Feature	**Spearman's ρ**	**Corrected p Value**
Interest	0.587	<0.001***
Formality	0.251	<0.001***
Meaningfulness/Emotionality	0.336	<0.001***
Emotion Intensity	0.273	<0.001***
Reaction Emotion Intensity	0.326	<0.001***
Battery Level (Arrival)	-0.013	1.000
Battery Level (Removal)	-0.031	0.578
Extraversion	0.028	0.078
Neuroticism	-0.069	<0.001***

Table 27.: Correlation between smartphone features and perceived importance analyzed using Pearson's χ^2 Test of Independence. Statistically significant results are marked: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Feature	**Pearsons χ^2**	**Korrigierter p Wert**	**Cramérs V**
Sozialer Kontext	9.033	0.859	0.049
Anwesenheit am Telefon	32.555	<0.001***	0.093
Internetverbindung	33.831	<0.001***	0.095
Benachrichtigungsempfinden	21.507	0.001**	0.160
Emotionswechsel	173.298	<0.001***	0.214

Table 28.: Association between smartphone features and perceived importance analyzed using the Kruskal-Wallis Test. Statistically significant results are marked: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Feature	**Kruskal-Wallis Test Statistik H**	**Korrgierter p Wert**
Ankunftszeitpunkt	23.557	0.135
Löschzeitpunkt	24.340	0.099
Ankunftsort	115.720	0.001**
Löschort	103.921	0.008**
Aktivität	88.373	<0.001***
Kategorie	227.379	<0.001***
Andere Partei	247.494	<0.001***
Emotion	228.492	<0.001***
Reaktionsemotion	477.091	<0.001***

Table 29.: Correlation between smartphone features and perceived importance analyzed using Spearman's ρ. Statistically significant results are marked: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Feature	**Spearmans ρ**	**Korrigierter p Wert**
Interesse	0.587	<0.001***
Formalität	0.251	<0.001***
Bedeutung/Emotionalität	0.336	<0.001***
Emotionsintensität	0.273	<0.001***
Reaktionsemotionintensität	0.326	<0.001***
Battery Level (Ankunft)	-0.013	1.000
Battery Level (Löschen)	-0.031	0.578
Extraversion	0.028	0.078
Neurotizismus	-0.069	<0.001***

In the following, we analyze the results in more detail.

TIME, LOCATION, ACTIVITY, AND SOCIAL CONTEXT According to Table 28, we can see that the importance of a notification is influenced by the location of its arrival and removal as well as the current activity with statistical significance. These features relate to a daily schedule. It can indicate where a person is and what kind of activity they are engaged in – influencing the interruptibility and receptivity of the user, but also the perceived importance of a notification.

In contrast to our expectations, the *social context* did not indicate significant relations (see Table 27). However, this can be explained by the nature of ESM study: many participants reported that they did not answer our questionnaires while they were still in company.

FORMALITY AND MEANINGFULNESS / EMOTIONALITY The *formality* and *meaningfulness / emotionality* of an activity showed correlations to the perceived importance of a notification with statistical significance (see Table 29). According to Cohen [58], *formality* showed to have a small effect size ($\rho > 0.1$) while *meaningfulness / emotionality* showed to have a medium effect size in our data ($\rho > 0.3$).

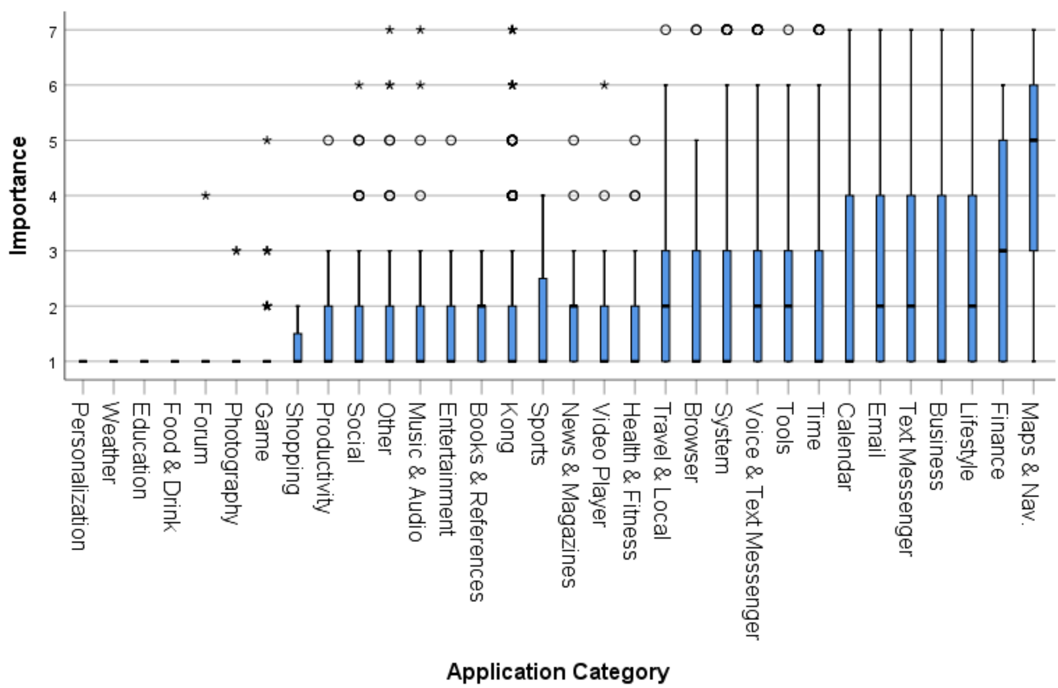


Figure 16.: Importance of notifications from different app categories sorted by their mean values (1 = low, 7 = high).

APP CATEGORY AND OTHER PARTY According to Table 28, there is a statistically significant relation between the app category and the perceived importance. Figure 16 depicts the rating for perceived importance plotted against app categories. It can be seen that apps belonging to certain categories tend to send more important notifications than others – confirming findings from related work [170] and our interviews (see Table 22). Surprisingly, apps which belong to the broader category *Communication* did not outrank the other categories, despite several participants mentioning *Communication* as very important. This could be due to a discrepancy in *objective* and *subjective importance*: objectively, notifications from close friends and family are considered trivial, but they have a personal meaning and thereby a subjective importance. Future studies should investigate these two types of importance in relation to the *other party* in more detail. Apps from the *Maps and Navigation* category were considered most important, probably due to their *situational importance*: users usually utilize such apps while navigating and receive notifications from them in form of navigation instructions. Hence, situational importance should be investigated in more detail as well.

PHONE ATTENDANCE, BATTERY LEVEL, AND INTERNET CONNECTIVITY None of the features *phone attendance*, *battery level* and *internet connectivity* showed a statistically significant effect on the perceived importance (see Table 27 and Table 29). Similar to the *social context* the reason might be that we could not assess the *situational importance*. For example, users probably did not respond to ESM prompts when their phone had a critical battery level, causing us to miss capturing their experience in this exact situation.

USER INTEREST Due to privacy concerns, we were unable to collect the notification content of all notifications (as mentioned in Section 8.1). Hence, the relation between the *individual interest* and perceived importance was not assessable for us. Instead, we let our participants rate their interest towards the notification, i.e., the *situational interest*. Considering the results depicted in Table 29, we notice that *interest* shows statistical significance and a large effect size ($\rho > 0.5$). Since related work found an interrelation between *situational* and *individual interest* [107, 167, 173], we generally recommend *interest* as an influential feature.

NOTIFICATION SENTIMENT Based on the results of the χ^2 test (see Table 27), we can see that the *notification sentiment* relates to the perceived importance with statistical significance and a small effect size according to Cohen [58]. Examining the sentiment, we notice that negative notifications are rated slightly more important. A possible explanation could be that humans rather tend to focus on negative information due to their potential indication of danger [47]. Positive notifications are rated slightly less important than negative ones but still personally important.

Based on another χ^2 test, we found statistically significant relations between the *notification sentiment* and the user's *interest* ($\chi^2(6)$ = 14.996, p = 0.020* and V = 0.134). Thus, positive notifications which are not perceived as highly important might still be of interest to the user.

USER EMOTION, CHANGE OF EMOTION, AND EMOTION INTENSITY

Table 30 shows the absolute distribution of *emotions* at arrival time. We can see that the emotion *Apathy* occurs fairly frequently (43.1%). One possible explanation is the limited amount of emotions from which the participants could choose. 5 participants mentioned that they fell back on *Apathy* as a default if they could not pinpoint their exact emotion to another option.

Table 30.: Number of occurrences of emotions before handling a notification.

Emotion	Frequency
Apathy	1626
Contentment	621
Joy	322
Interest	312
Amusement	209
Pleasure	184
Relief	97
Disappointment	75
Sadness	58
Anger	43
Fear	34
Pride	27
Compassion	27
Love	26
Regret	25
Admiration	21
Disgust	16
Shame	15
Contempt	15
Guilt	12
Hate	7

Table 31.: Number of occurrences of emotions after handling a notification.

Reaction Emotion	Frequency
Apathy	1267
Interest	662
Contentment	466
Joy	276
Amusement	226
Pleasure	123
Anger	122
Relief	111
Disappointment	109
Sadness	90
Love	52
Disgust	38
Admiration	33
Fear	32
Hate	30
Regret	28
Compassion	26
Pride	23
Guilt	22
Shame	18
Contempt	18

Considering the occurrences of *emotions* after reaction to a notification (see Table 31), we notice that the distribution of emotions changed – by 39.2%, overall. The most frequent change was from *Apathy* to *Interest* which occurred 231 times. In our data, the majority of notifications did not cause a change of emotion. Thus, it seems likely that if a notification does have an emotional impact it is because of the notification being important. According to the results gained from the χ^2 test (see Table 27), there is a statistically significant relation between *change of emotion* and the perceived importance of a notification. When analyzing the relationship of change of emotions and *situational interest*, we found statistical significance and a medium effect size ($\chi^2(6) = 398.655$, $p < 0.001^{***}$, V = 0.325). This suggests that interesting notifications are not necessarily important, but they might induce a change of emotion.

Based on Spearman's ρ (see Table 29), the *emotion intensity* shows statistically significant correlation to the perceived importance. When investigating the relation between change of intensity and interest towards a notification, we found a statistically significant correlation and a small effect size ($\rho = 0.152^{*}$, $p < 0.001^{***}$).This suggests that very interesting notifications could be perceived personally important due to personal interest of the user.

USER PERSONALITY In contrast to Mehrotra et al.'s work [139] the personality traits *Neuroticism* and *Extraversion* did not show any effect on the perceived importance in our data sample (see Table 29). In both their study and ours, the sample size was rather small and generalizations are difficult. It is possible that future studies with a larger sample will be able to reveal relations between personality traits and the perceived importance of a notification.

8.3.3 *Classification*

To evaluate if it is possible to predict notification importance based on smartphone features, we trained a generalized RandomForest model. It was not possible to evaluate personalized classifiers as the data for each participant was too sparse to build and evaluate a predictive model. We considered all features gathered within the user study and predicted importance on a scale from 1 to 7. This lead to a recognition accuracy of only 0.615. The classifier misclassified importance values that were close to each other, e.g., assigned an importance of 6 instead of 7. Hence, we examined the performance of the classifier if we consider two categories of notification importance only, i.e., all notifications with an importance value of up to 4 are considered to be of "low importance" and those with a higher value than 4 to be of "high importance". This way, it was easier for the classifier to distinct between the importance classes and the accuracy increased to 0.93. The accuracy is far better than a random prediction (estimated accuracy of 0.50), but only slightly better than returning the most frequent class label (estimated accuracy of 0.92). However, if we consider precision and recall, we notice that the identification of highly important notifications is not as accurate as desired: the prediction only yields a precision of 0.64 and a recall of 0.36. That means that many important notifications are wrongly classified as being not important and that some unimportant notifications are rated to be more important than they actually are. Based on our set of features, it is not yet possible to correctly classify a notifications importance automatically.

8.4 DISCUSSION

Within this chapter, we investigated features that might relate to perceived notification importance. Even though we found reasonable and significant correlations with small to large effect sizes, we were not able to build a reliable classifier to predict a notification's perceived importance. The classification accuracy was high, but precision and recall rather low. These results are caused by the imbalance of the dataset: there is a vast number of unimportant notifications and only a small number of actually important notifications. Moreover, we were not able to build personalized models due to a sparseness of the data per participants so that we had to build a generalized model. Since individuals have different preferences and priorities, a generalized model can only be considered as a basis, but in the end it is necessary to gain personalized models. This means, that either a personalized model has to be trained and used from the beginning or that an adapted model needs to be implemented that is based on a generalized model but learns user preferences over time and, thereby, customizes itself to the user.

The sparseness of the data, as mentioned above, might lie in the nature of our ESM study which allowed participants to answer questionnaires at a later point in time. Thus, when answering the questions, they might have been in a different context and did not assess the importance of a notification in relation to the situation they were in when the notification arrived. A scenario-based interview might help to gain information about user behavior in such situations in which the smartphone is usually not considered, such as being in company or being highly engaged in a task.

The number of answered ESM prompts might also be influenced by the interest-based notifications we introduced. These additional notifications rose the total number of notifications each participant received and might have led to an unintended overflow of the user with notifications. This might have inflicted a decrease in willingness to answer ESM prompts or might have caused a change in emotion, such as an increase in annoyance or anger, which influenced the results.

Due to some technical problems but also due to the length of the study, some participants also tended to answer less questionnaires towards the end of the study. This might not only affect the number of answered prompts but also the range of emotions, activities or locations found in our dataset. A gamification mechanism to keep the participants' motivation and compliance high throughout the course of the study might be a good inclusion in future studies.

A possible limitation of our work is the sample size and homogeneity of our sample. The sample size is larger or similar compared to related studies (e.g., [136, 155]), but still low and does not allow much generalization. Concerning the homogeneity of the sample, we only covered a certain age range and the majority of our participants work or study in scientific or technological fields (again similar to related work such as [136, 155]). Our findings might not be representative for elderly people or smartphone users with a different technical background. Though, the results can be projected to digital natives and technophiles who probably represent the majority of smartphone users or will within the next years.

8.5 SUMMARY

The results of our work show that both content and context-related features influence the perceived importance of a notification. Considering the actual number of important notifications, we notice that only a small amount of all notifications were actually rated as important while more than 50% of all notifications were not considered important at all. A classification system should put emphasis on recognizing the most important notifications.

Based on our findings, influencing factors are foremost: *time*, *location*, *activity*, *formality of an activity*, *meaningfulness/emotionality of an activity*, *user emotion*, *situational interest*, *app category*, *notification sentiment*, and *other party*. Researching the influence of each feature in more detail is a crucial step in future work in order to mitigate the fear of missing out important notifications mentioned by a few users. A classification system used to predict a user's perceived importance needs to be sensitive for the assessment of these features. Since these are, to a large degree, rather personal features, we suggest to either rely on personal classifiers at all or to follow a hybrid approach that starts with a generalized model which adapts to the personal preferences of the user over time. In addition, the classification system should be trained especially to detect important notifications and to clearly differentiate these from unimportant or neutrally important notifications. Due to the rareness of (highly) important notifications, it is worth to consider anomaly detection mechanisms.

Related to the identified features, we also noticed an interplay between perceived importance, interruptibility, and receptivity. Content and context-related features such as app category, other party, location, and social context that proved to be related to interruptibility and receptivity in literature [75, 77, 136, 155] also showed relations to the perceived importance. Qualitative feedback revealed that the relevance and urgency of the notification matter but also its content and interestingness for the user – characteristics of a notification's utility. This suggests interdisciplinary work for future investigations.

In addition to content and context-related features, qualitative feedback revealed that personal circumstances such as *age*, *culture*, and *occupation* may influence the perceived importance of a notification as well. Further user studies involving a larger user sample with representatives from different demographic groups might allow to gain more insights on the effects of personal circumstances on the perceived importance.

Based on study results, we extracted four kinds of importance from qualitative data: *subjective*, *objective*, *public* and *situational importance*. We advocate to consider them in future work in order to find relations between smartphone features and their impact on each kind of importance and the emphasis each user puts on these. A definition of perceived importance can be provided referring to these four kinds. Importance relates to the relevance and usefulness for the user (personal/situational importance), the notification urgency (situational importance), interestingness and other party (subjective/objective/public importance). Generally, content-related features can be used to infer personal, objective and public importance while context-related features yield indicators for situational importance. We recommend to build hybrid classifiers for notification importance with a generic component which considers objective and public importance, but also a personalized component which is sensitive to personal and situational preferences.

Part IV.

Perceptibility of Smartphone Notifications

9 Investigating the Perceptibility of Different Notification Modalities Depending on the Smartphone Position

Smartphones offer different modalities to inform us about incoming notifications: ringtone, vibration, notification LED, but also camera flashlight or smartwatch vibration. However, it is not always possible to notice the arrival of a notification as a vibration might be overheard or the blink of a notification LED might be overseen. The perceptibility depends on the user's context such as the smartphone position as the surrounding of the smartphone might inhibit the sound or visibility. For example, a smartphone within a trouser pocket or backpack might be easier to be overheard due to a dampening effect of the textile surrounding the smartphone. A smartphone being hold in the hand or lying on the table might allow to spot the blink of a notification LED that would be invisible if the smartphone is out of sight in the trouser pocket or in the backpack.

Several researchers found a relation between ringer mode and interruptibility [60, 88], a concept that is related to perceptibility as discussed in Chapter 2. We hypothesize that there is a correlation between the perceptibility of a notification, the notification modality and the smartphone position. We will focus on the notification modalities ringtone, vibration, and LED as they were used in related work [132, 134].

Smartphone positions were also investigated in related work before [56, 92]. Common positions, as identified in Chapter 5, are being held in the hand, lying on a table, being kept in a trouser pocket, or being stored in a backpack [60]. We include these positions into our considerations. We decided to not consider the smartphone being held in the hand as this is usually the case during position transitions or when the smartphone is in use and, potentially, every notification would be perceived directly.

Within this chapter we address the following questions:

- How perceptible are smartphone notifications depending on the notification modality and the smartphone position?
- Which notification modality is most suitable for which smartphone position?
- Which notification modality is most pleasant to the user?

9.1 RELATED WORK

Several researchers advocate that the perception of incoming notifications depends on the smartphone position and notification modality [60]. This seems reasonable: A backpack might reduce the sound of the ringtone while a notification LED can be recognized by the user if the smartphone is currently in use and held in the hand. Smartphone users have different preferences both for storing their device and for the default ringer mode and notification modality. For example, vibration and ringtone have a better perceptibility and should be used for rather important notifications [134]. Related work investigated preferences and effects of notifications as well, but rather with a focus on different locations and devices to display notifications. This includes the delivery of notifications via smartphone, nearby displays, and body-worn wearables to detect the most suitable delivery modality [191]. Other researchers examined the effects of notification delivery via smartwatch and smartphone on the driving behavior [92, 94]. To our best knowledge, the suitability of or preferences for certain notification modalities depending on the smartphone position was not investigated before. To benefit from this information, a smartphone would be required to know its storage position to automatically select a suitable notification modality. As investigated in Chapter 5, recognizing the position is already possible with a fairly high accuracy and can be improved even further with a position transition correction. Hence, we can assume that the smartphone position is known and focus on investigating user preferences for and the suitability of notification modalities depending on such positions.

9.2 USER STUDY

We ran a laboratory experiment to have similar environmental conditions for each participant. This setting allows us to have more control over external interruptions that might affect whether, how, and how fast incoming notifications are perceived. It also leads to a higher internal validity.

9.2.1 *Design Decisions*

We had two variables that were to be manipulated. On the one hand the notification modality: ringtone, vibration, and LED. On the other hand the smartphone positions: on the table, in the trouser front pocket, and in the backpack. We decided to run an experimental study with a mixed design. Every participant experienced all notification modalities, i.e., this experimental condition varied within-subject. The smartphone position was fix for each user, i.e., this experimental condition varied between-subject. This should create the feeling of having their own smartphone situated at a certain position. It is pretty unusual to have three smartphones at different positions or to have to move one smartphone to different positions while performing a task such as watching a movie.

To counteract carry-over effects, we randomized the order of the notification modalities. This resulted in six possible orders. As each of these orders has to be conducted with each of the three positions, we end up with 18 different orders.

9.2.2 *Smartphone app*

We created a smartphone app to be able to change the notification modality without the need for manual smartphone interaction. This also allowed us to send out notifications at fix points in time for each participant.

The app was controlled remotely via HTTP requests. The web interface allowed us to specify the content of the notification, the notification modality and the time of arrival. The app received the HTTP requests and reacted by sending out a notification. The smartphone display was always deactivated so that only the specified notification modality indicated an incoming notification. The app was designed to track the time between sending the notification and the user's reaction to the notification. The collected data is sent back to the server for storage, but also remains on the phone in case of a connection loss.

For the notification modalities we selected the following configuration:

- **Ringtone:** standard sound *"Tejat"* at full volume for approximately 250ms
- **Vibration:** standard haptic cues for 400ms with a cooldown of 300ms in between
- **LED:** blinking for 500ms in green (color code #00*FF*00) with a cooldown of 500ms in between

9.2.3 *Scenario*

The scenario was set to take place in a home environment. The smartphone user was supposed to relax while watching a movie. While watching the movie, several distractions might happen: the doorbell might ring, someone might knock at the door, a smartphone notification might come in.

We kindly asked the user to watch the movie thoroughly so that they are able to recall the storyline afterwards. We also asked them to react to distractions. If the phone is showing a notification the user was asked to tap the notification. To avoid the smartphone to be overheard we selected a movie that does not require sound to be understood. We selected "DUSTIN [9], a short animation movie about a dog and a robot.

9.2.4 *Room Setup*

We conducted the experiment in a neutral university office. We aimed at creating an "at home" feeling by selecting a room with pleasant temperature, a window that allows natural light, and a comfortable seat, among others. In addition, we made sure that the room had a good WiFi connection to ensure a working communication and data transfer between smartphone app and webserver.

9.2.5 *Procedure*

Each participant was invited to the experiment room and asked to sit down and to make themself comfortable. First, we explained the scenario, i.e., watching a movie and remembering the storyline while being exposed to external distractions. Next, we asked the participant to sign a consent form. As soon as the movie started, a timer was started automatically that was responsible for sending notifications to the smartphone at fixed time intervals: after 1, 3, and 5 minutes. The participant was watching the movie and, in parallel, reacted to incoming notifications whenever noticed. Each participant had 120 seconds to react to a notification before it was dismissed and labeled as "not perceived". After the movie had finished, the study lead revealed the true nature of the experiment, i.e., that our objective was to investigate the effects of different notification modalities on the perception of a notification. If the participant still gave their consent, we assessed demographic information and kept the gathered smartphone data. To conclude the experiment, we collected qualitative feedback. Each experiment took about 20 minutes in total.

9.2.6 *Participants*

Each participant joined the experiment voluntarily and without being paid. We acquired 36 participants to meet all 18 orders to allow randomized study conditions. Out of the 36 participants 12 were female. Participants were between 18 and 26 six years old. We focused on participants that are familiar with smartphones usage, i.e., who own a smartphone and use it on a daily basis.

9.3 RESULTS

The objective of this laboratory experiment was to investigate:

- The perceptibility of smartphone notifications depending on the notification modality and smartphone position
- The most suitable notification modality for each smartphone position
- The most pleasant notification modality

Each of these aspects will be discussed in the following.

9.3.1 *Perceptibility*

First of all, we investigated how perceptible smartphone notifications are depending on the notification modality and the smartphone position. We considered the reaction times per participant as an indicator for the perception of a smartphone notification. The reaction times per modality and position are visualized in Table 32.

Vibration and ringtone were noticed very often and fairly quickly. In contrast, LED was often not noticed at all, especially while the phone was out of sight, i.e., in the trouser pocket and the backpack.

To verify if these differences are statistically significant, we ran statistical tests. Since the data was not normally distributed, we chose the parameter-free Friedman test [89] instead of the classical t-test. P values were corrected using the Holm-Bonferroni method to counteract an inflation of type I errors [110].

The results, as presented in Table 33, show that there is a statistically significant difference between the reaction times of *LED and vibration* as well as between *LED and ringtone* for all positions. We could not find any significant differences between *vibration and ringtone*.

These results emphasize the high perceptibility of vibration and ringtone as notification modalities, especially in comparison to the notification LED. We conclude that these notification modalities should be used to inform about incoming notifications of high importance that should be noticed at any rate – e.g., ESM prompts for data labeling.

Table 32.: Reaction times for notifications depending on the notification modality and the smartphone position. Events labeled as "not perceived" were counted as 120s, the time limit before a notification was dismissed.

Notification Modality	Smartphone Position	Reaction Time in s
LED	Table	103.24 (±34.98)
	Trouser Pocket	115.95 (±14.02)
	Backpack	120 (±0)
	Average	113.06 (±22.33)
Vibration	Table	16.03 (±32.77)
	Trouser Pocket	9.08 (±2.40)
	Backpack	39.81 (±48.52)
	Average	21.64 (±35.46)
Ringtone	Table	6.68 (±2.42)
	Trouser Pocket	19.06 (±31.84)
	Backpack	30.01 (±42.20)
	Average	18.58 (±31.20)

Table 33.: Results of the Friedman tests to investigate distinctions between the reaction times for different combinations of notification modality and smartphone position. Statistically significant results are marked: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Notification Modality 1	Notification Modality 2	Smartphone Position	Friedman's χ^2	Corrected p Value
LED	Vibration	Table	11	<0.011*
		Trouser Pocket	12	<0.006**
		Backpack	9	0.032*
		All	48.238	<0.001***
LED	Ringtone	Table	12	0.006**
		Trouser Pocket	11	0.010*
		Backpack	10	0.019*
		All	50.628	<0.001***
Vibration	Ringtone	Table	0.33333	1.000
		Trouser Pocket	1.33333	1.000
		Backpack	0.33333	1.000
		All	28.923	1.000

9.3.2 *Suitability*

Next, we investigated if there is a most suitable notification modality for each considered smartphone position.

The results depicted in Figure 17 show that vibration and ringtone are most prominent. They were noticed most of the time for each position, though most often when the smartphone is close to the user, i.e., in the trouser pocket or on the table. They were the only notification modalities that were perceived while the phone was stored in the backpack, even though not by each participant. LED, however, was barely noticed, even when the smartphone was lying on the table.

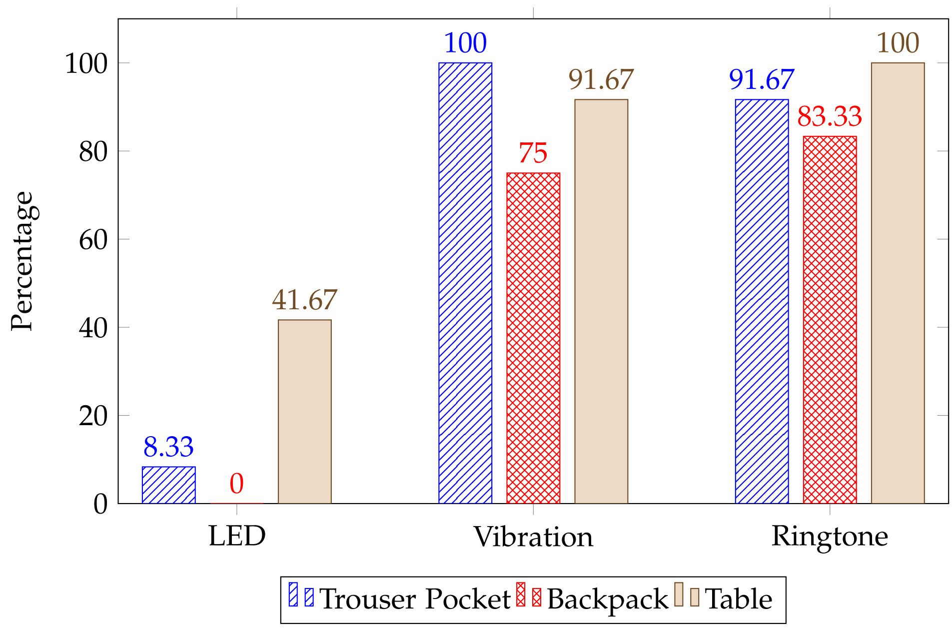


Figure 17.: Overview of percentage of perceived notifications per notification modality and smartphone position.

There are some outliers that probably occurred by coincidence: One participant was able to catch the blink of the LED while the phone was in the trouser pocket. Another one did not notice the vibration of the phone while it was lying on the table. A third participant did not hear the ringtone while the smartphone was in the trouser pocket.

In a qualitative feedback round participants stated that they have a preference of vibration over ringtone or the other way round depending on the context and importance of the notification.

We conclude that both vibration and ringtone are suitable to inform about incoming notifications at all smartphone positions. However, the importance of the notification and the context of the user, e.g., being at home vs. being in a meeting, should be taken into account. LED should only be used for notifications with normal or rather low importance and in contexts where sounds are not welcome, e.g., in a library or during a meeting.

9.3.3 *Pleasantness*

Last of all, we investigated which notification modality is perceived most pleasant. This rating is based on the qualitative feedback of the participants. We asked them about the most pleasant and most unpleasant notification modality.

23 of the participants stated that vibration was perceived as most pleasant. 10 of the participants pledged for ringtone and only 3 for LED as most pleasant. The reasons for their decision can be summarized as one or more of the following aspects:

1. Habit: I use to have the phone in this mode
2. Annoyance, volume and obtrusiveness: it does not annoy or disturb people around me as much as other notification modalities; it is less obtrusive; it does not annoy me all the time
3. High perceptibility, low distraction: I can still follow the movie in parallel; it is not as distracting as other notification modalities; I usually perceive it

We also investigated the least pleasant notification modality. 24 of the participants stated that ringtone was perceived as least pleasant. Each 6 of the participants stated that vibration and LED were least pleasant. The reasons for their decision can be summarized as follows.

1. Disturbance and distraction: it disturbed me; it distracted me; it annoys people in my environment
2. Volume and obtrusiveness: it is too loud; it is too obtrusive; it cannot be ignored easily
3. Privacy and environment: other perceive that I receive notifications

We conclude that it is important to take into account the habits of the user, their (social) context, their current task, and the importance of the notification when automatically selecting a notification modality based on pleasantness. These aspects might be related to each other.

9.4 DISCUSSION

While we did not examine all possible smartphone positions, we focused on the most prominent ones. Our investigations showed that notifications announced by ringtone or vibration are highly receptible rather independently from the smartphone position for a rather silent environment like the one in our laboratory setting. We cannot generalize these results to noisy environments such as trains or areas with construction sites, but use these findings as a first indicator for future investigations.

Another factor that might influence the perceptibility and that was not considered in our study is the hearing capability of the smartphone user. Our sample of participants consisted of digital natives only. It is possible that older persons would show different preferences, e.g., hearing-impaired elderly people.

We recommend to run a field experiment underlying daily noises to verify if the results still hold true. Field experiments might also cover different user activities and might include everyday activities such as riding a bicycle, sitting in the train, attending a meeting, or meeting friends at a restaurant. In addition, we suggest to consider a broader set of smartphone users and with occasional

9.5 SUMMARY

Within a lab experiment, we investigated the perceptibility, suitability, and pleasantness of three different smartphone notification modalities (vibration, ringtone, and LED) depending on three different smartphone positions (on the table, in the front trouser pocket, and in the backpack).

Our results indicate that vibration and ringtone are best perceptible independent from the smartphone position. While vibration is considered most pleasant, ringtone is considered rather annoying, disturbing, and obtrusive. Ringtone should only be used if the notification is very important and requires a quick and direct user interaction. Both notification modalities are suitable for perception of incoming notifications even while the smartphone is stored in the trouser pocket or backpack. LED, in contrast, is only perceptible while the phone is lying on the table. That is, LED is considered very unobtrusive and is therefore suitable for rather normal or unimportant notifications that do not need instant attention from the user.

Qualitative feedback reveals that the habits of the user, their (social) context, and their current task also influence the suitability of the notification modality. These factors also relate to the location and location-based activities: habits describe where we are and what we do, the (social) context influences the kind of activity we are engaged in. Hence, it is a reasonable step to investigate the influence of location and location-based activities next.

10 Investigating the Perception of Different Notification Modalities Depending on the Location-Based Activity

The perception of notifications depends on multiple contextual factors such as the smartphone position, the user's location, and their activity [60]. After considering the smartphone position in Chapter 9, this chapter investigates the perception of notifications and the preferences for different notification modalities depending on the location of the user and the location-based activity. Notification modalities include three alerting methods: auditory cues (e.g., ringtone), haptic cues (e.g., vibration), and visual (e.g., notification LED). Within this chapter, we consider the exemplary named modalities ringtone, vibration, and LED in addition to the silent mode of the smartphone which does not allow any alert. Locations can be described in different formats: as GPS coordinates [69], as Bluetooth or WiFi fingerprints [69, 155], or as place types [155]. Though, semantic interpretation that allows inferences of activities is only available for annotated locations such as place types. Jones et al. argue that (un)common activities are a factor that should be considered when investigating places [115]. Hence, it is reasonable to consider location in combination with location-based activities.

To infer location-based activities, it is recommendable to rely on a location representation with semantic meaning, e.g., place types. The semantic meaning of a place can be provided by the user [155] or inferred automatically from sensor measurements, e.g., using the Google Places API [75]. From the semantic meaning of a place, one or more activities can be inferred that are representative for this location, for example, "watching a movie" at the location "movie theater" or "drinking" at the location "bar" based on common knowledge. Findings from literature emphasize a relation between a user's location and the activity performed at this location [129, 130]. Such activities are also called "location-based activities" – a term which we will use further on. Location-based activities can also be influenced by the social context of a user which can relate to the user location as shown in Chapter 7. Depending on the location-based activity, the corresponding level of engagement, the disruptiveness of a notification, a user's receptivity and a user's interruptibility might vary [75, 155] which also influences the suitability of or preference for a notification modality [60, 77]. To our best knowledge, there is no known research that explicitly investigates preferences of notification modalities depending on the place type and location-based activities. Though, thanks to the omnipresence of smartphones, an automatic and unobtrusive assessment of place types is possible which would allow an automatic selection of suitable notification modalities. We will investigate place type-specific user preferences and relations between notification perception and place type.

Within this chapter we address the following questions:

- How disruptive are smartphone notifications at different place types depending on the smartphone modality?
- How receptive are smartphone users for incoming notifications at different place types depending on the smartphone modality?
- How engaged are users in the location-based activities at different place types?
- Which notification modality is preferred or undesired by users for different place types?

10.1 RELATED WORK

Since our focus is on notification perception, it is reasonable to examine related work in the areas of receptivity and interruptibility detection, especially on mobile phones and in ubiquitous environments. There is a large amount of literature that investigated breakpoints and activity changes [148, 147]. Switching from one task to another is considered a situation with increased user receptivity [86] and a suitable moment for interruptions [34, 109, 156]. These findings suggest relations between the interruptibility of the user and their current activity and task engagement – and possibly also between the user's receptivity for smartphone notifications. Not only the engagement with a task but the nature of the task itself influences the user and their interruptibility. Park et al. concluded that users do not want to be interrupted while being socially engaged [153]. The social context is a property that is related to the location of the user [79] which for itself proved to be a feature related to a user's interruptibility [75]. Social context, activity, and location were considered together in a set of features used for interruptibility detection by Pejovic et al. [155]. Location and activity are not only used as features for interruptibility detection, but also for identifying or predicting user preferences for notification reception [137, 139, 138] or selection of suitable notification modalities [60]. Due to their extensive consideration in literature, we will focus on location and location-based activities and investigate their influence on a user's perception of notifications.

10.2 online survey

We ran an online survey to get a first impression about preferences for notification modalities depending on the place type. The survey results were collected to serve as a basis for inferring hypotheses.

10.2.1 *Survey Design*

Within the survey, we considered the 20 place types that were suggested in Chapter 6 and in [75], respectively.

For each place type, we asked the following questions:

1. How disrupting are smartphone notifications at the following location?
2. How receptive are you for smartphone notifications at the following place?
3. How engaged are you in location-based activities at the following location?
4. Which notification modality do you prefer for the reception of incoming notifications at the following location?
5. What defines a notification modality as suitable at the following location?
6. Which notification modality must not be used to inform you about incoming notifications at the following location?
7. What defines a notification modality as unsuitable at the following location?

Question 1 to 3 were to be answered on a Likert scale ranging from 1 ("not at all") to 5 ("very much"). Question 4 and 6 were multiple choice questions and offered the following notification modalities as response options: silent, vibration, ringtone, LED. Question 5 and 7 were free text questions. In addition, we assessed demographic data.

Originally, the questionnaire was presented in German, the native language of the anticipated participants. It was translated to English for this dissertation to increase its understandability. The survey was created using Google Forms and distributed via university mailing lists. The survey was activated for two weeks.

10.2.2 *Participants*

Overall, 44 people participated in our online survey. Datasets of two participants had to be removed as the participants were under the age of 18 and we did not have proof of their parents' consent. Thus, 42 datasets were considered in the following analysis. These 42 participants were between 21 and 29 years old with an average of 26 years. Out of all participants, 17 were female and 25 were male.

10.2.3 *Survey Results*

The responses to survey question 1 to 3 were summarized in Table 34, representing mean and standard deviation for all three factors disruption, receptivity, and task engagement. The results are ordered in decreasing order of the disruptiveness, i.e., locations on top of the list are those for which notifications are most disruptive. Mostly, these are also those where participants are least receptive and the task engagement is the highest, respectively.

Table 34.: Overview of the average survey responses regarding the disruptiveness of a notification, the receptivity of the participant, and the task engagement of the participant per place type.

Place Type	**Disruptiveness**	**Receptivity**	**Task Engagement**
Movie Theater	4.69 (±0.87)	1.45 (±1.02)	3.83 (±1.5)
Library	4.45 (±1.06)	2.45 (±1.4)	4.14 (±1.14)
Restaurant	3.78 (±1.33)	2.48 (±1.23)	3.55 (±1.13)
Bank	3.57 (±1.31)	2.24 (±1.36)	3.57 (±1.25)
University	3.41 (±1.41)	2.95 (±1.34)	3.93 (±0.92)
Café	3.36 (±1.34)	2.93 (±1.28)	3.05 (±1.01)
Bar	3.12 (±1.27)	2.62 (±1.19)	3.24 (±1.19)
Gym	2.93 (±1.49)	2.17 (±1.4)	3.95 (±1.4)
Night Club	2.55 (±1.5)	2.41 (±1.27)	3.17 (±1.32)
Meal Takeaway	2.38 (±1.23)	3.5 (±1.09)	3.07 (±1.14)
Park	2.29 (±1.42)	3.64 (±1.14)	2.21 (±0.97)
Clothing Store	2.33 (±1.18)	3.43 (±1.17)	3.2 (±0.8)
Post Office	2.05 (±1.06)	3.71 (±1.2)	2.74 (±0.91)
Store	2 (±1.04)	3.6 (±1.21)	2.88 (±0.86)
Shopping Mall	1.88 (±0.99)	3.74 (±1.13)	2.76 (±0.91)
Gas Station	1.64 (±1.06)	3.64 (±1.27)	2.91 (±1.1)
Bakery	1.67 (±0.87)	3.74 (±1.11)	2.62 (±1.13)
Bus / Subway Station	1.69 (±1.1)	4.5 (±0.94)	1.52 (±0.89)
Grocery Store	1.69 (±0.92)	4.02 (±1.02)	3.1 (±1.03)
Parking	1.36 (±0.76)	3.88 (±1.23)	2.24 (±1.23)

The results show that many places with a high probability for notifications being disruptive are also places at which survey participants are less responsive and vice versa. There are places for which a relation between task engagement and notification disruptiveness / receptivity towards smartphone notifications seems to exist, e.g., "library" with a high task engagement and a high notification disruptiveness or "bus / subway station" with a low task engagement but high receptivity, respectively. However, there are also many places with medium task engagement or with a rather high standard deviation indicating that the place alone does not refer to one specific activity with a pre-definable task engagement, but that activities and task engagement can vary – among activities or even among participants.

The responses to survey question 4 and 6 are presented in Table 35 and 36, respectively. Responses to question 5 and 7 are considered in the following when discussing suitable and unsuitable notification modalities.

Table 35.: Survey responses regarding preferred notification modalities per place type. The option with the highest number of picks per place type is printed in bold.

Place Type	**Preferred Notification Modality**				
	Silent	*Vibration*	*Ringtone*	*LED*	*No Preference*
Bakery	4	**25**	12	1	0
Bank	17	**20**	4	1	0
Bar	7	**27**	5	3	0
Bus / Subway Station	4	**24**	13	1	0
Cafe	11	**22**	6	3	0
Clothing Store	5	**25**	11	1	0
Gas Station	5	**19**	17	1	0
Grocery Store	3	**23**	13	3	0
Gym	16	**19**	4	3	0
Library	**21**	16	0	5	0
Meal Takeaway	5	**26**	10	1	0
Movie Theater	**27**	12	1	2	0
Night Club	8	**22**	9	3	0
Park	6	**23**	11	2	0
Parking	4	**20**	16	2	0
Post Office	5	**27**	8	2	0
Restaurant	13	**23**	4	2	0
Shopping Mall	5	**23**	13	1	0
Store	4	**24**	12	2	0
University	16	**23**	0	3	0

Table 36.: Survey responses regarding unsuitable notification modalities per place type. The option with the highest number of picks per place type is printed in bold.

Place Type	**Undesired Notification Modality**				
	Silent	*Vibration*	*Ringtone*	*LED*	*No Preference*
Bakery	8	2	14	2	**16**
Bank	6	3	**25**	1	7
Bar	9	2	**19**	4	8
Bus / Subway Station	8	2	**17**	3	12
Café	8	1	**22**	2	9
Clothing Store	8	2	**18**	3	11
Gas Station	9	2	12	2	**17**
Grocery Store	9	2	**14**	3	**14**
Gym	7	2	**26**	1	6
Library	6	1	**34**	1	0
Meal Takeaway	7	2	**18**	3	12
Movie Theater	5	2	**30**	4	1
Night Club	7	4	**19**	2	10
Park	8	2	13	3	**16**
Parking	9	2	**14**	3	**14**
Post Office	7	3	14	2	**16**
Restaurant	7	1	**27**	2	5
Shopping Mall	9	2	**15**	4	12
Store	8	2	**16**	3	13
University	6	1	**28**	3	4

In general, rather unobtrusive modalities such as silent or vibration are preferred. Especially at locations such as "library" or "movie theater" the phone is supposed to be as silent as possible as a very obtrusive and loud notification modality might disturb other attendees. The majority of participants argued that the social context and environment is most influencing, other people shall not be disturbed by the notification modality. Besides, the preference depends on the task engagement, disruptiveness of the notification, the frequency of disruptions, and the subjective desired distraction. The location itself also plays a role, especially the environmental noise that keeps others from noticing incoming notifications and the importance of the place. Apart from these factors, participants report to select a notification modality out of habit and usually choose their default modality. In summary, participants to not want to bother others, but still want to be aware of their incoming notifications and calls.

For most participants and place types, rather obtrusive modalities such as ringtone are considered undesired. Ringtone was considered most unsuitable at almost all places except for "bakery", "gas station", "park", "parking", and "post office" for which there is no clear refusal of any notification modality. Qualitative feedback confirms that ringtone was perceived unsuitable due to its disruptive, obtrusive nature. Most participants are highly concerned about their companions or the social environment in general. They do not want others to be disturbed by incoming notifications due to obtrusive notification modalities. Some participants refuse ringtone in general as it is obtrusive and annoying, according to them. Though, some mention its usefulness to inform about important, urgent, or critical notifications. For some places, there is no clear refusal of certain modalities. This might be influenced by further factors such as the frequency of interruptions, the participant's momentary receptivity and interruptibility, the importance of the notification itself, or the current environmental noise, as mentioned by some participants.

The results suggest to focus on vibration and ringtone as notification modalities as one is pleasant while the other is considered to be obtrusive, but very receptive. Silent mode would also be an option worth to be considered, but it seems meaningless to investigate the perception of notifications arriving while the smartphone is in silent mode since users would probably not notice any incoming notification. Hence, we will consider vibration and ringtone when formulating hypotheses to be validated in the follow-up user study.

10.2.4 *Hypotheses*

Based on the survey results we formulated the following hypotheses listed in Table 37. We included "home" and "work" as place types as these were used in literature [79, 155].

H1 and H2 focus on rating the receptivity and perceived disturbance of a smartphone user based on their sentiment towards receiving a notification at a certain location with vibration or ringtone as notification modality. Both hypotheses are based on the assumption that most smartphone users are used to vibration and would not mind it as notification modality in contrast to the rather obtrusive notification modality ringtone. H3 and H4 consider the preference of a notification modality at a specific place. H3 is again based on the assumption that vibration is a default modality and preferred over ringtone. H4 is more specific and refers to "do not disturb" locations that inhibit any disruptions at all and which suggest due to their nature that the silent mode might be preferred by a smartphone user. H1 to H4 are primarily based on the quantitative and qualitative interview feedback of the survey participants and focus on the perception of notifications and preferences for notification modalities depending on the place type.

H5 to H7 refer to correlations among different considered concepts, namely receptivity, disruptiveness, and task engagement. H5 is based on the assumption that a high task engagement relates to a high workload and a user desire to focus on a task and its fulfillment. Any interruption is undesired as it would draw attention and, thereby, the willingness to be interrupted (i.e., the receptivity [48]) is low. H6 is similar to H5, but describes that an interruption, e.g., caused by an incoming notification, would cause a highly unpleasant disruption (i.e., high disruptiveness) as it takes away attention resources from the user against their will. H7 relates to the definitions of receptivity and disruptiveness, i.e., the willingness of a smartphone user to be interrupted (i.e., receptivity) and the negative sentiment induced by an interruption (i.e., disruptiveness). The more willing a smartphone user is to be interrupted, the less they will mind a disruption. H8 describes that these three concepts differ depending on the place type, i.e., for each concept the measurements vary among different place types. This seems natural, as place types introduce different activities with varying levels of task engagement and different social environments. H5 to H8 are based on the quantitative survey responses.

Table 37.: Hypotheses about place types and notification reception concepts inferred from survey responses.

	Hypotheses	**Considered Place Types**
H1	Receiving notifications is considered less unpleasant with *vibration* as modality compared to *ringtone*.	All locations
H2	Being disrupted is considered less unpleasant with *vibration* as modality compared to *ringtone*.	All locations
H3	*Vibration* is the preferred notification modality at these locations.	All locations excluding library, movie theater, restaurant, and work
H4	*Silent mode* is the preferred notification modality at these locations.	Library, movie theater, restaurant, and work
H5	The higher the task engagement the lower the receptivity.	All locations
H6	The higher the task engagement the higher the disruptiveness.	All locations
H7	The higher the receptivity the lower the disruptiveness.	All locations
H8	Receptivity, disruptiveness, and task engagement differ at all locations.	All locations

10.3 USER STUDY

The objective of the user study was to validate the hypotheses inferred from the results of the online survey. We decided to run the study in a laboratory setting to have more control about external interruptions and to increase the internal validity. In case that statistically significant results were yielded in such a controlled environment, it would be possible to go into the field as future work.

10.3.1 *Design Decisions*

As notification modalities, we considered ringtone (standard sound *"Tejat"* at full volume for approximately 250ms) and vibration (pattern of 300ms off, 400ms on, 300ms off, and again 400ms on). We excluded LED as it is not supported by all smartphones and hence not a common notification modality. In addition, we excluded the silent mode as it would not have resulted in any notification reception at all.

For the course of the study, we handed out a smartphone to the participants to ensure that each participant is notified by the exact same ringtone and vibration pattern. However, the participants were free to place the smartphone at a convenient position to make the carriage of the smartphone feel as natural as possible.

The study followed a mixed design: each participant experienced each location (within-subject), but was informed about incoming notifications at a certain location either via ringtone or vibration (between-subject). The notification modality changed at each location, i.e., the first group experienced vibration at place type 1, 3, 5, etc. and ringtone at place type 2, 4, 6, etc. and the second group vice versa. Each participant was interrupted by a smartphone notification at the same place and during the same location-based activity. There was only one interruption per place type to keep the study short and as too many interruptions proved to have a negative impact on the participant [70, 151]. Notifications were triggered by the study lead by sending HTTP requests to a smartphone app. This mechanism was already introduced and used before in Chapter 9.

10.3.2 *Room Setup*

We conducted the experiment in a neutral university office. We selected a room with pleasant temperature, a window that allows natural light, and a comfortable seat, among others. We ensured that there were no distractions in the room so that the participants can focus on the study and its scenarios. In addition, we made sure that the room had a good WiFi connection to ensure a working communication and data transfer between smartphone app and webserver.

10.3.3 *Procedure*

At the beginning of the meeting, we informed the participants about their tasks within the study. We provided an example scenario to increase the participants understanding. We explained that there will be interruptions caused by smartphone notifications. Finally, if the participant had no more questions, we asked them to sign a consent form and handed out the smartphone.

The scenarios of the study were designed to fit into an exemplary daily schedule of a student having a student job in a company. We considered the days Friday to Sunday and let the participant experience common activities, each represented by a scenario. After each scenario, i.e., after each experienced location-based activity and interruption, we ran a short structured interview to assess the participant's sentiment towards the interruption. The questions were asked in German, the native language of the anticipated participants, and can be translated as follows:

- On a scale from 1 to 7 with 1 being "not unpleasant at all" and 7 being "very unpleasant": How unpleasant was the reception of the notification?
- On a scale from 1 to 7 with 1 being "not unpleasant at all" and 7 being "very unpleasant": How unpleasant was it to be interrupted during the location-based activity? Why?
- On a scale from 1 to 7 with 1 being "not engaged at all" and 7 being "highly engaged": How engaged were you in the location-based activity?
- Would you have preferred a different notification modality? (yes/no) Why?
- On a scale from 1 to 7 with 1 being "not at all" and 7 being "very much": How much do you agree that ringtone was not a suitable notification modality for incoming notifications at this place type? Why?

At the end of the study, we assessed demographic data and thanked the participants for their time and effort. Overall, one full study walkthrough took between 40 and 90 minutes.

10.3.4 *Scenarios*

We used scenarios to let the participants imagine to be at a certain location performing a certain activity. Our procedure was similar to the one of Turner et al. [187] who proposed to select scenarios first, to collect data next (including to decide when and how to interrupt) and finally process the data, in their case for prediction and in our case for data analysis. The scenarios were provided in German but any examples in this section will be translated into English for better understanding.

At the beginning, we provided an introduction:

> "In this study, you will visit different locations introduced by the scenarios. We will ask you to explain the activities you perform at these locations just like you would perform them in everyday life. Please use the first-person perspective and explain exactly which activities you perform. Please say at least three sentences about location-based activities for each place type."

In addition, we provided a short example scenario:

> "Scenario: You are at home and want to go to the *swimming pool* to meet friends." Start: You leave your home. Activities: going to the swimming pool by bicycle, being at the swimming pool (looking for friends, applying sun screen, go swimming). End: You leave the swimming pool."

The example shows that each scenario contains a place type printed in italics (swimming pool), an objective (meet friends), a beginning and an end, plus possible activities in brackets of which one will trigger an interruptive notification. For this exemplary scenario we also provided an exemplary response to facilitate the participants understanding of how to answer.

> "I take my bicycle and go to the swimming pool. After arrival, I pay the entrance fee and look for a changing room. I put on my bathing togs. Next, I look for my friends. After finding them, I greet them and place my towel next to them. I sit down and put on some sun screen. After a short chat with my friends, I decide to go swimming. After swimming a few lengths, I leave the pool."

In this example scenario, an interruption might have occurred during the search for the friends, e.g., because they send a message indicating their position. As mentioned before, there is only one interruption per place type, so only one predefined activity within each scenario will trigger a notification.

10.3.5 *Place Types and Location-Based Activities*

The selected places and activities are mostly based on the results of the online survey and the inferred hypotheses (cf. Section 10.2) as well as related work [75]. We decided to merge some places with similar location-based activities to reduce the number of test cases in the following study – as it would only prolong the study and would cause repeated activities. Namely, these place types are "shopping mall" and "clothing store" (considered together as "clothing store", because in the intended scenario the participant will go into the shopping mall to try on an outfit), "grocery store" and "store" (considered together as "grocery store", because both offer articles), and "bar" and "club" (considered together as "bar", because both serve drinks and drinking is the intended location-based activity in our scenario). As already mentioned in the hypotheses section, we included "home" and "work", because they were considered in literature [79, 155]. We selected activities that are, based on common knowledge, representative for a place type. In addition, we considered the probability of activities being performed alone or in groups as investigated in Chapter 7. Table 38 gives an overview of all place types and location-based activities that were included into our scenarios.

Table 38.: Overview of different place types and location-based activities during which study participants were disrupted by a smartphone notification.

Place Type	**Location-Based Activity**
Bakery	Ordering food
Bank	Withdrawing money
Bar	Drinking beverages
Bus / Subway Station	Waiting for the train
Café	Chatting with a friend
Clothing Store	Trying on new clothes
Gas Station	Refueling the car
Grocery Store	Looking for groceries
Gym	Performing weight training
Home	Cooking food
Library	Studying
Meal Takeaway	Ordering food
Movie Theater	Watching a movie
Park	Relaxing on a picnic blanket
Parking	Parking a car
Post Office	Waiting in line
Restaurant	Eating food
University	Attending a lecture
Work	Giving a presentation

10.3.6 *Participants*

Participants were recruited out of the personal and professional circles of the study lead, but also acquired by sending announcements to university mailing lists. Most of the participants were current or former computer science students or working in IT-related fields. 40 people joined our study, 19 of them female and 21 of them male. The participants were aged between 19 and 41 with an average age of 26. Each participant took part in the study voluntarily. Among all participants, we raffled two €10 Amazon gift coupons.

10.4 RESULTS

The objective of this laboratory experiment was to investigate:

- The receptivity of a smartphone user for smartphone notifications depending on the notification modality and the location-based activity
- The disruptiveness of a smartphone notification depending on the notification modality and the location-based activity
- The task engagement of a smartphone user depending on the location-based activity
- The preferred notification modality depending on the location-based activity

Each of these aspects will be discussed in more detail in the following.

10.4.1 *Receptivity*

First of all, we investigated how receptive our participants are for smartphone notifications that disrupt them during their location-based activity. We considered the responses given to the first item of the questionnaire, i.e., about the (un)pleasantness of the reception of smartphone notifications. The responses range from 1 ("not unpleasant at all") to 7 ("very unpleasant"), i.e., a low value represents a high receptivity. Table 39 gives an overview of the interview responses in relation to the applied notification modality. To verify if differences between the notification modalities ringtone and vibration are statistically significant, we ran statistical tests. As the data is not normally distributed, we chose the parameter-free Mann-Whitney U test [133]. To avoid an inflation of type I errors, p values were corrected using the Holm-Bonferroni method [110]. We reported the p value that resulted from each test together with an indicator for statistical significance, if applicable.

It is visible that our participants are more receptible for smartphone notifications at locations where they have to perform rather actionless activities such as waiting ("bank", "bus / subway station", "store", "gas station") or relaxing ("bar", "home"). They are rather not perceptive for smartphone notifications at places where disturbances are not generally tolerated ("library", "movie theater", "work"). It is also visible that for some places the receptivity depends on the notification modality. In general, ringtone is perceived as more unpleasant and many participants are not willing to receive notifications at the "post office", "restaurant", or "university" if the notification arrival is announced by the ringtone instead of vibration.

Table 39.: Responses to the interview questions about the receptivity of a participant at a specific place type and during a specific location-based activity depending on the notification modality. Statistically significant results are marked: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Place Type	Modality	Receptivity	Corrected p Value
Bakery	Vibration	2.45 (±1.5)	0.494
	Ringtone	3.60 (±1.57)	
Bank	Vibration	1.80 (±1.06)	1.000
	Ringtone	2.45 (±1.07)	
Bar	Vibration	1.60 (±1.19)	1.000
	Ringtone	2.65 (±2.08)	
Bus / Subway Station	Vibration	1.25 (±0.72)	0.760
	Ringtone	2.25 (±1.55)	
Café	Vibration	3.40 (±1.64)	0.171
	Ringtone	4.85 (±1.50)	
Clothing Store	Vibration	1.95 (±1.44)	1.000
	Ringtone	2.85 (±1.75)	
Gas Station	Vibration	1.35 (±0.67)	1.000
	Ringtone	1.90 (±1.29)	
Grocery Store	Vibration	1.90 (±1.29)	1.000
	Ringtone	2.95 (±1.79)	
Gym	Vibration	2.10 (±1.14)	<0.001***
	Ringtone	4.60 (±1.85)	
Home	Vibration	1.75 (±1.07)	1.000
	Ringtone	1.40 (±0.68)	
Library	Vibration	4.80 (±1.85)	0.133
	Ringtone	6.30 (±1.03)	
Meal Takeaway	Vibration	2.10 (±1.33)	0.190
	Ringtone	3.75 (±2.05)	
Movie Theater	Vibration	4.75 (±1.97)	0.019*
	Ringtone	6.75 (±0.55)	
Park	Vibration	2.05 (±1.43)	1.000
	Ringtone	3.35 (±2.23)	
Parking	Vibration	3.50 (±1.53)	1.000
	Ringtone	4.00 (±2.00)	
Post Office	Vibration	1.55 (±0.94)	<0.001***
	Ringtone	3.55 (±1.61)	
Restaurant	Vibration	3.35 (±1.81)	0.019*
	Ringtone	5.35 (±1.50)	
University	Vibration	3.15 (±1.66)	<0.001***
	Ringtone	5.90 (±0.85)	
Work	Vibration	5.90 (±1.52)	1.000
	Ringtone	6.80 (±0.52)	

Statistical tests emphasize that the difference between ringtone and vibration is statistically significant for "gym", "movie theater", "post office", "restaurant", and "university". These are mostly places where other people are present who might be disturbed by a ringtone, but would probably not notice or mind the vibration of a smartphone. We did not expect statistically significant differences between ringtone and vibration for "movie theater" as both already have high values indicating the unpleasantness of interruptions. Though, the imagination of an alert by ringtone appears to be much more unpleasant than an alert by vibration at this place type, causing a significantly large difference between the mean response values. Absence of statistical significance for other place types, e.g., "library", might be caused by a similar effect of both vibration and ringtone on the participant independent from the location or the location-based activity. Though, it is also possible that opinions vary too much among participants or that the sample was too small or not representative enough.

10.4.2 *Disruptiveness*

Next, we investigated the disruptiveness of our study participants, i.e., their sentiment towards the interruption caused by the reception of smartphone notifications during their location-based activity. We considered the responses given to the second item of the questionnaire, i.e., about the (un)pleasantness of the interruptions caused by the smartphone notifications. The responses range from 1 ("not unpleasant at all") to 7 ("very unpleasant"), i.e., a low value represents a low disruptiveness. Table 40 provides an overview of the interview responses in relation to the applied notification modality. Again, we ran Mann-Whitney U tests [133] to test for statistical significance. P values were corrected using the Holm-Bonferroni method to counteract an inflation of type I errors [110].

It is visible that our participants perceive interruptions as less disruptive at locations where they have to perform rather actionless activities such as waiting ("bank", "bus / subway station", "post office", "clothing store", "grocery store", "gas station") or relaxing ("bar", "home") – fairly similar to places at which the receptivity is high. They are rather not interruptible at places where disturbances are rather not tolerated ("library", "movie theater", "work"). It is also visible that for some places the disruptiveness depends on the notification modality. In general, ringtone is perceived as more unpleasant, especially at places where participants are not willing to be interrupted by an incoming smartphone notification announced via ringtone, e.g., at the "gym", "restaurant", or "university".

Statistical tests did not reveal any significant difference. A reason might be that our participants feel disrupted by both vibration and ringtone independent from the location or the location-based activity. Though, it is also possible that opinions vary too much among participants or that the sample was too small or not representative enough.

Table 40.: Responses to the interview questions about the disruptiveness of a participant at a specific place type and during a specific location-based activity depending on the notification modality. Statistically significant results are marked: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Place Type	**Modality**	**Receptivity**	**Corrected p Value**
Bakery	Vibration	3.20 (±1.58)	0.494
	Ringtone	4.55 (±1.91)	
Bank	Vibration	1.95 (±1.40)	1.000
	Ringtone	2.40 (±1.82)	
Bar	Vibration	2.00 (±1.41)	1.000
	Ringtone	2.60 (±1.88)	
Bus / Subway Station	Vibration	1.30 (±0.73)	1.000
	Ringtone	1.35 (±1.14)	
Café	Vibration	4.25 (±1.86)	1.000
	Ringtone	4.95 (±1.24)	
Clothing Store	Vibration	2.30 (±1.84)	1.000
	Ringtone	2.80 (±1.74)	
Gas Station	Vibration	1.45 (±0.83)	1.000
	Ringtone	1.50 (±0.83)	
Grocery Store	Vibration	2.35 (±1.50)	1.000
	Ringtone	2.30 (±1.22)	
Gym	Vibration	3.25 (±1.97)	0.285
	Ringtone	4.80 (±2.12)	
Home	Vibration	2.25 (±1.52)	1.000
	Ringtone	1.85 (±1.50)	
Library	Vibration	5.10 (±1.62)	1.000
	Ringtone	5.80 (±1.40)	
Meal Takeaway	Vibration	2.50 (±1.64)	1.000
	Ringtone	3.40 (±1.93)	
Movie Theater	Vibration	5.55 (±1.76)	0.190
	Ringtone	6.65 (±0.67)	
Park	Vibration	2.60 (±1.93)	1.000
	Ringtone	3.65 (±2.28)	
Parking	Vibration	4.30 (±1.98)	1.000
	Ringtone	4.65 (±1.63)	
Post Office	Vibration	1.15 (±0.37)	1.000
	Ringtone	1.90 (±1.33)	
Restaurant	Vibration	4.20 (±1.67)	1.000
	Ringtone	5.05 (±1.64)	
University	Vibration	3.80 (±2.02)	1.000
	Ringtone	4.60 (±1.79)	
Work	Vibration	6.65 (±0.81)	1.000
	Ringtone	6.75 (±0.64)	

10.4.3 *Task Engagement*

Last, we investigated the task engagement of our participants in location-based activities. We considered the responses given to the third item of the questionnaire, i.e., about the engagement with location-aware activities at the moment of the interruption. The responses range from 1 ("not engaged at all") to 7 ("highly engaged") and are summarized in Table 41.

Table 41.: Responses to the interview questions about the task engagement of a participant at a specific place type and during a specific location-based activity.

Place Type	**Task Engagement**
Bakery	6.75 (±0.50)
Bank	4.10 (±1.60)
Bar	3.08 (±1.35)
Bus / Subway Station	3.50 (±1.75)
Cafe	5.88 (±0.97)
Clothing Store	4.75 (±1.57)
Gas Station	5.38 (±1.43)
Grocery Store	1.23 (±0.53)
Gym	3.93 (±1.54)
Home	5.03 (±1.61)
Library	3.98 (±1.33)
Meal Takeaway	2.40 (±1.66)
Movie Theater	5.45 (±1.38)
Park	1.30 (±0.61)
Parking	4.53 (±1.50)
Post Office	3.18 (±1.15)
Restaurant	2.58 (±1.34)
University	4.35 (±1.35)
Work	3.85 (±1.17)

It is visible that our participants show low task engagement during activities such as waiting ("bank", "bus / train station", "post office", "grocery store", "gas station") or relaxing ("bar", "home", "park"). They are rather engaged in activities if they need concentration and focus to perform an activity, e.g., to learn at the "library", watch a movie at the "movie theater", parking a car at a "parking" lot, or do their business at "work". There is no differentiation regarding the notification modality for the task engagement as the notification modality only affects the interaction between a smartphone and its user and does not influence the general engagement with an activity that is not influenced by the smartphone or its notification modality. Hence, we did not run comparative, statistical tests.

10.4.4 *Correlations Among Receptivity, Disruptiveness, and Task Engagement*

Interview responses for the three aspects receptivity, disruptiveness and task engagement are listed in Table 39 to 41. It is important to remember that low values in Table 39 and 40 represent a high receptivity and low disruptiveness, respectively. Based on the similarity of many of the results presented above, we decided to run correlation tests to investigate relations between receptivity and disruptiveness, receptivity and task engagement, and disruptiveness and task engagement, respectively. Considering the nature of these concepts, it seems reasonable that there is a correlation among them. As explained in Chapter 2, they are linked to each other. A relation between these concepts, as assessed by our questionnaire, was investigated and confirmed by Pearson's correlation coefficient [154], as visualized in Table 42. As Table 42 shows, the correlations are statistically significant and, according to Cohen [58], they show large effect sizes (r > 0.5) and medium effect sizes (r > 0.3).

Table 42.: Results of the correlation tests to investigate correlations between receptivity, disruptiveness, and task engagement. Statistically significant results are marked: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Concept 1	**Concept 2**	**Correlation Coefficient r**	**p Value**
Receptivity	Disruptiveness	0.828	<0.001***
Receptivity	Task Engagement	0.494	<0.001***
Disruptiveness	Task Engagement	0.616	<0.001***

These correlations seem reasonable as the engagement with a task influences the sentiment towards a disruption during the task and the willingness to be interrupted. Also, as mentioned before, the willingness to be interrupted and the willingness to receive a disruption are also related as they both describe concepts about the expectation and sentiment towards a disruption. Hence, it is comprehensible that the places at which our participants tend to be more receptive are also those with a lower disruptiveness and a lower task engagement, respectively. The same applies the other way around: participants tend to decline the reception of notifications if the disruptiveness is high and they are highly engaged in their location-based activity.

10.4.5 *Correlations Between Place Type and Perception-Related Concepts*

To investigate if the values for receptivity, disruptiveness, and task engagement vary with statistical significance between different place types, we performed a multivariate analysis of variance (MANOVA). The results of the analysis are depicted in Table 43. The analysis confirms statistical significance differences with $p < 0.01$ and, according to Cohen [58], large effect sizes with a partial $\eta^2 > 0.14$. This means that the task engagement level varies among different location-based activities and that the participants' receptivity and the perceived disruptiveness also vary at different locations.

If we include the notification modality in addition to the place type and perform another MANOVA (see Table 44), we can see that the effect sizes increase even more. This might indicate an effect of the notification modality, leading to the conclusion that both place type and notification modality should be considered when investigating receptivity, disruptiveness, or task engagement.

Table 43.: Results of the MANOVA performed to analyze the differences of values for receptivity, disruptiveness, and task engagement depending on the place type.

Concept	**F Value**	**p Value**	**Partial η^2**
Receptivity	31.413	<0.001***	0.433
Disruptiveness	39.830	<0.001***	0.492
Task Engagement	48.248	<0.001***	0.540

Table 44.: Results of the MANOVA performed to analyze the differences of values for receptivity, disruptiveness, and task engagement depending on the place type and the notification modality.

Concept	**F Value**	**p Value**	**Partial η^2**
Receptivity	23.888	<0.001***	0.550
Disruptiveness	21.327	<0.001***	0.522
Task Engagement	24.256	<0.001***	0.554

10.4.6 *Preferred Notification Modality*

Lastly, we investigated the preference for a notification modality depending on the place type and the location-based activity.

First of all, we examined if our participants would have preferred a different notification modality than the one they were confronted with. Table 45 shows an overview of how many participants would have kept vibration or ringtone or modality and how many participants would have preferred to switch to silent, vibration or ringtone, respectively.

The percentage of participants who received ringtone alerts and who would have preferred a different modality is 76.32%, on average. This already shows that ringtone is usually not a preferred notification modality. Especially at the place types "library", "movie theater", "restaurant", "university", and "work" ringtone is undesired by every participant. Reasons for that are that the ringtone is perceived very obtrusive, interrupting and annoying. It is also considered to be inadequate in public and inappropriate when being in company. In addition, many participants stated to use vibration or silent mode as default and would switch from ringtone to another modality out of habit. Place types at which less than 50% would have preferred a different modality than ringtone are "bank", "gas station", and "home". It seems that most participants would not mind to be disrupted by a ringtone at these places, possibly because the corresponding activities do not require much attention and do not involve others who could feel bothered by the ringtone. In addition, the environment might be noisy or the participant outside, so that the ringtone would not be perceived by others and therefore not be as disturbing as at other places.

For vibration, 49.74% of all participants would have preferred a different notification modality. Even this less obtrusive notification modality seems to be too disruptive at places such as "library", "movie theater", or "work". Reasons for that are that they do not want to be disturbed or interrupted themselves or do not want others to be disturbed at these places. For more than half of the considered place types, less than 50% would have preferred a different modality. Participants mentioned that they preferred the pre-selected vibration modality, because it is unobtrusive and only perceived by themselves and it can be ignored, if necessary. In addition, vibration is a common default setting and preferred out of habit by many participants.

In summary, all participants agreed that they wanted neither ringtone nor vibration but preferred the silent mode for the place types "library", "movie theater", or "work". This seems reasonable, since these locations and the corresponding location-based activities relate to a high task engagement, a low receptivity of the smartphone user, and a high disruptiveness caused by incoming notifications.

Our participants do not want to be disturbed at these places and prefer the silent mode which is also enabled by the suitably-named "do not disturb" function of many smartphones.

Table 45.: Preferences for keeping a notification modality or switching to another based on the place type and the location-based activity.

Place Type	**Keep Vibration**	**Switch from Vibration to**		**Keep Ringtone**	**Switch from Ringtone to**		**Preferred Modality**
		Silent	*Ring-tone*		*Silent*	*Vibra-tion*	
Bakery	10	9	1	2	5	13	Vibration
Bank	16	3	1	14	1	5	Vibration
Bar	18	2	0	6	4	10	Vibration
Bus / Subway Station	16	3	1	9	2	9	Vibration
Café	5	15	0	1	9	10	Silent
Clothing Store	13	4	3	4	7	9	Vibration
Gas Station	15	3	2	14	2	4	Vibration
Grocery Store	15	4	1	10	3	7	Vibration
Gym	7	13	0	3	15	2	Silent
Home	12	4	4	16	2	2	Ringtone
Library	3	17	0	0	18	2	Silent
Meal Takeaway	15	5	0	2	10	8	Vibration
Movie Theater	0	20	0	0	19	1	Silent
Park	11	6	3	7	7	6	Vibration
Parking	5	13	2	1	12	7	Silent
Post	18	1	1	3	4	13	Vibration
Restau-rant	7	13	0	0	9	11	Silent
Univer-sity	8	12	0	0	10	10	Silent
Work	0	20	0	0	19	1	Silent

10.4.7 *Hypotheses Validation*

Based on the online survey, we inferred eight hypotheses that were validated based on results from the laboratory study. H1 and H2 are answered based on the results presented in Subsection 10.4.1 and Subsection 10.4.2, respectively. A hypothesis is confirmed if the differences between receptivity and disruptiveness values for vibration and ringtone as modality showed a statistical significance for the considered place types. H3 and H4 are answered based on the results discussed in Subsection 10.4.6. If the preferred modality per place type was vibration for H3 and silent for H4 then these hypotheses can be considered confirmed. H5 to H7 are answered by the correlation coefficients presented in Subsection 10.4.4. A correlation exists if the correlation is statistically significant and there is at least a small effect size. H8 is answered by the results in Subsection 10.4.5. Differences are confirmed if there is a statistical significance for differences among the place types for each concept, respectively. In the following, we will examine each hypothesis in more detail.

H1: Receiving notifications is considered less unpleasant with vibration as modality compared to ringtone as modality at all locations. This hypothesis could be confirmed for some place types, but not for all. Based on the mean values, vibration is perceived as less unpleasant than ringtone for all place types except home. However, the differences between both notification modalities only show statistical significance for the place types "bakery", "bus / subway station", "café", "gym", "library", "meal takeaway", "movie theater", "post office", "restaurant", and "university". This means that there is an overall tendency that vibration is perceived less unpleasant, but we cannot generalize this for each place type. For home, we found an exception as ringtone is considered useful at this location as it allows to perceive notifications while not disturbing anyone else or while not violating any rules of conduct.

H2: Being disrupted is considered less unpleasant with vibration as modality compared to ringtone as modality at all locations. Similar to H1, the hypothesis H2 could be confirmed for some place types only. Based on the mean values, disruptions caused by haptic alerts are perceived less unpleasant than auditory alerts for all place types except home. Only for a few place types, namely "bakery", "gym", and "movie theater", the differences between the disruptiveness of haptic and auditory alert showed statistical significance. Again, the mean values indicate a tendency, but the missing statistical significance does not allow a generalized interpretation. Similar to H1, the reason why ringtone is perceived less disrupting at home is again that it allows to perceive notifications while not disturbing other people. Our participants seem not to mind being disturbed by ringtone if they are the only ones who are affected by the alert.

H3: Vibration is the preferred notification modality at all locations excluding library, movie theater, restaurant, and work. This hypothesis is, again, only true for some place types and cannot be used to infer generalized findings. In contrast to our assumptions, our participants preferred silent mode instead of vibration at the places "café", "gym", "parking", and "university". Apparently, our participants are very sensitive about avoiding to disturb other people around them and about the location-based primary task such as talking to a friend in a "café" or attending a lecture at the "university". Anyway, it is obvious that ringtone was preferred at no place at all, indicating that at most vibration should be used as a default, but at some places even the silent mode. It might be recommendable to include further information about the notification (e.g., its importance) or the smartphone (e.g., the smartphone position) to select a suitable notification modality.

H4: Silent mode is the preferred notification modality at the locations library, movie theater, restaurant, and work. This hypothesis complements H3. While this hypothesis H4 could be confirmed, our investigations also reveal that there are more places with a user preference for silent mode and not only the four considered by H4. What is obvious for these four place types is that at all of them no one wanted to keep the ringtone modality or switch to it: it is perceived as too obtrusive to be selected or tolerated as a notification modality at these locations. At such places, it might be recommendable to rely on vibration as a fallback option to inform about rather important notifications or ESM prompts, but to avoid ringtone-related alerts at any time.

H5: The higher the task engagement the lower the receptivity. This hypothesis was confirmed by our results. There is a correlation of $r = 0.494$ (medium effect size) between the two concepts with $p < 0.01$, confirming that a high task engagement relates to a low receptivity. This seems natural as a high task engagement is defined by the user being involved in a task. Any reception of a notification would represent a distraction and is therefore not desired if the engagement with the primary task is high.

H6: The higher the task engagement the higher the disruptiveness. This hypothesis was confirmed by our results. There is a correlation of $r = 0.616$ (large effect size) between the two concepts with $p < 0.01$, confirming that a high task engagement relates to a high disruptiveness. Similar to H5, this is reasonable as an interruption caused by the reception of a notification would represent a distraction from the primary task. If this task is challenging or the user highly engaged in it, then distractions inflict a negative sentiment towards the notification and lead to a high disruptiveness of incoming notifications.

H7: The higher the receptivity the lower the disruptiveness. This hypothesis was confirmed by our results. There is a correlation of $r = 0.828$ (large effect size) between the two concepts with $p < 0.01$, confirming that a high receptivity relates to a low disruptiveness. The reason behind is the nature of the concepts: a high receptivity means that a smartphone user is willing to receive notifications, i.e., they are open to interruptions and incoming notifications are perceived less disruptive.

H8: Receptivity, disruptiveness, and task engagement differ at all locations. This hypothesis was confirmed by our results. Based on the MANOVA, the differences between the gathered measurements for receptivity, disruptiveness, and task engagement depending on different place types showed statistical significance with $p < 0.01$ and large effect sizes with $\eta^2 > 0.14$.

To conclude, H1 and H2 showed a tendency indicating that ringtone is not as obtrusive as expected and tolerated at more locations than expected. However, we did not found statistical significance for all places so that these hypotheses could not be confirmed. For H3 and H4, we noticed that ringtone should be avoided as a notification modality, especially at presumably "do not disturb" places where the silent mode is preferred. Vibration seem to be a good default for most locations with exception of these "do not disturb" locations. Hypotheses H5 to H8 could be confirmed with statistical significance and medium to large effect sizes. They confirm that there are interrelations among perception-related concepts and between them and place types.

These results confirm the suitability of vibration and silent mode as a default notification modality, depending on the place type. In addition, our results emphasize correlations among receptivity, disruptiveness, and task engagement and confirm that these concepts differ significantly among place types. These findings should be investigated further when examining methods to automatically select a suitable notification modality.

10.5 DISCUSSION

Our investigations showed that smartphone users have different preferences for notification modalities. These preferences relate to a general habit of the user but also to their current location, activity, and task engagement. However, visited places and related activities might vary among individuals.

On the one hand, personal characteristics and hobbies might also have an influence. Our sample consisted of digital natives only. It is possible that other participants perform alternative activities and that they have different preferences. Older participants such as hearing-impaired elderly people or pensioners might prefer ringtone as notification modality as it is easier to perceive and as they might not be in environments such as "work" where the ringtone can disturb other attendants.

On the other hand, we only considered one exemplary activity per place type. We recommend to run a field experiment with participants undergoing daily activities to assess further location-related activities and to verify if our hypotheses per place type still hold true.

In addition, further aspects such as the perceived importance of a notification should be considered when investigating preferred notification modalities in more detail.

For now, our findings can serve as a basis for an automatic selection of a suitable notification modality that can be enhanced by further information about user activities or customized to personal preferences.

10.6 SUMMARY

We conducted an online survey to gain first insights into notification preferences of smartphone users depending on the notification modality and the location-based activity. Based on the survey results we inferred hypotheses about the concepts receptivity, disruptiveness, and task engagement, all related to the perception of smartphone notifications. Within a lab experiment, we collected data to validate the hypotheses.

Our results indicate that smartphone users are usually receptive for notifications if the notification modality is vibration, independent from the place type. This is mostly caused by the habit of the smartphone user, i.e., vibration as default modality, or due to its compromise between unobtrusiveness and perceptibility. Ringtone as modality is perceived as rather unpleasant for many places types, but also tolerated at places where the environment is noisy, at outdoor locations or if the user is alone. At the places "café", "gym", "library", "movie theater", "parking", "restaurant", "university", and "work" users are rather not receptive for any notifications and prefer the silent mode of the smartphone as notification modality.

Interruptions caused by notifications with vibration modality are usually not considered unpleasant, except for the locations "library", "movie theater", and "work". For the ringtone modality, interruptions are mostly perceived unpleasant except for place types at which the interruption caused by the notification is not too obtrusive or if the primary task has a rather low level of engagement.

We found correlations between the three concepts that are related to the perception of a notification. If a user's receptivity is high then the perceived disruptiveness of a notification is rather low, i.e., if the user is willing to receive a smartphone notification then they are also willing to be disturbed by this notification. If the user is highly engaged in the primary task then they are not receptible for smartphone notifications and consider them disruptive, i.e., if the user is engaged in a task then they are not willing to shift their attention from this task to a smartphone notification or any related interruption.

As shown by the results of a MANOVA analysis, receptivity, disruptiveness, and task engagement of our participants vary with statistical significance depending on the place type. This can be explained with the nature of these locations and activities: place types differ in environmental factors, social context, and location-based activities, among others, while the activities themselves vary in task engagement and, possibly, also social context. These factors influence the willingness of smartphone user's to receive notifications and their preference for different notification modalities. Another influencing factor seems to be the importance of a notification as concluded by Mashhadi et al. [134] and as mentioned by participants in an earlier study of ours presented in Chapter 9. Hence, further subjective factors should be considered when investigating the preferred notification modality and the user's perception of smartphone notifications.

11 Investigating Designs to Highlight Important Notifications

The growth in smartphone usage has its benefits for many users such as keeping them up to date or maintaining social relations. However, smartphone usage can also be a burden as it introduces dangers such as information overflow, overchoice, or digital burnout due to permanent availability. Many users trust their phones to inform them about important events such as calls from the partner, messages from good friends, notification about an email confirming a successful job interview. However, with the growing number of potential smartphone apps, the number of potential notifications also increases. Important notifications might drown in the flood of notifications that a user receives throughout the day. Missing important notifications can lead to frustration. Researchers already investigated methods to filter notifications or postpone them to act interruptibility-aware [155]. These options are a good start, but they might not allow the user to actually spot the important notifications. We propose to use visual highlights to enhance the perception of important notifications. We will investigate and evaluate different designs and their effect on the user.

11.1 Related Work

Several researches investigated solutions to address the information overflow and the large amount of notifications that some users tend to receive. These solutions mostly focus on the user's receptivity and interruptibility and usually delay the delivery of unimportant notifications [155] or filter them [138] (cf. Section 2.3). However, a high amount of delayed notifications might still produce an overfull and unclear notification drawer while filtered notifications might cause fear of missing out. Visual highlighting of important notification within the possibly crowded notification drawer might be an alternative solution.

Related work already investigated notification highlights and adaptions. Yoon et al. propose to categorize notifications depending on their content to allow adaptions per category [200]. Possible categorizations might depend on the importance of the source app. Both options would allow further highlights. The results of Aranda et al. show that smartphone users seek a way to identify the source app and the content of a notification without reading the full content [41]. The authors propose colored codes for different notifications, e.g., colors depending on the importance of a notification or the app that sent the notification. Gomes et al. showed that urgent notifications such as alarm or call work best with adaptions of the whole screen [97]. This idea might be adapted to important smartphone notifications: they could be visualized bigger than other notifications.

Of course, there is also fundamental psychological research that investigated the impact of optical adaptions, e.g., on the visual search and selective visual attention. Wang et al., for example, showed that highlights have a statistically significant positive effect on the required search time [192]. They investigated adaptions such as bold and underlined as well as different font sizes for highlighted text. These ideas are also possible highlights for smartphone notifications, especially a bigger font size or bold text. Gluck et al. [96] examined ten different notification signals and varying levels of notification utility and their effect on the user. They noticed that a signals with a high utility known by the user lead to less annoyance and an increased positive perception of the notification. This idea can be transferred to smartphone notifications as well: if the user is aware of highlights which are related to important notifications, the can adapt their expectations towards incoming notifications. This might lead to more positive experiences and a pleasant perception of smartphone notifications.

To our best knowledge, an actual adaption of these highlighting ideas and an application to smartphone notifications was not yet investigated and evaluated in terms of the notification perception so far. We take a first step into that direction by changing the design of important notifications, evaluating how users perceive such highlightings, and identifying user preferences.

11.2 METHODOLOGY

We needed to clarify how to proceed to create and evaluate suitable notification designs. As a first step, we had to define how to collect ideas for design adaptions, e.g., through interviews with a pilot group or in a participatory design workshop. Next, we had to select a method to evaluate designs created based on these ideas and to infer design preferences of smartphone users. Design adaptions in different contexts were investigated and evaluated in related work before.

To investigate notifications for smart TV, Weber et al. [194] formed two focus groups. Based on their responses, they developed designs and evaluated these in a follow-up online survey. Within a laboratory setting, they finalized their investigations and inferred findings concerning their designs.

Yoon et al. investigated correlations between subjective stress due to notification reception and the notification properties within a messenger app [200]. They started with pilot interviews with four participants to get a first impression. Next, they ran an online survey responded by 95 participants.

We adapted both their methodologies. In a first step, we conducted pilot interviews to get an impression of users' ideas for design adaptions to improve the perceptivity of important smartphone notifications. In a second step, we ran an online survey asking participants to rate different notification designs that resulted from the interviews and findings from related work.

11.3 INTERVIEWS

To gain a broader view on the participants' ideas for alternative notification designs and to allow them to freely sketch improvements and adaptions, we ran semi-structured interviews. Such interviews include a basic set of questions, but they can be answered freely and allow follow-up discussions in every direction [182].

11.3.1 *Course of the Study*

In the first part, we assessed demographic data, the general experience with smartphones, and the number of notifications per day. In the second part, we asked for desired optical adaptions of notification designs. All participants had the option to draw their desired designs on paper using, e.g., colored pens, post-its, or glue dots, or to create their designs in the IntelliJ IDEA editor [15] and, optionally, test the designs directly on a Moto E 2nd Gen device with Android 6.0 that we provided.

11.3.2 *Participants*

Five participants joined the interview sessions, three of them male and two female. They were between 19 and 37 years old with an average age of 24. Each of them used their smartphone on a daily basis. Three participants stated to receive up to 30 notifications per day, one between 90 and 100, and one between 300 and 500 notifications per day.

11.3.3 *Results*

None of our participants knew about Androids possibility to adapt the notification design. This implies that none of them used an app so far that uses this option.

All participants shared wishes about how to improve the notification design for important notifications. The suggestions are summarized in Table 46. On the one hand, participants requested more options to select a notification modality based on the app that sends a notification. On the other hand, participants wished for temporal adaptions such as postponing of unimportant or neutrally important notifications. Overall, our participants wanted more control over their notifications which is in line with findings from related work [158, 134, 200].

Concerning notification importance, our participants desired that important notifications do not drown in a flood of unimportant ones. One solution is to have separate spaces for important, neutral, and unimportant notifications. Alternatively, it is possible to adapt the design: highlight important notifications; hide or gray out unimportant notifications.

The designs that were sketched by our participants are listed in Appendix A.

Table 46.: Overview of suggestions based on participants' responses.

Aspect	**Participant**				
	1	**2**	**3**	**4**	**5**
General Suggestions					
Different LED colors	x				
Batch of important notifications	x				
Time-sensitive muting of notifications	x				
Importance-aware notification modality	x				
Reminder for missed notifications	x	x			
Automatic filtering of unimportant notifications		x		x	
Temporal postponing of unimportant notifications			x		x
Suitable actions when swiping away a notification				x	
Overview of removed notifications				x	
Ringtone / vibration for important notifications while phone is in silent mode				x	
Design Suggestions					
More visible content for important notifications	x				
Different colors depending on the importance			x		
Sorting w.r.t to the notification importance			x		

11.4 ONLINE SURVEY

Within an online survey, we evaluated eleven notification designs which were either based on findings drawn from literature (cf. Section 11.1) or based on the interview responses (cf. Section 11.3). Design adaptions consisted of a bigger font size, colored background, alternate text color or increased amount of visible text. In addition, the designs were based on the sketches created by participant 1 to 5, e.g., a sorting by importance. Each mock-up was created in German as this is the native language of our participants and represents their natural smartphone usage. For each mock-up, important notifications originate from messenger and email app while unimportant notifications originate from gaming apps – similar to findings from related work [170] and our own investigations (cf. Chapter 8).

11.4.1 *Set-up*

Similar to the interviews, the online survey was two-fold. First, we introduced the objective of the survey and checked that participants have smartphone experience. In addition, demographic information was collected. In a second part, participants were confronted with one notification design at a time. For each design, a short text described the special property of the design in contrast to the standard design. The participant was asked to rate the design in terms of "I found it easy to spot the important notification" and "I found it easy to distinguish between the important and unimportant notifications" on a 5 point Likert scale ranging from 1 ("I do not agree") to 5 ("I agree"). Afterwards, the participants were asked for their subjective impression of benefits and drawbacks of the design. In addition, we asked if they would use this design themselves (usage likeliness) and why (not).

After all eleven designs were displayed and rated as described above, we asked the participants to rank the designs. We presented all eleven designs again and the participants sorted them from "best" to "worst". To conclude the survey, we assessed additional comments in a free text field, e.g., suggestions about possible design combinations.

The whole survey was conducted in German, but questions and task were translated into English for this thesis.

11.4.2 *Notification Designs*

In the following, we introduce the designs. The corresponding graphics (Figure 18 to 28) are English equivalents to the German designs presented in the survey.

Design 1 (Figure 18) highlights an important notification using a green background.

In **Design 2** (Figure 19) the text color of the important notification was changed to green. The color selection in both cases is based on the feedback of interview participant 3 (cf. Figure 32b).

For **Design 3** (Figure 20) the font size of the important notification was increased.

In **Design 4** (Figure 21) the important notification contains more text in the preview than the other notifications.

Design 5 (Figure 22) is two-fold: in the normal state the app icon is in the focus; when expanded, all notifications belonging to this app are shown with the important notification being highlighted by font size and contrast. It is based on the suggestion of interview participant 1 (cf. Figure 29).

Design 6 (Figure 23) displays the icons of apps, which were sending notifications, in a carousel. The user might switch between them by swiping and sees all notifications of an app by tapping on the app icon in the front. The design is based on the second suggestion of interview participant 2 (cf. Figure 31).

In **Design 7** (Figure 24a and 24b) the unimportant notification was modified by a gray filter and a minimalist design. It is based on the first design suggestion of participant 2 (cf. Figure 30).

Design 8 (Figure 25) shows the categorization and separated storage of notifications as "important", "neutral", or "unimportant". It is based on the first design suggestion of interview participant 3 (cf. Figure 32a).

In **Design 9** (Figure 26), notifications are categorized and sorted based on their source app. For each app, notifications are sorted in reversed chronological order. It is based on the second design suggestion of interview participant 3 (cf. Figure 32b).

Design 10 (Figure 27) presents notifications on form of a "hub". Tapping an app icon will expand all notifications of this app on the right side while the remainder of the hub is darkened. Each icon shows the number of notification per app. For each app, notifications are sorted in reversed chronological order. This design is based on interview participant 4's design suggestion (cf. Figure 33).

In **Design 11** (Figure 28) all notifications are shown in form of a live ticker that switches from one notification to another every few seconds, browsing through them from left to right. Each notification is bigger than the standard Android notification. This design is based on the suggestion made by interview participant 5 (cf. Figure 34).

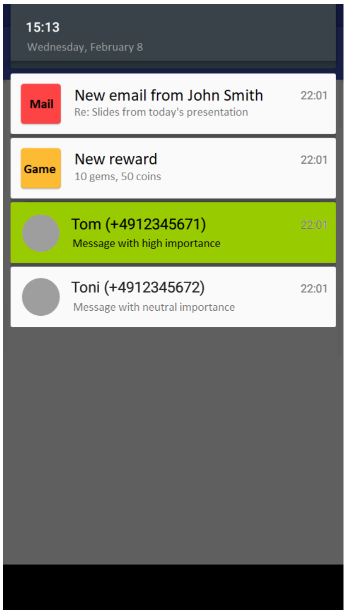


Figure 18.: Design 1: green background for important notifications.

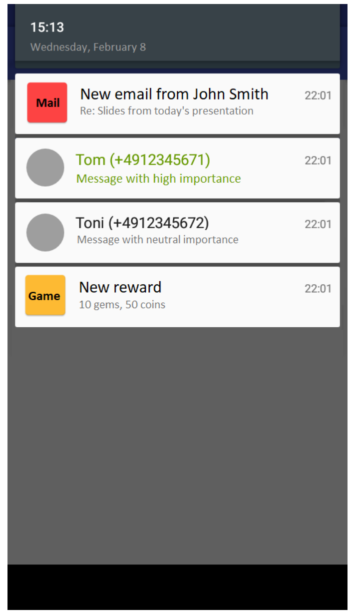


Figure 19.: Design 2: green font size for important notifications.

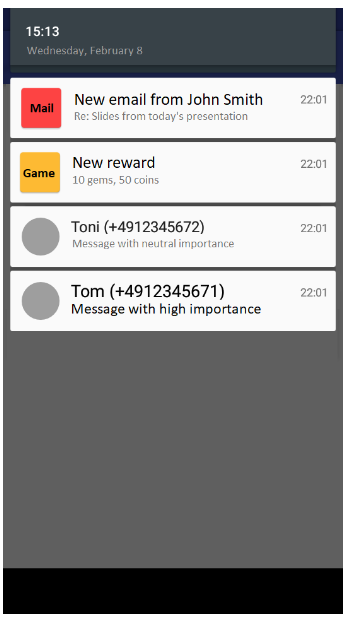


Figure 20.: Design 3: increased font size for important notifications.

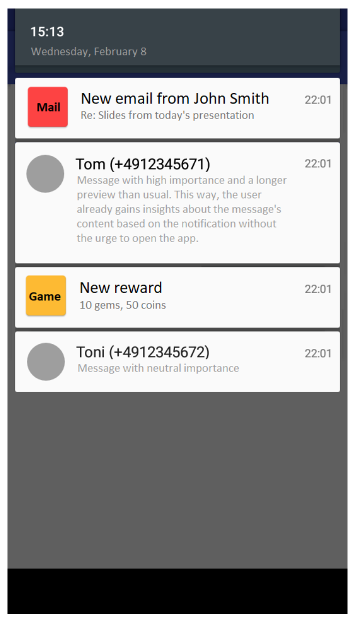
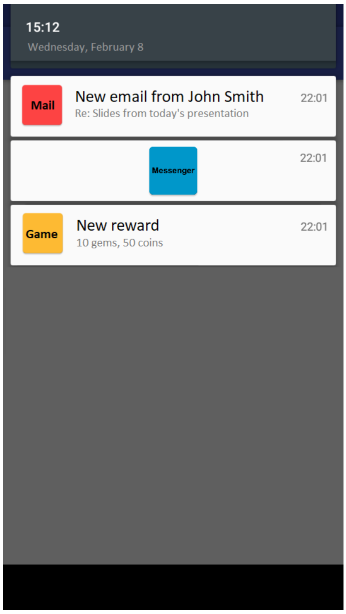
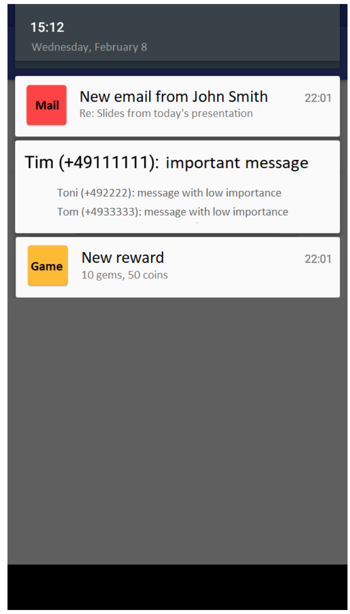


Figure 21.: Design 4: more content is visible for important notifications.



(a) Normal notification view



(b) Expanded notification view

Figure 22.: Design 5: clustering of multiple notifications per app. When extended, the important notifications are indicated by a large font size in contrast to unimportant notifications with small and grey font.

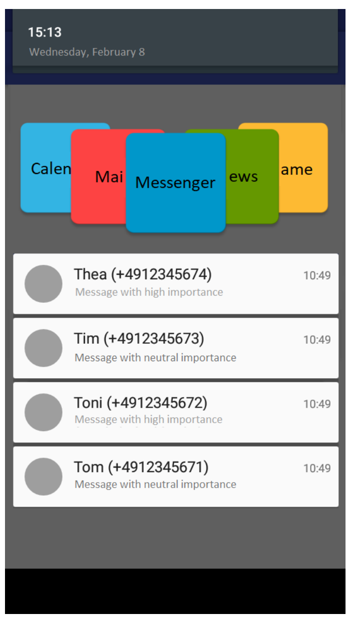
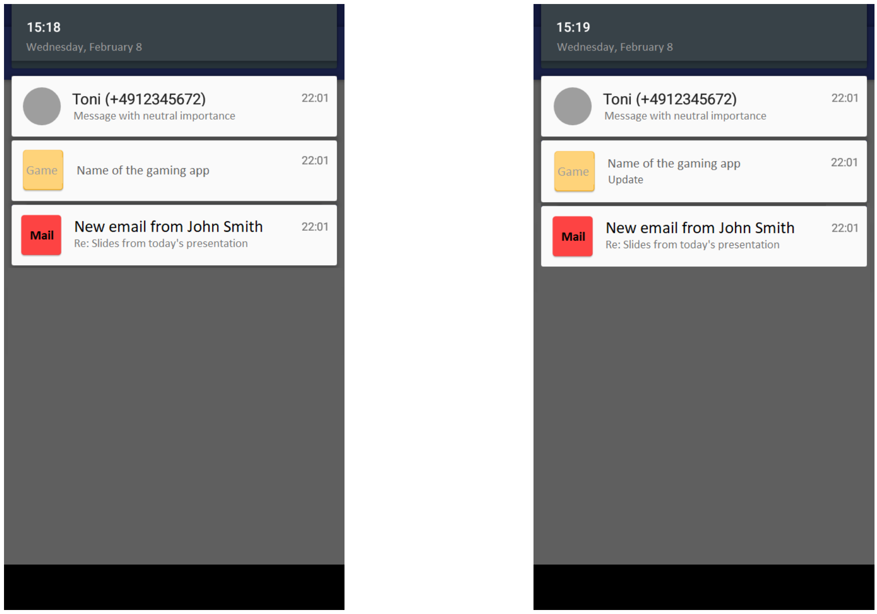


Figure 23.: Design 6: app icons arranged in a carousel. For each app in the foreground, notifications are listed without a pre-defined order.



(a) Normal notification view (b) Expanded notification view

Figure 24.: Design 7: unimportant notifications are greyed out and their content is only revealed after expansion.

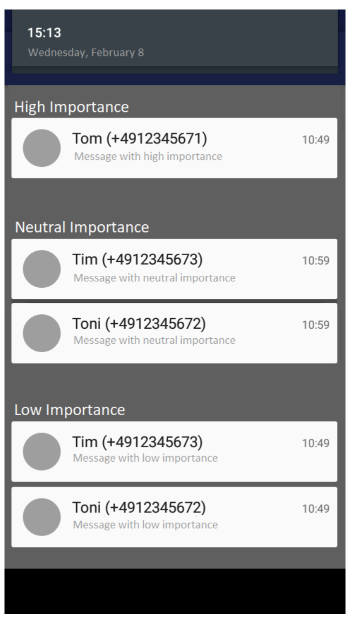


Figure 25.: Design 8: notifications categorized as important, neutral, or unimportant.

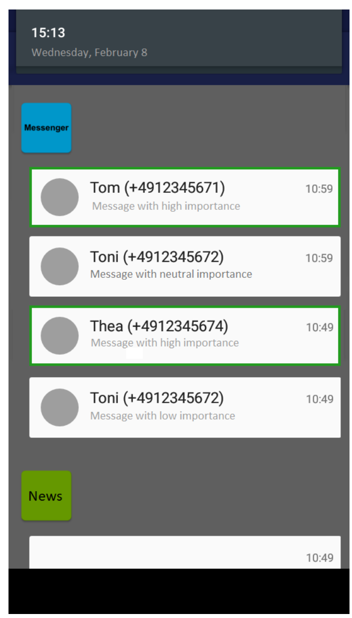


Figure 26.: Design 9: notifications categorized by app. Green frames highlight important notifications.

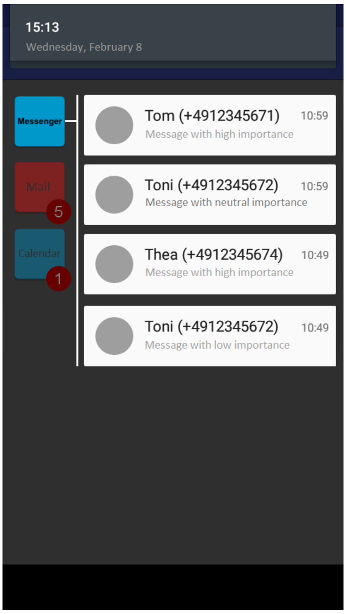


Figure 27.: Design 10: notifications sorted by app and stored in a hub.

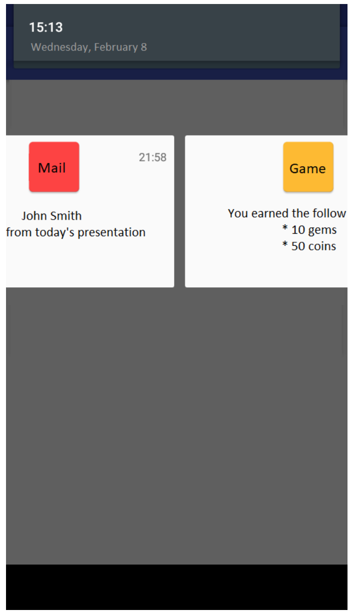


Figure 28.: Design 11: notifications sorted by app in a live ticker style.

11.4.3 *Participants*

Everyone with access to the survey URL was able to participate. We spread the link via social media and mailing lists of the institute. Participants were free to respond to the survey from any location. However, due to the mock-ups, we recommended using a desktop PC.

54 person started the survey. However, only 37 completed it. 18 of them stated to be male and 18 to be female, while one participant preferred not to reveal their gender. The participants were between 19 and 59 years old with a mean age of 29 years. The most frequent background was technical and IT (12), followed by business (9), management (7), education (3), arts (1), law (1), and others (4) such as police or architecture. 35 participants stated to use their smartphone very often (i.e., multiple times a day) and 2 stated to use it often (i.e., once per day). A daily smartphone usage was a requirement that had to be fulfilled so that participants could continue the survey.

11.5 RESULTS

11.5.1 *Ratings*

The mean values for the responses to the questions "I found it easy to spot the important notification" (*easy to find*) and "I found it easy to distinguish between the important and unimportant notifications" (*easy to distinguish*) are visualized in Table 47. In addition, Table 47 lists how many participants would like to use a specific design. It is already visible that some designs are more popular than others. For many designs, easiness-to-find and easiness-to-distinguish seem to be associated with each other and with a participant's desire to use a design.

Table 47.: Ratings for *easy to find* and *easy to distinguish* (1 = "I do not agree", 5 = "I agree") and the number of participants who would like to use a specific design in practice.

Design	Easy to Find	Easy to Distinguish	Number of Participants Who Can Imagine to Use This Design in Practice
Design 1	4.16 (±1.14)	4.05 (±0.97)	13
Design 2	2.97 (±1.14)	2.81 (±1.02)	7
Design 3	1.73 (±0.93)	1.73 (±0.99)	2
Design 4	3.51 (±1.28)	3.38 (±1.28)	20
Design 5	2.77 (±1.17)	2.91 (±1.4)	5
Design 6	2.22 (±1.02)	2.08 (±1.18)	9
Design 7	2.43 (±1.24)	2.22 (±1.23)	3
Design 8	4.62 (±0.95)	4.43 (±1.01)	23
Design 9	3.35 (±0.95)	3.32 (±1.06)	11
Design 10	3.03 (±1.04)	2.84 (±1.12)	13
Design 11	1.52 (±1.00)	1.42 (±0.75)	1

11.5.2 *Influencing Factors*

Based on mentioned benefits and drawbacks of a design as well as reasons to consider using a design in practice or not, we figured out that four aspects influence the design choice the most.

One important aspect is the *color usage*. 16 participants stated that a colored highlighting is beneficial or that it is a drawback of a design if there is no colored highlight of important notifications. However, four participants noted that it is easy to use too many colors and to make a design "too colorful". In addition, the choice of the color is important: for some participants, green is not associated with importance and green and red highlights might be difficult to see for a user suffering from color blindness.

Another important aspect is *space consumption*, mentioned by 11 participants. Many participants are afraid that notifications consume too much space within their notification drawer. They welcome each design that reduces the space consumption.

19 participants also preferred a clear *separation of important and unimportant notifications* or a *visual highlighting of important notifications*. This aspect is mentioned positively if applied to a design and is mentioned negatively if missing.

For many participants the *sorting of the notification* matters. This includes a *sorting by app* (mentioned by nine participants), a *sorting by importance* (mentioned by seven participants), as well as *chronologically inverse sorting* (mentioned explicitly by one participant, but at several occasions also by other participants). If there is no sorting at all, participants criticize this aspect. Sorting is an aspect that was already proven useful by related work, e.g., sorting by content or app [200]. It is important to mention that a sorting by app and importance at the same time is not mutually exclusive, but the order depends on the participant and is very subjective. It is possible to have a categorization by importance and another app overview per category or to have a categorization by app and for each app notifications are ordered by importance – as suggested by several participants. The suitability of this combination as well as the first selection criterion of app or importance might also heavily depend on the number of overall notifications and the number of distinct apps sending notifications.

11.5.3 *Ranking*

At the end of the online survey, we asked the participants to rank the designs from rank 1 ("best") to rank 11 ("worst"). To get a better overview of the ranking results, we chose two analysis methods: frequency count in buckets (first three ranks, middle, and last three ranks) or rating in points, either linear or exponential. Table 48 shows how often each design scored in each of the three buckets. Table 49 shows the results if we assign points to each rank, either linear (10 points for the first rank, 9 for the second, [...], and 0 for the last rank) or exponential (25 points for rank 1, 18 for rank 2, 15 for rank 3, 12 for rank 4, 10 for rank 5, 8 for rank 6, 6 for rank 7, 4 for rank 8, 2 for rank 9, 1 for rank 10 and 0 for rank 11), and add up all points per design.

Table 48.: Frequency of rankings per design if we consider three ranking buckets.

	Design										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Rank 1-3	23	9	0	17	3	11	2	22	11	12	1
Rank 4-8	12	25	23	16	29	17	15	12	18	13	5
Rank 9-11	2	3	14	4	5	9	20	3	8	12	31

Table 49.: Rank and points for each design based on a linear or exponential rating.

Points	**Linear Rating**		**Exponential Rating**	
Rank 1	Design 1	322	Design 8	603
Rank 2	Design 8	316	Design 1	589
Rank 3	Design 4	281	Design 4	450
Rank 4	Design 2	262	Design 2	389
Rank 5	Design 9	236	Design 10	360
Rank 6	Design 10	230	Design 6	355
Rank 7	Design 6	229	Design 9	344
Rank 8	Design 5	195	Design 5	255
Rank 9	Design 3	150	Design 3	163
Rank 10	Design 7	139	Design 7	152
Rank 11	Design 11	82	Design 11	77

Both point-based ratings show that design 1 and 8 are considered the best while design 11 is considered to be the worst one. Some participants reported that two designs rank similar and that it was difficult to decide which one to rank higher. For one participant, only four designs were considered good and the remaining seven bad, which caused a rather random ranking of the worse designs.

11.5.4 *Correlation Analysis*

Based on the results depicted in Table 47 and 49, we assumed associations between design features. To see if there are statistically significant correlations, we ran different analyses.

We calculated the correlation coefficient for different design features, namely:

- Important notifications are easy to find (*easy to find*)
- Important notifications are easy to distinguish from other notifications (*easy to distinguish*)
- Likeliness to use a design in practice (*usage likeliness*)
- Ranking position (*linear ranking position* and *exponential ranking position*)

At first, we calculated Pearson's correlation coefficients to investigate correlations between *easy to find* and *easy to distinguish*, visualized in Table 50. To avoid an inflation of type I errors, p values were corrected using the Holm-Bonferroni method [110]. All designs show statistically significant correlations. This means, that when it is easy for participants to spot an important notification than it is usually also easy to distinguish between important notifications and other notifications.

Table 50.: Overview of Pearson's correlation coefficients calculated for *easy to find* and *easy to distinguish*. Statistically significant results are marked: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Design	Correlation Coefficient r	Corrected p Value
Design 1	0.467794	0.039*
Design 2	0.7561713	<0.001***
Design 3	0.8212881	<0.001***
Design 4	0.9124752	<0.001***
Design 5	0.7622509	<0.001***
Design 6	0.8646302	<0.001***
Design 7	0.8878148	<0.001***
Design 8	0.9204813	<0.001***
Design 9	0.8812513	<0.001***
Design 10	0.8155056	<0.001***
Design 11	0.9443302	<0.001***

Next, we calculated Pearson's correlation coefficients to investigate correlations between *easy to find* and the *usage likeliness*, visualized in Table 51. P values were corrected using the Holm-Bonferroni method to counteract an inflation of type I errors [110]. Statistically significant results were only gained for those comparisons that achieved a high linear correlation (i.e., design 2, 4, and 9). All other designs did not show any statistically significant correlation between *easy to find* and the *usage likeliness*. This means that the easiness to find important notifications does not necessarily mean that this will cause a design to be used in practice – there is still the component of subjective preference which might be influenced by other aspects as well.

Finally, we averaged the response values for each question and each design and calculated Pearson's correlation coefficients. P values were, again, corrected using the Holm-Bonferroni method [110]. The results are summarized in Table 52 and show statistically significance for each combination. That means, on average, that the easiness to spot an important notification leads to a higher position in the ranking and a higher usage likeliness. This is not fully identical with the findings from the previous two analyses. This is due to the averaging: Pearson's correlation coefficient seeks linear correlations with might not be exactly linear due to scattered user responses as shown by the standard deviation. If we only consider the average values, like in this case, it might be more likely to actually approach linear correlation.

Table 51.: Overview of Pearson's correlation coefficients calculated for *easy to find* and *usage likeliness*. Statistically significant results are marked: $*p < 0.05$; $**p < 0.01$; $***p < 0.001$

Design	Correlation Coefficient r	Corrected p Value
Design 1	-0.2456556	1.000
Design 2	-0.5627874	0.003**
Design 3	-0.3301601	0.506
Design 4	-0.5027244	0.017*
Design 5	-0.4366889	0.096
Design 6	-0.2812424	1.000
Design 7	-0.381634	0.217
Design 8	-0.1602062	1.000
Design 9	-0.5137985	0.013*
Design 10	-0.4219165	0.102
Design 11	-0.086711	1.000

Table 52.: Overview of Pearson's correlation coefficients calculated for correlations between two features each with mean values per feature. Statistically significant results are marked: $*p < 0.05$; $**p < 0.01$; $***p < 0.001$

Feature 1	Feature 2	Correlation Coefficient r	Corrected p Value
Easy to find	Linear ranking	-0,9111982	<0.001***
Easy to find	Exponential ranking	-0,8799084	0.002**
Easy to find	Usage likeliness	0,8663965	0.003**
Linear ranking	Usage likeliness	-0,8393291	0.006**
Exponential ranking	Usage likeliness	-0,8812956	0.002**

Based on the results, we infer that "easy to find" is no guarantee for the later usage of a design in practice, even though it increases the probability. If it is easy to find an important notification due to the notification design than it is usually also easy to differentiate between important and unimportant notification.

11.5.5 *Suggestions for Design Combinations*

Each participant was free to add further comments in a free text field at the end of the study, for example, to propose optimizations for designs, combinations of designs, or own designs. An overview of all requested combinations, represented by their essential aspects, are visualized in Table 53. Table 54 lists the frequency of how often a combination of two aspects was requested.

Table 53.: Overview of desired design combinations based on comments in the free text field at the end of the survey.

Aspect	**Combinations**										**Frequency**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	
Colored highlighting	x		x	x	x	x	x	x			7
Expanded preview	x			x	x			x	x	x	6
Sorting by importance			x	x		x	x		x		5
Categorization by app		x	x		x				x		4
Number of unread notifications		x						x			2
Chronological sorting		x									1
Hub	x										1
General highlight		x									1

11.6 DISCUSSION

Our results are a good basis for future investigations, but they have to be treated with care. First of all, we ran an online survey only. This was useful to gain first insights into user preferences and desired properties. As a next step, it is necessary to actually implement the designs that we mocked-up and to evaluate them first in a laboratory setting and later in-field.

Table 54.: Overview of the frequency of named desired combinations.

	Expanded preview	Sorting by importance	Categorization by app	Number of unread notifications	Chronological sorting	Hub	General highlight
Colored highlighting	4	4	2	1	0	1	0
Expanded preview		2	2	1	0	1	0
Sorting by importance			2	0	0	0	0
Categorization by app				1	1	0	1
Number of unread notifications					1	0	1
Chronological sorting						0	1
Hub							0

In addition, both our interview as well as our survey participants were young people with smartphone experience. They probably represent the majority of smartphone users and, therefore, were a good fit to gain first insights. We assume that design preferences differ for smartphone users who receive less notifications or who use their phone less frequent. The same applies to elderly people or people with defective vision who possibly prefer other kinds of visualizations.

Within this chapter, we focused on adaptions of the notification design. We recommend to investigate adaptions of the notification modality in combination with the notification design. In addition, other approaches such as an importance classifier or an interruptibility-aware system are perfect complements to our designs and should be implemented and evaluated together.

So far, we restricted our investigations to smartphone notifications. It is also possible to expand the notification delivery to wearables or nearby displays and apply highlighting methods for important notifications for such presentation methods as well. We suggest to examine if the same preferences as identified by us for smartphone displays are suitable for notification presentation via these alternative interfaces as well.

11.7 SUMMARY

We investigated different design adaptions to facilitate finding important notifications. In an interview session with five participants, we let the participants sketch a few designs for a better notification presentation. Based on findings from literature and the interview session, we created 11 designs which were rated by 37 participants in an online survey.

Our analyses revealed statistically significant correlations between the design features "important notifications are easy to find" and "important notifications are easy to distinguish from other notifications". In addition, the probability of participants using a design is positively correlated with the easiness to find important notifications.

Ratings of the eleven designs show that there is no overall perfect design and that each user has their own subjective preference. However, there are two designs that were ranked the highest: design 1 (green background to highlight an important notification) and design 8 (categorized notifications by importance). Both favorite designs own properties that were positively mentioned several times such as: colored highlight; low space consumption; separation of important and unimportant notifications; sorting by importance or app.

Based on the subjective preference of each user, we suggest that mobile operating systems implement different options to adapt the notification design to make this process most pleasant for its users. Design-specific properties are a good start, e.g., a personal selection of a color for a colored background or frame.

In future work, different designs should be implemented and tested in a laboratory study (similar to [194]) or later in a field experiment. Related work already investigated ways to postpone unimportant notifications [155] or to select different notification modalities [77]. It is recommendable to investigate combinations of different strategies to manage notifications. In the end, single strategies or combinations of them should be implemented and offered to the smartphone users as options to choose from.

Part V.

Conclusion

12 Discussion

In this chapter, we discuss different aspects investigated in this dissertation in a broader context. We review assumptions we made and limitations we were facing, discuss the generalization of results, and consider ESM-related issues.

12.1 Assumptions and Limitations

During our investigations, we made some assumptions – some implicitly, some explicitly. For example, when investigating ways to assess factors of interest, we collected labels out of a pre-defined pools of labels. This holds true for the position detection as well as the location detection. It is possible that the device will be stored at an unknown position or that the user visits a location that is not covered by the 20 selected place types, respectively. Moreover, we did not consider moments between position state changes, e.g., if a hand is wrapped around a smartphone while the device is still in the backpack, or if a smartphone user stays at locations that belong to multiple place types or that might overlap, e.g., "library" and "university" for a university-internal library. One solution might be to apply weighted approaches that compute a score based on the values assigned to multiple positions or locations that apply to the current situation. For the detection of being in company, for example, this is possible by multiplying the a priori probability for being in company of each place type among the top 5 possible places identified by the Google Places API with the probability of being the current location that the API assigns to each of these place types.

Considering the identification of important notifications, there might be features related to the perceived importance that would yield a higher predictive power but which we did not consider. In addition, not only the interest of a user but also the relation to the other party might underlie fluctuations: an acquaintanceship might become a partner, a colleague might become a superior, or a friendship might end. It might be difficult for a classification system to cover such cases. Still, our findings are a first basis that indicate future research directions. An important finding are the four kinds of importance we identified – they might be investigated in relation to considered smartphone features separately in more detail and in terms of their predictive power.

When focusing on the perceptibility of notifications and the effect of different notification modalities on the user, we could confirm different preferences depending on the smartphone position and location. However, not every possible smartphone position or user location was included in our user studies, neither are combinations. We did not consider what might happen if conflicts arise, e.g., someone is at home but having a formal meeting with their superior. Not every situation can be covered by an automatically acting system like a smartphone. We provide recommendations for common situations and propose to use adaptive classifiers that ask for feedback in case of uncertainty to learn how to behave in uncommon situations.

Another aspect is that we focused on Android as mobile OS. However, there are other platforms such as iOS or Windows phone. We cannot argue about the suitability of designs or modalities for these users. There is a chance that they have varying preferences due to the habit of using the non-Android OS. Even though it is possible that there are desirable design adaptions not yet supported by Android that we missed, our results are a good complement for the upcoming Android 8.0 version and its offered functionality to customize notifications. App designers as well as the Android developers themselves can benefit from our findings and start to change the design of notifications.

12.2 GENERALIZABILITY

For all user studies conducted on the context of this dissertation, the sample size was rather small and homogeneous, especially in terms of age and occupation. Although both are comparable to samples reported in related work, this implies that the results have to be treated with care. All findings, even if they showed statistical significance and small to large effect sizes, primarily apply to the considered kind of smartphone users.

Most of our participants were young and most of them were students of a STEM subject. Assumably, they have certain technical affinity. We cannot transfer our findings to persons without any technical knowledge or experience at all. In addition, we could not consider every type of smartphone user. There was a variance in the number of received notifications per day, but a higher absolute number does not mean that the user suffers from negative effects – there are subjective differences. When evaluating notification designs, some users mentioned to be afraid of too many colored notifications. However, what is the expected number of important notifications? Is there a chance that someone would receive too many important notifications and be overwhelmed by the highlighting? We cannot answer these questions directly, but they should be investigated further.

The proliferation of smartphones suggest that in a few years everyone will own such a device and will be a more or less experienced smartphone user. It is possible that the findings we yielded for rather young smartphone users nowadays will hold true for a much wider range of persons in the future.

Another aspect is that most of our studies were conducted as online surveys or lab experiments which grant a higher internal validity, but a low external validity. It is not yet known if the results are replicable in the wild – this aspect remains to be investigated in follow-up field studies. The perceptibility of smartphone notifications, including both different notification modalities as well as different designs, might change with external influences such as ambient sound or light. There are also aspects which might influence the perception but that cannot be covered or detected by a smartphone by now, e.g., the presence of a sleeping child or an indisposition of a user due to illness. Still, our results are a good basis that can be expanded in the future. With growing smartness of smartphones, more and more situations will be covered by context recognition systems.

12.3 ESM-RELATED ISSUES IN CONTEXT-AWARE DATA ASSESSMENT

It lies in the nature of user-centered study types such as ESM studies that participants might be prompted in situations in which they cannot or do not want to react to a notification and answer self-report questionnaires. This can be the case, for example, if the user is highly engaged in an activity or involved in a social interaction. However, this keeps us from capturing the true ground truth for situations such as those of low interruptibility or low receptivity as well as certain locations or activities. We have to trust the study participants to accurately report such kind of information retrospectively. This might require to clarify the importance of accurate reporting – for the sake of the investigation but, in the best case, also with an additional benefit for the user.

In addition, the quality of data in ESM studies is highly user-dependent. If participants are annoyed by prompts then their compliance might drop, possibly leading to an increased churn rate. Hence, it is recommendable to apply mechanisms to keep the participant motivated. As a first counteracting measure, we introduced inter-notification times and inquiry limits. Further, it is recommendable to consider gamification techniques. This might include rewarding the participation of the user in form of achievements or increasing reputation or embedding the self-reporting into a small game. Hsieh et al. [111] as well as Berkel at al. [190] already mentioned "motivational elements" as beneficial additions to ESM studies. It should be investigated if and how such elements improve the assessment of data labels.

Another issue that arises when discussing ESM studies is whether the knowledge about the participation in a user study influences the behavior of the participant. The simple fact of knowing that they are part of a user study might lead to a different behavior of a participant. This might be less critical if the assessed items relate to objective measurements such as smartphone position or activity labels. However, it might have an impact on the assessment of rather personal items such as social activity or personal interest. Willingly or unwillingly, participants might provide incorrect answers, e.g., they might report to be in company more often than usual or deny to be interested in a notification that actually matched their interests. This is also known as "social desirability bias" [72]. In addition, knowing that there will be ESM prompts might lead to a different receptivity: participants might actively wait for incoming notifications – in general or in certain situations such as location changes. However, this is a general issue that every user study has to face and not only the ones we conducted. Possible solutions include to mask the true nature of the user study, similar to what we did in our laboratory experiment in Chapter 9. For field studies, such a procedure is more challenging than for laboratory studies. If applicable, an option might be to distribute an app via the Google Play Store and infer findings from user feedback received directly or via app ratings.

13 Conclusion

In this dissertation, we investigated approaches to support the collection of labels for annotating smartphone data in ESM studies. A special focus lay on the perceptibility of smartphone notification which are the essential means to deliver ESM prompts to the user. At first, we revisit the challenges that we identified at the beginning of this dissertation. Next, we review the contribution and impact of our research to the community. Finally, we consider emerging future tasks.

13.1 Addressed Challenges

In Chapter 1, we presented four challenges that arise when using ESM for the assessment of data to label smartphone measurements.

Challenge 1 was to *prompt in situations of interest*. To achieve such a prompting, it was necessary to be able to interpret physical and virtual sensor values to decide if an event took place or not. We proposed ESMAC, a tool that facilitates the creation of ESM apps for Android and that addresses requirements mentioned in related work [60, 87, 140]. It provides different question types to choose from and offers access to a large set of sensors for automatic data logging as well as methods to detect events that enable event-based triggering of prompts. Within our evaluations, ESMAC performed comparable to a state of the art platform – with the difference, that it focuses more on context-awareness and offers more event triggers to prompt in situations of interest.

In an in-field user study, we confirmed the usefulness of event-triggers exemplified on an ESM study focusing on location and activity changes – two factors that are of interest for a variety of research fields [60, 155, 184]. The usage of event-prompts seems most suitable if the situations of interest that serve as event-triggers are in line with the study objective and the questionnaire content. We conclude that it is possible to prompt in situations of interest – if these events are covered by the ESM app or creation tool, respectively. Since ESMAC is free-to-use and open source, it allows continuous adaption to recent developments and new sensors available on Android and, thereby, an inclusion of new events.

Challenge 2 was to *reduce the burden of labeling tasks to the user* in ESM studies by *restricting the number of prompts* and *reducing the length of the self-report questionnaire*. These aspects can be addressed by context recognition and the utilization of smartphone capabilities. The restriction in number of prompts can be realized by introducing an inter-notification time and inquiry limit. Both are implemented in ESMAC and can be set when designing an ESM study. The length of the questionnaire can be reduced by automatic assessment of information via smartphone sensors instead of asking the user for this information – as suggested by Berkel et al. [50]. We equipped ESMAC with a large range of sensor access to realize this idea. In addition, ESMAC offers different question types, including single or multiple choice as well as closed-ended questions. According to Consolvo and Walker [60], avoiding open-ended questions reduces the length of the questionnaire or at least the time the user requires to fill out the self-report.

Challenge 3 was to *support the perceptibility of smartphone notifications*. That included the tasks to *find methods to assess factors that relate to the perceptibility* and, in a next step, to *investigate the relation between perception and the notification modality*.

We focused on the factors smartphone position, a combination of location and related (social) activity, and the perceived notification importance, since these showed a relation to receptivity and interruptibility – two concepts related to perception and perceptibility – in related work [60, 136, 155]. There is existing related work that investigated the assessment of each one of these aspects. However, each approach had its drawbacks such as a recognition accuracy with room for improvement or such as privacy concerns about the location assessment requiring a privacy-sensitive alternative. We successfully found and applied methods to assess these features with satisfying accuracy and in a more privacy-aware manner – by running a position transition correction and by using place types instead of raw GPS, respectively.

We were able to implement a classification system with satisfying recognition accuracy and a position-transition-correction algorithm to further improve the recognition accuracy for the smartphone position detection. For the location assessment, we found place types to be a useful and privacy-sensitive solution that fulfilled our requirements and which, in a next step, already served as a basis for the estimation of social activity. In an evaluation performed on data gathered in an in-field user study, the detection of being in company based on these place types showed satisfying recognition accuracy. Concerning the perceived importance of notifications, we were able to identify smartphone features that correlate with the perceived importance in general, but we also found that importance has four facets that need to be investigated separately in more detail.

Assuming that we know the current smartphone position or location, we ran two experimental lab studies to investigate the perception of smartphone notifications depending on the notification modality as the first independent variable and (a) the smartphone position or (b) the location in combination with the location-based activity as the second independent variable. In both experiments, we were able to identify suitable notification modalities per position or location, respectively, and to derive user preferences that can serve as a basis for automatic selection or recommendation of a notification modality.

Challenge 4 is related to Challenge 3. It considered the design of smartphone notifications and aimed at *improving their visibility*. Based on findings from interview sessions and literature, we created different designs to highlight important notifications in a way to increase their perceptibility. In a laboratory study, we evaluated how potential smartphone users perceive these designs. We were able to identify favorite designs and desired features, namely categorization by notification importance or source app and colored highlighting of important notifications. For Android 8.0 and possibly above, such highlightings might be realizable using the Android functionality to customize notifications.

Overall, we were able to successfully address the identified challenges and, thereby, to contribute to our research community. These contributions will be reviewed in more detail in the following section.

13.2 CONTRIBUTION TO THE RESEARCH COMMUNITY

Context-Aware Experience Sampling

There are several requirements a tool for creating ESM apps has to fulfill. One aspect is to provide a set of different question types, sensors for logging, and events for prompting. Another aspect is to allow settings for inter-notification time and inquiry limit. Moreover, the tool should be intuitive and easy-to-use and, in the best case, free-to-use and published open source to allow further development and individual adaptions. Last but not least, the tool should be able to provide easy-to-install apps for all available mobile OS, in the best case. It seems almost impossible to fit all these expectations into one tool, leading to several restrictions.

Many tools only offer premium services with high functionality for a remarkable amount of money that might not be available to researchers. Free alternatives often lack access to a required minimum of sensors or event-triggers or require specific user knowledge about programming or markup languages. By providing ESMAC, we tried to address as much of the mentioned expectations as possible, but lack to support all operating systems: we had to restrict ourselves to Android 4.4.4 and above. The choice of supported sensors and events is debatable and depends on the study objective. We tried to cover a wide range based on a survey among experts. We wanted to ensure that their requests are addressed and that requested features are implemented in ESMAC. Concurrent products might benefit from our results as well, since the survey results are published and allow them to examine the responses of the survey participants and implement desired sensors and events by themselves. In our opinion, ESMAC is a good alternative to pay-for-use tools and especially beneficial for junior researchers who have basic requirements in terms of sensors and events and cannot afford premium solutions. Due to its open source nature it is also suitable for researchers and ESM study designers who are interested in enhancing an existing system and adapting it to their needs, e.g., by including wearables as new sensor sources or by implementing new events.

Perception-Related Research

Different concepts proved to be related to the perception of a smartphone notification: the receptivity and interruptibility of the user [109, 155], the task engagement and current activity [147, 151], and the relevance and importance of the notification content [136, 138]. Related work examined smartphone features that relate to these concepts, evaluated classification methods, or investigated user preferences for notification delivery. However, to our best knowledge, perception was not yet considered in relation to ESM studies with the objective of assessing data labels via event-triggered prompts.

Researchers already found relations between the preferred notification modality and the smartphone position, user location, and user activity, respectively [60]. Though, an explicit investigation of these features in relation to the perceptibility of smartphone notifications and the user preference was not yet conducted and addressed by us. Our findings allow to infer first recommendations for automatic selection of suitable notification modalities and are a good basis for follow-up in-field user studies.

The classification mechanisms that we developed as a basis for the perception research can be applied in other areas as well. For example, they can be used for an automatic assessment of desired features such as place type and social activity indicator as alternative to user-reported values in interruptibility detection [155]. It is possible to rely on single classifiers only or to combine their predictive power. A combination might be reasonable, e.g., since our social activity recognition is based on place types which requires to assess the user location in any case. Automatically collecting multiple features at a time might be relevant for specific research applications, e.g., automatic assessment of movement and activity information to support the monitoring of patients suffering from affective disorders.

Due to the interrelation of perception, receptivity, interruptibility, etc., it is recommendable to collaborate with international scientists and seek joint solutions.

For example, it is worth to consider building a smart notification management system that is aware of user needs and acts in accordance with the mentioned concepts and the user's individual characteristics.

Relevancy for Future Developments on Mobile OS

The results that we gained can influence future developments on mobile OS in terms of new features that could be implemented and provided to the users. Recent notification customization options available for Android [3] already suggest that mobile OS vendors are interested in adapting their system to the need of the user and the needs of app designers. We only considered the user-side in our research, but it is reasonable that app developers are interested in user-friendly notification delivery as well, e.g., to avoid annoying the user what could cause a deinstallation of their app. Our research might serve as a basis for adaptions of the notification modality, including the ringer mode [136] and alert type [139], the design of a smartphone notification but also for the scheduled delivery of smartphone notifications.

A first step might be to automatically select notification modalities based on the smartphone position or the current place type. The smartphone could switch into silent or "do not disturb" mode automatically when its user enters a silent area or a "do not disturb" location such as "cinema" or "library". Moreover, the smartphone could choose ringtone as alert type for very important notifications which rarely occur but usually require immediate attention. It is crucial to implement adaptive classifiers and not generic ones: even though there is a common basis for preferences, smartphone users are individuals and act differently. It is necessary to provide a basis but allow users to adjust the settings to their needs.

A second step might be to adapt the design of notifications and to change their appearance in the notification drawer. We investigated different designs to highlight important notifications or to group notifications either by app or by importance. Some of the design changes are already realizable using the custom notifications offered by Android 8.0 and higher. Hence, it seems promising that customized styles will be available to a large set of devices within a few years. Mobile OS should offer ways to change the design of the notification but also to provide ways to group notifications based on certain criteria. Again, it is important to support individual preferences and to offer multiple options. The question arises how and when to change the design of a notification. This might depend on the perceived importance like in our case. Though, at first, it might be easier to let the user define simple rules for design adaptions, e.g., depending on the app or the other party that sent a notification. Grouped notifications might exceed pure sorting by app, e.g., by providing an overview of multiple instant messages sent from one person but via different platforms or by summarizing multiple notifications with similar content or certain keywords.

Finally, notifications might be filtered or delivered based on a specific schedule according to the estimated perceived importance (cf. Chapter 8), the anticipated user reaction [138], or the predicted interruptibility [155]. There is much research about these aspects that can be taken into account and included into further investigations. It could be interesting to examine the effects on the user if multiple of those approaches are combined or if all of them are offered to the user and they can select them at their whim. While related work rather focused on filtering unimportant notifications [138], we suggest to define rules that indicate important notifications. Based on our experience (cf. Chapter 8), the number of important notifications is much smaller than the number of unimportant or neutral notifications. Hence, it is reasonable to investigate means to filter for those notifications and apply special announcements such as ringtone alert or visual highlight instead of trying to identify and filter unimportant notifications. This would allow the user to use their usual, rather unobtrusive notification modality as default without the need of fearing to miss important notifications.

Overall, automatic modality selection and design adaptions could prove to be beneficial for users who desire to customize their smartphone setting and adapt the device usage to their needs. This might lead to reduced technostress and increased user experience. However, at first, it will require the user to manage many settings. The mobile OS should aim at providing intuitive and easy settings and, possibly, not offer all functionalities at once but release them one after another to allow the user to become familiar with them. It is recommendable not to have one fixed design, but to adapt to the user needs either automatically, adaptively by taking user feedback into account, or manually by offering options or a set of rules the user can choose from.

13.3 FUTURE WORK

THE ESMAC TOOL In the future, the ESMAC system might be enhanced by further sensors or events. Due to their rising number, wearables such as smartwatches are good candidates for complementary sensor sources, e.g., heart rate. It is recommendable to implement interfaces to include externally trained classifiers which allow further contextual interpretation of sensor data and, thereby, provide additional event-triggers.

LOCATION AND ACTIVITY CHANGES AS EVENT-TRIGGERS Keeping location and activity changes as a use case, we suggest to design and conduct studies with patients suffering from affective disorders as participants – similar to Grünerbl et al. [100]. This might allow to gain insights about their location change behavior and to evaluate the usefulness of our location detection for episode diagnosis. It is also recommendable to run further ESM studies with alternative study objectives to evaluate the usefulness of event-triggers for other application areas.

ASSESSMENT OF PERCEPTIBILITY-RELATED FACTORS Detection mechanisms for the investigated factors might prove to be useful in related research areas. Automatic detection of smartphone position, location, and social activity indicator can be used as an alternative to manual assessment of such factors as conducted in previous interruptibility detection research by Pejovic et al. [155] or Mehrotra et al. [138]. These detection mechanisms might also be useful in ambulatory assessment: for example, the location assessment might be beneficial for tracking the mobility of depressive patients [164].

APPLICATION OF PERCEPTIBILITY-RELATED FINDINGS These aspects might be investigated in future studies or possibly even implemented in future mobile OS versions. Moreover, it might be worth to consider alternative ways to present notifications to the user such as ambient displays [143], nearby displays, or smartwatches [191]. In any case, we recommend that the underlying system offers multiple options for the user to choose from instead of one fix solution or that the system includes an adaptive classifier that learns specific habits and preferences of the user over time instead of using one generic classifier for all users.

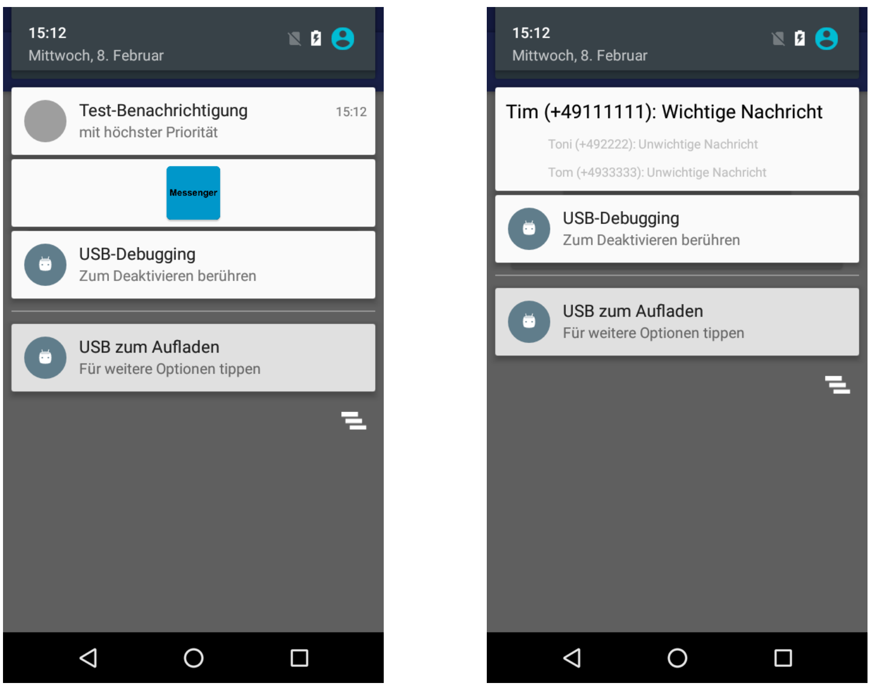
COMBINATION OF DIFFERENT APPROACHES In the future, the identified detection mechanisms and user preferences for notification modalities might be embedded into the ESMAC tool to provide it not only with further event-triggers, but also with additional awareness, e.g., for a user's receptivity or interruptibility. Study designers have to ensure that they do not overwhelm their study participants with features or settings. The number of selected sensors, event-triggers and settings adjustable by the user should be kept as low as possible and only as high as necessary and always be related to the study objective. Since smartphone users have individual preferences, it might be inevitable that a notification management system monitors the users for a certain time to learn their habits and to be able to correctly adapt generalized classifiers to individual peculiarities. It is questionable if ESM studies allow for such a training period or if this is a general task that requires the support of the mobile OS. In any case, we recommend to fuse the results and insights gained in this dissertation to support context-aware ESM studies, to improve the user experience when receiving notifications, and, possibly, to reduce negative effects that a smartphone might induce on its user.

Part VI.

Appendix

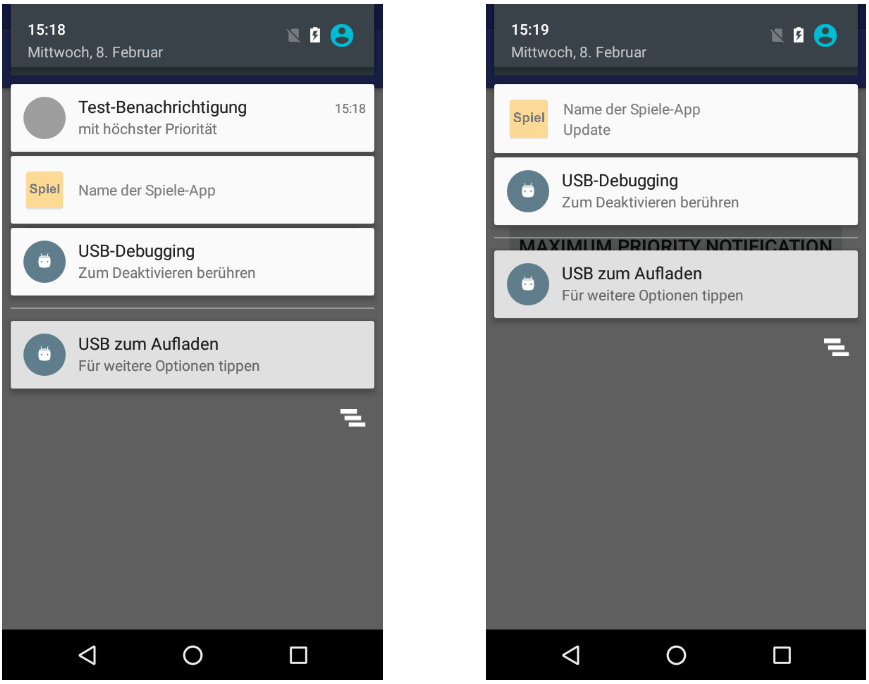
A Designs to Highlight Important Notifications

In the following, we present designs to highlight important notifications. The designs were sketched by the five participants who attended the interviews we presented in Chapter 11.



(a) Normal Notification (b) Expanded Notification

Figure 29.: Participant 1's design of a notification originating from a messenger app.



(a) Normal Notification (b) Expanded Notification

Figure 30.: Participant 2's design of a notification originating from a gaming app.

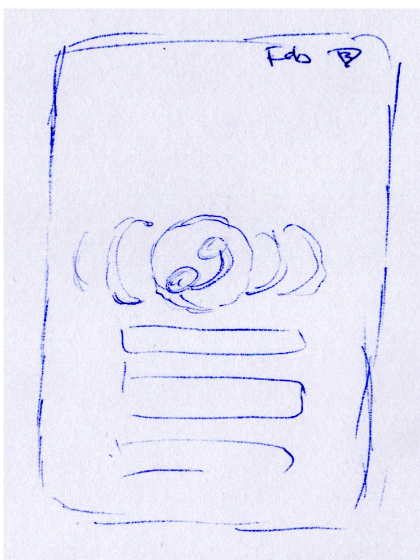
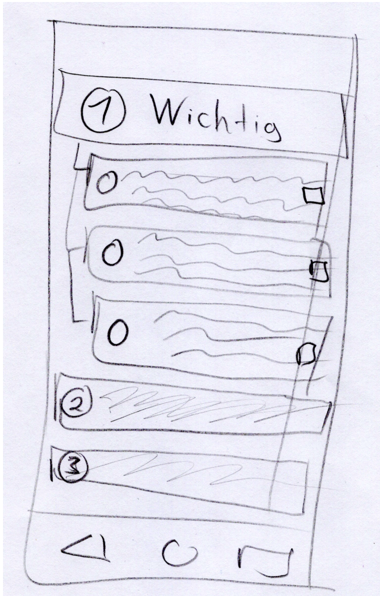
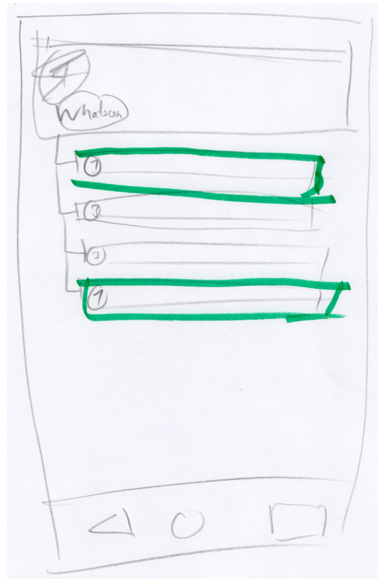


Figure 31.: Another design suggestion by participant 2: a carousel of app icons, showing the notifications per app if the icon is tapped.



(a) Notifications ordered by importance. ("Wichtig" = "important")



(b) Notifications sorted by application with colored highlight of important notifications.

Figure 32.: Notification designs created by participant 3.

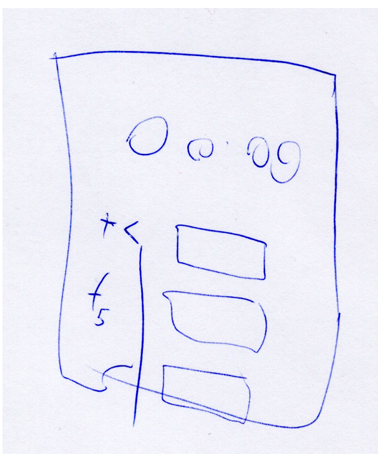


Figure 33.: Notification design by participant 4: presentation of notifications per app.

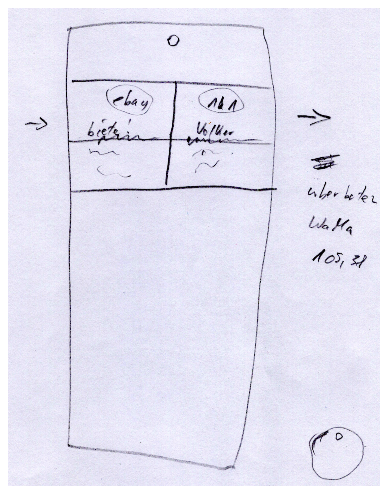


Figure 34.: Notification design of participant 5: visualization of notifications in form of a live ticker.

Bibliography

[1] Society for ambulatory assessment. http://www.saa2009.org. Accessed: 2018-04-25.

[2] Android bluetooth access. https://support.google.com/nexus/answer/2819579?hl=en-GB, . Accessed: 2018-04-25.

[3] Create a custom notification layout | android developer. https://developer.android.com/training/notify-user/custom-notification.html, . Accessed: 2018-05-28.

[4] Android developer. https://developer.android.com/index.html, . Accessed: 2018-04-25.

[5] Notifications overview | android developer. https://developer.android.com/guide/topics/ui/notifiers/notifications.html, . Accessed: 2018-04-25.

[6] Sensors overview | android developer. https://developer.android.com/guide/topics/sensors/sensors_overview, . Accessed: 2018-04-25.

[7] Aware - open-source context instrumentation framework for everyone. http://www.awareframework.com/. Accessed: 2018-04-25.

[8] Create notification | android developer. https://developer.android.com/training/notify-user/build-notification. Accessed: 2018-05-29.

[9] Dustin short movie. http://www.cgmeetup.net/home/dustin-short-film/. Accessed: 2018-04-25.

[10] Esm capture. http://esmcapture.com/, . Accessed: 2018-04-25.

[11] Github - esmac. https://github.com/teco-kit/ESMAC, . Accessed: 2018-06-22.

[12] Flash notification. https://play.google.com/store/apps/details?id=com.istudio.flashnotification&hl=en_US. Accessed: 2018-08-27.

[13] Detected activity | google apis for android |google developers. https://developers.google.com/android/reference/com/google/android/gms/location/DetectedActivity. Accessed: 2018-04-25.

[14] Mobile ecological momentary assessment software. http://mema.ilumivu.com/. Accessed: 2018-04-25.

[15] Designer tool window - help | intellij idea. https://www.jetbrains.com/help/idea/2017.1/designer-tool-window.html. Accessed: 2018-07-12.

[16] Life data @ experience sampling applications for research. https://www.lifedatacorp.com/. Accessed: 2018-04-25.

[17] Metricwire inc. - breakthrough researchmetricwire inc. https://www.metricwire.com/. Accessed: 2018-04-25.

[18] Samsung galaxy s6 edge: How to manage and use quick responses feature in messages app in android 5.1.1 lollipop. http://inside-galaxy.blogspot.de/2016/04/samsung-galaxy-s6-edge-how-to-manage.html. Accessed: 2018-04-25.

[19] movisensXS. http://www.movisens.com/en/products/movisensxs/, . Accessed: 2018-04-25.

[20] movisens Products. https://www.movisens.com/en/products/, . Accessed: 2018-04-25.

[21] the myexperience tool. http://myexperience.sourceforge.net/. Accessed: 2018-04-25.

[22] ohmage | open data collection. http://ohmage.org/index.html. Accessed: 2018-04-25.

[23] Paco app. https://www.pacoapp.com/. Accessed: 2018-04-25.

[24] Software - society for ambulatory assessment. http://www.saa2009.org/?page_id=57. Accessed: 2018-04-25.

[25] Global mobile os market share in sales to end users from 1st quarter 2009 to 2nd quarter 2017. https://www.statista.com/statistics/266136/global-market-share-held-by-smartphone-operating-systems/, . Accessed: 2018-04-25.

[26] Global number of public hotspots from 2009 to 2015. https://www.statista.com/statistics/218596/global-number-of-public-hotspots-since-2009/, . Accessed: 2018-06-22.

[27] Weka 3 - data mining with open source machine learning software in java. http://www.cs.waikato.ac.nz/ml/weka/. Accessed: 2018-04-25.

[28] disruption - wiktionary. https://en.wiktionary.org/wiki/disruption, . Accessed: 2018-04-25.

[29] distracted - wiktionary. https://en.wiktionary.org/wiki/distracted, . Accessed: 2018-04-25.

[30] distraction - wiktionary. https://en.wiktionary.org/wiki/distraction, . Accessed: 2018-04-25.

[31] disturbance - wiktionary. https://en.wiktionary.org/wiki/disturbance, . Accessed: 2018-04-25.

[32] disturbing - wiktionary. https://en.wiktionary.org/wiki/disturbing, . Accessed: 2018-04-25.

[33] Saeed Abdullah, Mark Matthews, Ellen Frank, Gavin Doherty, Geri Gay, and Tanzeem Choudhury. Automatic detection of social rhythms in bipolar disorder. *Journal of the American Medical Informatics Association*, 23(3):538–543, 2016.

[34] Piotr D Adamczyk and Brian P Bailey. If not now, when?: the effects of interruption at different moments within task execution. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 271–278. ACM, 2004.

[35] Khaled Alanezi and Shivakant Mishra. Design, implementation and evaluation of a smartphone position discovery service for accurate context sensing. *Computers & Electrical Engineering*, 44(0):307 – 323, 2015. ISSN 0045-7906. doi: http://dx.doi.org/10.1016/j.compeleceng.2015.01.015. URL http://www.sciencedirect.com/science/article/pii/S004579061500021X.

[36] William Albert and Thomas Tullis. *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Newnes, 2013.

[37] AlgoSnap Inc. Crowdsignals.io: A massive new mobile data collection campaign. http://crowdsignals.io/, 2016. Accessed: 2018-07-02.

[38] AlgoSnap Inc. Algosnap - crowdsignals.io pilot dataset reference. published as part of the final dataset; not yet available online, 2016.

[39] Mohamed R Amer, Dan Xie, Mingtian Zhao, Sinisa Todorovic, and Song-Chun Zhu. Cost-sensitive top-down/bottom-up inference for multiscale activity recognition. In *European Conference on Computer Vision*, pages 187–200. Springer, 2012.

[40] Stephen A. Antos, Mark V. Albert, and Konrad P. Kording. Hand, belt, pocket or bag: Practical activity tracking with mobile phones. *Journal of Neuroscience Methods*, 231(0):22 – 30, 2014. ISSN 0165-0270. doi: http://dx.doi.org/10.1016/j.jneumeth.2013.09.015. URL http://www.sciencedirect.com/science/article/pii/S0165027013003269. Motion Capture in Animal Models and Humans.

[41] Julie Aranda, Noor Ali-Hasan, and Safia Baig. I'm just trying to survive: an ethnographic look at mobile notifications and attention management. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 564–574. ACM, 2016.

[42] Anja Bachmann, Christoph Klebsattel, Matthias Budde, Till Riedel, Michael Beigl, Markus Reichert, Philip Santangelo, and Ulrich Ebner-Priemer. How to use smartphones for less obtrusive ambulatory mood assessment and mood recognition. MCSS'15. ACM, 2015.

[43] Anja Bachmann, Robert Zetzsche, Till Riedel, Michael Beigl, Markus Reichert, Philip Santangelo, and Ulrich Ebner-Priemer. Identification of relevant sensor sources for context-aware esm apps in ambulatory assessment. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pages 265–268. ACM, 2015.

[44] Anja Bachmann, Robert Zetzsche, Andrea Schankin, Till Riedel, Michael Beigl, Markus Reichert, Philip Santangelo, and Ulrich Ebner-Priemer. Esmac: a web-based configurator for context-aware experience sampling apps in ambulatory assessment. In *Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare (MobiHealth)*, pages 15–18, 2015.

[45] Brian P Bailey and Joseph A Konstan. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in human behavior*, 22(4):685–708, 2006.

[46] Brian P Bailey, Joseph A Konstan, and John V Carlis. Measuring the effects of interruptions on task performance in the user interface. In *Systems, Man,*

and Cybernetics, 2000 IEEE International Conference on, volume 2, pages 757–762. IEEE, 2000.

[47] Roy F Baumeister, Ellen Bratslavsky, Catrin Finkenauer, and Kathleen D Vohs. Bad is stronger than good. *Review of general psychology*, 5(4):323, 2001.

[48] James Bo Begole, Nicholas E Matsakis, and John C Tang. Lilsys: sensing unavailability. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 511–514. ACM, 2004.

[49] Wicher Bergsma. A bias-correction for cramér's v and tschuprow's t. *Journal of the Korean Statistical Society*, 42(3):323–328, 2013.

[50] Niels Van Berkel, Denzil Ferreira, and Vassilis Kostakos. The experience sampling method on mobile devices. *ACM Computing Surveys (CSUR)*, 50 (6):93, 2017.

[51] Scott Bradner. Key words for use in rfcs to indicate requirement levels. 1997.

[52] Kevin Brennan et al. *A guide to the business analysis body of knowledge (Babok Guide)*. International Institute of business analysis, 2009.

[53] John Brooke. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189, 1996.

[54] Scott Carter and Jennifer Mankoff. When participants do the capturing: the role of media in diary studies. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 899–908. ACM, 2005.

[55] Yongguang Chen and Hisashi Kobayashi. Signal strength based indoor geolocation. 1, 2002.

[56] Jungchan Cho, Inhwan Hwang, and Songhwai Oh. Vibration-based surface recognition for smartphones. In *Embedded and Real-Time Computing Systems and Applications (RTCSA), 2012 IEEE 18th International Conference on*, pages 459–464. IEEE, 2012.

[57] Wongun Choi, Khuram Shahid, and Silvio Savarese. Learning context for collective activity recognition. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 3273–3280. IEEE, 2011.

[58] Jacob Cohen. Statistical power analysis for the behavioral sciences 2nd edn, 1988.

[59] Jonathan S Comer and Philip C Kendall. *The Oxford handbook of research strategies for clinical psychology*. Oxford University Press, 2013.

[60] Sunny Consolvo and Miriam Walker. Using the experience sampling method to evaluate ubicomp applications. *IEEE Pervasive Computing*, 2(2): 24–31, 2003.

[61] Harald Cramér. *Mathematical Methods of Statistics (PMS-9)*, volume 9. Princeton university press, 2016.

[62] M Csikszentmihalyi and R Larson. Validity and reliability of the experience-sampling method. *The Journal of nervous and mental disease*, 175(9):526, 1987.

[63] Edward Cutrell, Mary Czerwinski, and Eric Horvitz. Notification, disruption, and memory: Effects of messaging interruptions on memory and performance. In *Human-computer Interaction: INTERACT'01: IFIP TC. 13 International Conference on Human-Comupter Interaction, 9th-13th July 2001, Tokyo, Japan*, page 263. IOS Press, 2001.

[64] Mary Czerwinski, Edward Cutrell, and Eric Horvitz. Instant messaging: Effects of relevance and timing. In *People and computers XIV: Proceedings of HCI*, volume 2, pages 71–76, 2000.

[65] Gerald C Davidson and John M Neale. *Abnormal psychology*. Wiley publishing co., 1996.

[66] Luigi De Russis and Alberto Monge Roffarello. On the benefit of adding user preferences to notification delivery. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1561–1568. ACM, 2017.

[67] G. Demiroz and A. Guvenir. Classification by voting feature intervals. In *9th European Conference on Machine Learning*, pages 85–92. Springer, 1997.

[68] Genevieve F Dunton, Jimi Huh, Adam M Leventhal, Nathaniel Riggs, Donald Hedeker, Donna Spruijt-Metz, and Mary Ann Pentz. Momentary assessment of affect, physical feeling states, and physical activity in children. *Health Psychology*, 33(3):255, 2014.

[69] Nathan Eagle and Alex (Sandy) Pentland. Reality mining: Sensing complex social systems. *Personal Ubiquitous Comput.*, 10(4), 2006.

[70] Ulrich W Ebner-Priemer and Günther Sawitzki. Ambulatory assessment of affective instability in borderline personality disorder: The effect of the sampling frequency. *European Journal of Psychological Assessment*, 23(4), 2007.

[71] Ulrich W Ebner-Priemer, Michael Eid, Nikolaus Kleindienst, Simon Stabenow, and Timothy J Trull. Analytic strategies for understanding affective (in) stability and other dynamic processes in psychopathology. *Journal of abnormal psychology*, 118(1):195, 2009.

[72] Allen L Edwards. The social desirability variable in personality assessment and research. 1957.

[73] Moustafa Elhamshary and Moustafa Youssef. Checkinside: A fine-grained indoor location-based social network. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 607–618. ACM, 2014.

[74] David H Epstein, Matthew Tyburski, Ian M Craig, Karran A Phillips, Michelle L Jobes, Massoud Vahabzadeh, Mustapha Mezghanni, Jia-Ling Lin, C Debra M Furr-Holden, and Kenzie L Preston. Real-time tracking of neighborhood surroundings and mood in urban drug misusers: application of a new method to study behavior in its geographical context. *Drug & Alcohol Dependence*, 134:22–29, 2014.

[75] Anja Exler, Marcel Braith, Andrea Schankin, and Michael Beigl. Preliminary investigations about interruptibility of smartphone users at specific place types. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pages 1590–1595. ACM, 2016.

[76] Anja Exler, Matthias Urschel, Andrea Schankin, and Michael Beigl. Smartphone-based detection of location changes using wifi data. In *International Conference on Wireless Mobile Communication and Healthcare*, pages 164–167. Springer, 2016.

[77] Anja Exler, Christian Dinse, Zeynep Günes, Nadim Hammoud, Steffen Mattes, and Michael Beigl. Investigating the perceptibility different notification types on smartphones depending on the smartphone position. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pages 970–976. ACM, 2017.

[78] Anja Exler, Sebastian Kramer, Miguel Angel Meza Martinez, Christian Navolskyi, Michael Vogt, and Michael Beigl. Suitability of event-based prompts in experience sampling studies focusing on location changes. In *Proceedings of the 7th EAI International Conference on Wireless Mobile Communication and Healthcare (MobiHealth)*, 2017.

[79] Anja Exler, Marcel Braith, Kristina Mincheva, Andrea Schankin, and Michael Beigl. Smartphone-based estimation of a user being in company or alone based on place, time, and activity. In *International Conference on Mobile Computing, Applications, and Services*, pages 74–89. Springer, 2018.

[80] Anja Exler, Christoph Michel, and Michael Beigl. On-body smartphone position detection with position transition correction based on the hand state. In

International Conference on Mobile Computing, Applications, and Services, pages 271–276. Springer, 2018.

[81] Hans Jürgen Eysenck and SGB Eysenck. The eysenck personality inventory. 1965.

[82] Louise Farnworth, Elphine Mostert, Susan Harrison, and Diane Worrell. The experience sampling method: Its potential use in occupational therapy research. *Occupational Therapy International*, 3(1):1–17, 1996.

[83] Maria Faurholt-Jepsen, Mads Frost, Maj Vinberg, Ellen Margrethe Christensen, Jakob E Bardram, and Lars Vedel Kessing. Smartphone data as objective measures of bipolar disorder symptoms. volume 217, pages 124–127, 2014.

[84] Andy Field. *Discovering statistics using IBM SPSS statistics*. Sage, 2013.

[85] Joel E Fischer, Nick Yee, Victoria Bellotti, Nathan Good, Steve Benford, and Chris Greenhalgh. Effects of content and time of delivery on receptivity to mobile interruptions. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, pages 103–112. ACM, 2010.

[86] Joel E Fischer, Chris Greenhalgh, and Steve Benford. Investigating episodes of mobile phone activity as indicators of opportune moments to deliver notifications. In *Proceedings of the 13th international conference on human computer interaction with mobile devices and services*, pages 181–190. ACM, 2011.

[87] Cynthia D Fisher and March L To. Using experience sampling methodology in organizational behavior. *Journal of Organizational Behavior*, 33(7):865–877, 2012.

[88] Robert Fisher and Reid Simmons. Smartphone interruptibility using density-weighted uncertainty sampling with reinforcement learning. In *Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on*, volume 1, pages 436–441. IEEE, 2011.

[89] Milton Friedman. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the american statistical association*, 32(200):675–701, 1937.

[90] Jon Froehlich, Mike Y Chen, Sunny Consolvo, Beverly Harrison, and James A Landay. Myexperience: a system for in situ tracing and capturing of user feedback on mobile phones. In *Proceedings of the 5th international conference on Mobile systems, applications and services*, pages 57–70. ACM, 2007.

[91] Kaori Fujinami. On-body smartphone localization with an accelerometer. *Information*, 7(2):21, 2016.

[92] Kaori Fujinami and Satoshi Kouchi. Recognizing a mobile phone's storing position as a context of a device and a user. In *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services*, pages 76–88. Springer, 2012.

[93] Shravan Gaonkar, Jack Li, Romit Roy Choudhury, Landon Cox, and Al Schmidt. Micro-blog: sharing and querying content through mobile phones and social participation. In *MobiSys*, pages 174–186. ACM, 2008.

[94] Wayne CW Giang, Liberty Hoekstra-Atwood, and Birsen Donmez. Driver engagement in notifications: a comparison of visual-manual interaction between smartwatches and smartphones. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 58, pages 2161–2165. Sage Publications Sage CA: Los Angeles, CA, 2014.

[95] Tony Gillie and Donald Broadbent. What makes interruptions disruptive? a study of length, similarity, and complexity. *Psychological research*, 50(4): 243–250, 1989.

[96] Jennifer Gluck, Andrea Bunt, and Joanna McGrenere. Matching attentional draw with utility in interruption. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 41–50. ACM, 2007.

[97] Antonio Gomes, Andrea Nesbitt, and Roel Vertegaal. Morephone: a study of actuated shape deformations for flexible thin-film smartphone notifications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 583–592. ACM, 2013.

[98] Google. Place Types | Google Places API| Google Developers. `https://developers.google.com/places/supported_types`. Accessed: 2018-06-22.

[99] Dawud Gordon, Jan-Hendrik Hanne, Martin Berchtold, Ali Asghar Nazari Shirehjini, and Michael Beigl. Towards collaborative group activity recognition using mobile devices. *Mobile Networks and Applications*, 18(3):326–340, 2013.

[100] Agnes Gruenerbl, Venet Osmani, Gernot Bahle, Jose C Carrasco, Stefan Oehler, Oscar Mayora, Christian Haring, and Paul Lukowicz. Using smart phone mobility traces for the diagnosis of depressive and manic episodes in bipolar patients. In *Proceedings of the 5th Augmented Human International Conference*, page 38. ACM, 2014.

[101] Agnes Grünerbl, Amir Muaremi, Venet Osmani, Gernot Bahle, Stefan Oehler, Gerhard Tröster, Oscar Mayora, Christian Haring, and Paul Lukowicz. Smartphone-based recognition of states and state changes in bipolar disorder patients. *IEEE Journal of Biomedical and Health Informatics*, 19(1): 140–148, 2015.

[102] Franz Halberg, Yung Liang Tong, and EA Johnson. Circadian system phase - an aspect of temporal morphology; procedures and illustrative examples. In *The cellular aspects of biorhythms*, pages 20–48. Springer, 1967.

[103] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[104] Peter A Hancock. A dynamic model of stress and sustained attention. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 31(5): 519–537, 1989.

[105] Sandra G Hart and Lowell E Staveland. Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183, 1988.

[106] Stefan Hey, Panagiota Anastasopoulou, André Bideaux, and Wilhelm Stork. Recent developments of ambulatory assessment methods. *Zeitschrift für Neuropsychologie*, 25(4), 2014.

[107] Suzanne Hidi and Judith M Harackiewicz. Motivating the academically unmotivated: A critical issue for the 21st century. *Review of educational research*, 70(2):151–179, 2000.

[108] Tomoya Hirano and Takuya Maekawa. A hybrid unsupervised/supervised model for group activity recognition. In *Proceedings of the 2013 International Symposium on Wearable Computers*, pages 21–24. ACM, 2013.

[109] Joyce Ho and Stephen S Intille. Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. In *CHI '05*, 2005.

[110] Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979.

[111] Gary Hsieh, Ian Li, Anind Dey, Jodi Forlizzi, and Scott E Hudson. Using visualizations to increase compliance in experience sampling. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 164–167. ACM, 2008.

[112] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM, 2004.

[113] Scott Hudson, James Fogarty, Christopher Atkeson, Daniel Avrahami, Jodi Forlizzi, Sara Kiesler, Johnny Lee, and Jie Yang. Predicting human interruptibility with sensors: A wizard of oz feasibility study. In *CHI '03*, 2003.

[114] Shamsi T Iqbal and Brian P Bailey. Effects of intelligent notification management on users and their tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 93–102. ACM, 2008.

[115] Quentin Jones, Sukeshini A Grandhi, Steve Whittaker, Keerti Chivakula, and Loren Terveen. Putting systems into place: a qualitative study of design requirements for location-aware community systems. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 202–211. ACM, 2004.

[116] Nicky Kern and Bernt Schiele. Context-aware notification for wearable computing. In *null*, page 223. IEEE, 2003.

[117] Mikkel Baun Kjærgaard, Martin Wirz, Daniel Roggen, and Gerhard Tröster. Detecting pedestrian flocks by fusion of multi-modal sensors in mobile phones. UbiComp'12. ACM, 2012.

[118] John G Kreifeldt and ME McCarthy. Interruption as a test of the user-computer interface. 1981.

[119] William H Kruskal and W Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621, 1952.

[120] Lun-Wei Ku and Hsin-Hsi Chen. Mining opinions from the web: Beyond relevance retrieval. *Journal of the American Society for Information Science and Technology*, 58(12):1838–1850, 2007.

[121] K. Kunze and P. Lukowicz. Using acceleration signatures from everyday activities for on-body device location. In *Wearable Computers, 2007 11th IEEE International Symposium on*, pages 115–116, Oct 2007. doi: 10.1109/ISWC.2007.4373794.

[122] Kai Kunze, Paul Lukowicz, Holger Junker, and Gerhard Tröster. Where am i: Recognizing on-body positions of wearable sensors. In Thomas Strang and Claudia Linnhoff-Popien, editors, *Location- and Context-Awareness*, volume 3479 of *Lecture Notes in Computer Science*, pages 264–275. Springer Berlin

Heidelberg, 2005. ISBN 978-3-540-25896-4. doi: 10.1007/11426646_25. URL http://dx.doi.org/10.1007/11426646_25.

[123] Kostadin Kushlev, Jason Proulx, and Elizabeth W Dunn. Silence your phones: Smartphone notifications increase inattention and hyperactivity symptoms. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1011–1020. ACM, 2016.

[124] Neal Lathia, Kiran K Rachuri, Cecilia Mascolo, and Peter J Rentfrow. Contextual dissonance: Design bias in sensor-based experience sampling methods. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 183–192. ACM, 2013.

[125] Bettina Laugwitz, Theo Held, and Martin Schrepp. Construction and evaluation of a user experience questionnaire. *HCI and Usability for Education and Work*, 2008. p. 63-76.

[126] Yu-Kang Lee, Chun-Tuan Chang, You Lin, and Zhao-Hong Cheng. The dark side of smartphone usage: Psychological traits, compulsive behavior and technostress. *Computers in Human Behavior*, 31:373–383, 2014.

[127] Shengnan Li, Zheng Qin, and Houbing Song. A temporal-spatial method for group detection, locating and tracking. *IEEE Access*, 4:4484–4494, 2016.

[128] Guanqing Liang, Jiannong Cao, and Weiping Zhu. Circlesense: A pervasive computing system for recognizing social activities. In *Pervasive computing and communications (percom), 2013 ieee international conference on*, pages 201–206. IEEE, 2013.

[129] Lin Liao, Dieter Fox, and Henry Kautz. Location-based activity recognition. In *Advances in Neural Information Processing Systems*, pages 787–794, 2006.

[130] Lin Liao, Dieter Fox, and Henry Kautz. Extracting places and activities from gps traces using hierarchical conditional random fields. *The International Journal of Robotics Research*, 26(1):119–134, 2007.

[131] Lin Liao, Dieter Fox, and Henry Kautz. Hierarchical conditional random fields for gps-based activity recognition. In *Robotics Research*, pages 487–506. Springer, 2007.

[132] Hugo Lopez-Tovar, Andreas Charalambous, and John Dowell. Managing smartphone interruptions through adaptive modes and modulation of notifications. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, pages 296–299. ACM, 2015.

[133] Henry B Mann and Donald R Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60, 1947.

[134] Afra Mashhadi, Akhil Mathur, and Fahim Kawsar. The myth of subtle notifications. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 111–114. ACM, 2014.

[135] Daniel C McFarlane and Kara A Latorella. The scope and importance of human interruption in human-computer interaction design. *Human-Computer Interaction*, 17(1):1–61, 2002.

[136] Abhinav Mehrotra, Mirco Musolesi, Robert Hendley, and Veljko Pejovic. Designing content-driven intelligent notification mechanisms for mobile applications. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 813–824. ACM, 2015.

[137] Abhinav Mehrotra, Jo Vermeulen, Veljko Pejovic, and Mirco Musolesi. Ask, but don't interrupt: the case for interruptibility-aware mobile experience sampling. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pages 723–732. ACM, 2015.

[138] Abhinav Mehrotra, Robert Hendley, and Mirco Musolesi. Prefminer: mining user's preferences for intelligent mobile notification management. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 1223–1234. ACM, 2016.

[139] Abhinav Mehrotra, Veljko Pejovic, Jo Vermeulen, Robert Hendley, and Mirco Musolesi. My phone and me: understanding people's receptivity to mobile notifications. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 1021–1032. ACM, 2016.

[140] Geoffrey Miller. The smartphone psychology manifesto. *Perspectives on Psychological Science*, 7(3), 2012.

[141] E. Mok and G. Retscher. Location determination using wifi fingerprinting versus wifi trilateration. *Journal of Location Based Services*, 1(2), 2007.

[142] Christopher A Monk, Deborah A Boehm-Davis, and J Gregory Trafton. The attentional costs of interrupting task performance at various stages. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 46, pages 1824–1828. SAGE Publications Sage CA: Los Angeles, CA, 2002.

[143] Heiko Müller, Jutta Fortmann, Martin Pielot, Tobias Hesselmann, Benjamin Poppinga, Wilko Heuten, Niels Henze, and Susanne Boll. Ambix: Designing ambient light information displays. In *Proceedings of Designing Interactive Lighting workshop at DIS*. Citeseer, 2012.

[144] Darren Newtson and Gretchen Engquist. The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, 12(5):436–450, 1976.

[145] Jakob Nielsen and Thomas K Landauer. A mathematical model of the finding of usability problems. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, pages 206–213. ACM, 1993.

[146] Petteri Nurmi. Identifying meaningful places. 2009.

[147] Mikio Obuchi, Wataru Sasaki, Tadashi Okoshi, Jin Nakazawa, and Hideyuki Tokuda. Investigating interruptibility at activity breakpoints using smartphone activity recognition api. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pages 1602–1607. ACM, 2016.

[148] Tadashi Okoshi, Julian Ramos, Hiroki Nozaki, Jin Nakazawa, Anind K Dey, and Hideyuki Tokuda. Attelia: Reducing user's cognitive load due to interruptive notifications on smart phones. In *Pervasive Computing and Communications (PerCom), 2015 IEEE International Conference on*, pages 96–104. IEEE, 2015.

[149] Tadashi Okoshi, Jin Nakazawa, and Hideyuki Tokuda. Interruptibility research: opportunities for future flourishment. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pages 1524–1529. ACM, 2016.

[150] Tadashi Okoshi, Anja Exler, Alexandra Voit, Dominik Weber, Martin Pielot, Benjamin Poppinga, Niels Henze, Sven Gehring, Matthias Böhmer, SeungJun Kim, et al. Ubittention 2017: 2nd international workshop on smart & ambient notification and attention management. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pages 893–897. ACM, 2017.

[151] Tadashi Okoshi, Kota Tsubouchi, Masaya Taji, Takanori Ichikawa, and Hideyuki Tokuda. Attention and engagement-awareness in the wild: A large-scale study with adaptive notifications. In *2017 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 100–110. IEEE, 2017.

[152] Jeongyeup Paek, Joongheon Kim, and Ramesh Govindan. Energy-efficient rate-adaptive gps-based positioning for smartphones. In *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pages 299–314. ACM, 2010.

[153] Chunjong Park, Junsung Lim, Juho Kim, Sung-Ju Lee, and Dongman Lee. Don't bother me. i'm socializing!: A breakpoint-based smartphone notification system. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 541–554. ACM, 2017.

[154] Karl Pearson. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 58:240–242, 1895.

[155] Veljko Pejovic and Mirco Musolesi. InterruptMe: Designing Intelligent Prompting Mechanisms for Pervasive Applications. *UbiComp '14*, 2014.

[156] Veljko Pejovic, Mirco Musolesi, and Abhinav Mehrotra. Investigating the role of task engagement in mobile interruptibility. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 1100–1105. ACM, 2015.

[157] Marianne Graves Petersen, Kim Halskov Madsen, and Arne Kjær. The usability of everyday technology: emerging and fading opportunities. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 9(2):74–105, 2002.

[158] Martin Pielot and Luz Rello. The do not disturb challenge: a day without notifications. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1761–1766. ACM, 2015.

[159] Martin Pielot, Karen Church, and Rodrigo De Oliveira. An in-situ study of mobile phone notifications. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, pages 233–242. ACM, 2014.

[160] Martin Pielot, Rodrigo de Oliveira, Haewoon Kwak, and Nuria Oliver. Didn't you see my message?: predicting attentiveness to mobile instant messages. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3319–3328. ACM, 2014.

[161] Swadhin Pradhan, Lili Qiu, Abhinav Parate, and Kyu-Han Kim. Understanding and managing notifications. In *INFOCOM 2017-IEEE Conference on Computer Communications, IEEE*, pages 1–9. IEEE, 2017.

[162] Harry T Reis and Shelly L Gable. Event-sampling and other methods for studying everyday experience. *Handbook of research methods in social and personality psychology*, 2000.

[163] Robert Remus, Uwe Quasthoff, and Gerhard Heyer. Sentiws-a publicly available german-language resource for sentiment analysis. In *LREC*, 2010.

[164] Brenna N Renn, Abhishek Pratap, David C Atkins, Sean D Mooney, and Patricia A Areán. Smartphone-based passive assessment of mobility in depression: Challenges and opportunities. *Mental Health and Physical Activity*, 2018.

[165] Daniele Riboni and Claudio Bettini. Owl 2 modeling and reasoning with complex human activities. *Pervasive and Mobile Computing*, 7(3):379–395, 2011.

[166] Stephanie Rosenthal, Anind K Dey, and Manuela Veloso. Using decision-theoretic experience sampling to build personalized mobile phone interruption models. In *International Conference on Pervasive Computing*, pages 170–187. Springer, 2011.

[167] Jerome I Rotgans and Henk G Schmidt. How individual interest influences situational interest and how both are related to knowledge acquisition: A microanalytical investigation. *The Journal of Educational Research*, pages 1–11, 2017.

[168] Daniel Rough and Aaron Quigley. Jeeves-a visual programming environment for mobile experience sampling. In *Visual Languages and Human-Centric Computing (VL/HCC), 2015 IEEE Symposium on*, pages 121–129. IEEE, 2015.

[169] MS Ryoo and JK Aggarwal. Stochastic representation and recognition of high-level group activities. *International journal of computer Vision*, 93(2):183–200, 2011.

[170] Alireza Sahami Shirazi, Niels Henze, Tilman Dingler, Martin Pielot, Dominik Weber, and Albrecht Schmidt. Large-scale assessment of mobile notifications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3055–3064. ACM, 2014.

[171] Jeff Sauro and James R. Lewis. Quantifying the user experience : practical statistics for user research, 2012. URL `http://lib.myilibrary.com/detail.asp?id=361223`.

[172] Klaus R Scherer, Vera Shuman, Johnny RJ Fontaine, and Cristina Soriano. The grid meets the wheel: Assessing emotional feeling via self-report. *Components of emotional meaning: A sourcebook*, pages 281–298, 2013.

[173] Ulrich Schiefele. Situational and individual interest. *Handbook of motivation at school*, pages 197–222, 2009.

[174] M Schrepp, S Olschner, and U Schubert. User experience questionnaire benchmark praxiserfahrungen zum einsatz im business-umfeld. *Usability Professionals*, 13, 2013.

[175] Florian Schulze and Georg Groh. Conversational context helps improve mobile notification management. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 518–528. ACM, 2016.

[176] Christie Napa Scollon, Chu-Kim Prieto, and Ed Diener. Experience sampling: promises and pitfalls, strength and weaknesses. In *Assessing well-being*, pages 157–180. Springer, 2009.

[177] Yue Shi, Yuanchun Shi, and Jie Liu. A rotation based method for detecting on-body positions of mobile devices. In *Proceedings of the 13th International Conference on Ubiquitous Computing*, UbiComp '11, pages 559–560, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0630-0. doi: 10.1145/2030112.2030212. URL `http://doi.acm.org/10.1145/2030112.2030212`.

[178] Paul J Silvia. *Exploring the psychology of interest*. Oxford University Press, 2006.

[179] Jeremiah Smith and Naranker Dulay. Ringlearn: Long-term mitigation of disruptive smartphone interruptions. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014 IEEE International Conference on*, pages 27–35. IEEE, 2014.

[180] Arthur A Stone and Saul Shiffman. Ecological momentary assessment (ema) in behavorial medicine. *Annals of Behavioral Medicine*, 1994.

[181] Arthur A Stone, Ronald C Kessler, and Jennifer A Haythomthwatte. Measuring daily events and experiences: Decisions for the researcher. *Journal of personality*, 59(3):575–607, 1991.

[182] Heather L Stuckey et al. Three types of interviews: Qualitative research methods in social health. *Journal of Social Health and Diabetes*, 1(2):56, 2013.

[183] G Henri Ter Hofte. Xensible interruptions from your mobile phone. In *MobileHCI '07*, 2007.

[184] Timothy J Trull and Ulrich Ebner-Priemer. Ambulatory assessment. *Annual review of clinical psychology*, 9:151, 2013.

[185] Timothy J Trull and Ulrich Ebner-Priemer. The role of ambulatory assessment in psychological science. volume 23. SAGE Publications, 2014.

[186] Timothy J Trull and Ulrich W Ebner-Priemer. Using experience sampling methods/ecological momentary assessment (esm/ema) in clinical assessment and clinical research: introduction to the special section. 2009.

[187] Liam D Turner, Stuart M Allen, and Roger M Whitaker. Interruptibility prediction for ubiquitous systems: conventions and new directions from a growing field. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pages 801–812. ACM, 2015.

[188] Gašper Urh and Veljko Pejović. Taskyapp: inferring task engagement via smartphone sensing. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pages 1548–1553. ACM, 2016.

[189] A. Vahdatpour, N. Amini, and M. Sarrafzadeh. On-body device localization for health and medical monitoring applications. In *Pervasive Computing and Communications (PerCom), 2011 IEEE International Conference on*, pages 37–44, March 2011. doi: 10.1109/PERCOM.2011.5767593.

[190] Niels Van Berkel, Jorge Goncalves, Simo Hosio, and Vassilis Kostakos. Gamification of mobile experience sampling improves data quality and quantity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):107, 2017.

[191] Alexandra Voit, Tonja Machulla, Dominik Weber, Valentin Schwind, Stefan Schneegass, and Niels Henze. Exploring notifications in smart home environments. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 942–947. ACM, 2016.

[192] Hsiu-Feng Wang, Pei-Yu Wang, and Hsin-Chung Ting. The influence of highlighting and font size on visual search performance. In *Contemporary Ergonomics and Human Factors 2012: Proceedings of the International Conference on Ergonomics & Human Factors*, pages 91–98, 2012.

[193] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. 2014.

[194] Dominik Weber, Sven Mayer, Alexandra Voit, Rodrigo Ventura Fierro, and Niels Henze. Design guidelines for notifications on smart tvs. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*, pages 13–24. ACM, 2016.

[195] Julie S Weber, Mark W Newman, and Martha E Pollack. Multi-format notifications for multi-tasking. In *IFIP Conference on Human-Computer Interaction*, pages 247–260. Springer, 2009.

[196] Tilo Westermann, Sebastian Möller, and Ina Wechsung. Assessing the relationship between technical affinity, stress and notifications on smartphones. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 652–659. ACM, 2015.

[197] Tilo Westermann, Ina Wechsung, and Sebastian Möller. Smartphone notifications in context: A case study on receptivity by the example of an advertising service. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2355–2361. ACM, 2016.

[198] Jason Wiese, T Scott Saponas, and AJ Brush. Phoneprioception: enabling mobile phones to infer where they are kept. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2157–2166. ACM, 2013.

[199] Zhixian Yan, Jun Yang, and Emmanuel Munguia Tapia. Smartphone bluetooth based social sensing. UbiComp'13 Adjunct. ACM, 2013.

[200] SungHyuk Yoon, Sang-su Lee, Jae-myung Lee, and KunPyo Lee. Understanding notification stress of smartphone messenger app. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pages 1735–1740. ACM, 2014.

[201] Fengpeng Yuan, Xianyi Gao, and Janne Lindqvist. How busy are you?: Predicting the interruptibility intensity of mobile users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 5346–5360. ACM, 2017.

[202] Zhen Yue, Eden Litt, Carrie J Cai, Jeff Stern, Kathy K Baxter, Zhiwei Guan, Nikhil Sharma, and Guangqiang George Zhang. Photographing information needs: the role of photos in experience sampling method-style research. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 1545–1554. ACM, 2014.

[203] Vincent W Zheng, Yu Zheng, Xing Xie, and Qiang Yang. Collaborative location and activity recommendations with gps history data. In *Proceedings of the 19th international conference on World wide web*, pages 1029–1038. ACM, 2010.

[204] Fred RH Zijlstra, Robert A Roe, Anna B Leonora, and Irene Krediet. Temporal factors in mental work: Effects of interrupted activities. *Journal of Occupational and Organizational Psychology*, 72(2):163–185, 1999.

List of Abbreviations

API	Application Programming Interface; An interface that allows one program or application to share information with another one
App	A smartphone application
C4.5	A decision tree algorithm
EMA	Ecological Momentary Analysis; A method for in-situ assessment of user behavior information
ESM	Experience Sampling Method; A method for in-situ assessment of user behavior information
ESMAC	ESM App Configurator; A tool to create and configure ESM apps for Android
FFT	Fast Fourier Transformation; An algorithm to extract frequency components from a digital signal
GPS	Global Positioning System; A method to identify the geolocation of a device
GSM	Global System for Mobile Communications; A standard for mobile communication and digital cellular networks
GUI	Graphical User Interface; Serves as an interface between machine and user and presents information visually
HCI	Human Computer Interaction; A research field in computer science that focuses on the interaction between humans/users and machines/devices
HMM	Hidden Markov Model; A method to model a system in form of a Markov model with hidden states

HTTP	Hypertext Transfer Protocol; A protocol to transfer information in information systems
ID	Identification number; A unique number used to differentiate between persons or objects
IBk	A Weka implementation of the k-nearest neighbors algorithm
J48	A Weka implementation of the C4.5 decision tree algorithm
JSON	JavaScript Object Notation; A human-readable format used to transfer data
KIT	Karlsruhe Institute of Technology; University of Karlsruhe and research institute
LED	Light-Emitting Diode; A light source available on several smartphones that might emit visual cues to indicate incoming notifications
MAC	Media Access Control, often referred to as MAC address Unique identifier of a network adapter
NASA-TLX	NASA Task Load Index; A standardized questionnaire used to evaluate the task load associated with the usage of a system
OS	Operating System; In this thesis often mentioned in the context of "mobile OS", the operating system running on mobile devices
PTC	Position Transition Correction; An algorithm used to correct predicted labels for a sequence of smartphone positions
R	GNU R; A tool for statistical computing and graphics
SMO	Sequential Minimal Optimization; An optimization algorithm for SVMs

SMS	Short Message Service; A communication methods used by mobile phones
SSID	Service Set Identifier; Usually the name of a wireless network
STEM	Science, Technology, Engineering, and Mathematics; English equivalent to the German MINT used to summarize study subjects in natural sciences and technology
SUS	System Usability Scale; A standardized questionnaire used to evaluate the usability of a system
SVM	Support Vector Machine; A classification algorithm
UEQ	User Experience Questionnaire; A standardized questionnaire used to evaluate the user experience after using a system
URL	Uniform Resource Locator; An identifier for a resource such as a website
VFI	Voting Feature Intervals; A classification algorithm
Weka	Waikato Environment for Knowledge Analysis; A data mining tool
WiFi	A technology that allows devices to connect to wireless LAN
WLAN	Wireless LAN; A wireless computer network
XML	Extensible Markup Language; Human-readable markup language usually used for descriptions
XSD	XML Schema definition; Specifies the elements of an XML file