Non-Intrusive and Physiologically Informed Methods and Interfaces for Notification Research

Pascal E. Fortin

Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

September 2021

A thesis submitted to McGill University in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Acknowledgments

The research presented in this thesis was financially supported by a Fonds de Recherche du Québec Nature et Technologies Doctoral Fellowship (FRQNT B2X-211767), McGill Engineering Doctoral Award (MEDA), McGill Engine TechAccelR innovation award and the National Sciences and Engineering Research Council of Canada (NSERC RGPIN-2017-05013).

I would like to thank my supervisor, Prof. Jeremy Cooperstock, for his guidance and support through each stage of the research process. I also want to acknowledge the contributions of my advisors Prof. Ilja Frissen, Frank Ferrie as well as Prof. Stephanie Blain-Moraes for the discussions that helped in inspiring and shaping the research I am presenting in this thesis.

Thanks to the Shared Reality Lab's members for their honest feedback and critique of my work. Our team's interdisciplinary backgrounds and perspectives indubitably allowed me to expand my social and research skills. I want to specifically acknowledge the support offered by my incredible officemates, colleagues and friends, Jeffrey Blum and Antoine Weill-Duflos without who my graduate studies could have been shorter (thank you for the many side projects!..), but definitely not as enriching on a personal and professional level.

Thanks to my parents, France Boily and Marc Fortin, for raising me in a loving and supportive environment where creativity, curiosity and risk taking were always actively encouraged. I also want to acknowledge my grand parents, Ronald Fortin, Jocelyne Théroux, Laurent Boily and Anne Gagnon, who inspired me and taught me the value of knowledge in all its forms. All of you may no longer be physically with us, but you directly shaped how I continue to evolve as a human being, researcher, and how I hope to become as a father.

Finalement, je suis extrêmement reconnaissant envers ma conjointe Violaine pour son amour, son soutien et sa compréhension lors des hauts, mais particulièrement des moments difficiles qui ont accompagné cette longue réalisation.

None of this work would have been possible without McGill University's amazing facilities and student services. Knowing that one of Canada's top academic institutions, and one of Montreal's top employers, systematically cares about your well-being , your future, and protects your interests as a student and employee means the world when facing the uncertainty associated with doing your graduates studies during a global pandemic. To McGill: *"So long and thanks for all the fish"*.

Contents

List of Figures

List of Tables

Abstract

Recent advances in mobile, wearable and telecomunication technologies have encouraged the growth of our desire for constant interconnectedness. Indeed, the ubiquity of electronic devices in our lives is such that we not only rely on them to support our desires for human-human social interaction, but also sometimes, fulfil these needs through interactions with virtual agents. As such, whether it consists of an important phone call or a reminder from your mood tracking app, our devices are constantly fighting to capture our attention, oftentimes causing significant disruption in professional and social contexts, as well as having negative impact on our physical and mental well-being.

This thesis argues that notification systems' deficiencies are largely due to our lack of understanding of users' notification experience in the wild, and that by focusing on users' behavior or relying on questionnaires, current notification research methodologies cannot capture the full picture of smartphone users' experience.

This thesis pursues two mutually reinforcing tracks that advance the notification research and engineering state of the art:

First, it explores the use of psychology concepts and methodologies, and physiological signal processing to the collection of information on users' state and perception of notifications. This is achieved by documenting the impact of notification perception on smartphone users' electrodermal activity, heart rate, heart rate variability and wrist-motion. Based on these physiological changes, a novel interaction-less notification perception prediction system is introduced and evaluated, expanding the breadth and depth of possible notification research methodologies.

Second, it presents two innovative methods allowing for an increase in out-of-the-lab self-reports and physiological signals data collection capabilities. This is enabled by the introduction and validation of a new gestural self-reporting interface, reducing the intrusiveness of subjective data collection, and a wearable tightness estimation system based on the raw optical heart rate signal.

Résumé

Les progrès récents en technologies mobiles, portables et de télécommunication ont favorisé la croissance de nos désirs d'interconnections sociales. En effet, l'omniprésence des appareils électroniques dans nos vies est telle que nous comptons sur eux pour répondre à nos besoins sociaux, mais aussi, parfois, pour satisfaire ces besoins par des interactions avec des agents virtuels. Qu'il s'agisse d'un appel téléphonique important, d'un rappel qu'une rencontre aura lieu ou d'une nouvelle mention "j'aime", nos appareils se battent constamment pour capter notre attention, causant souvent des perturbations importantes dans les contextes professionnels et sociaux, et ayant ultimement un impact négatif sur notre bien-être physique et mental.

Cette thèse soutient que les lacunes des systèmes de notification sont principalement attribuable à un manque de compréhension de l'expérience utilisateurs en matière de notification. Ce dernier est partiellement dû à l'utilisation de méthodologies de recherche qui ne sont pas en mesure de capturer l'ensemble des facettes de l'expérience des notifications.

Cette thèse poursuit deux directions qui se renforcent mutuellement et qui contribuent à faire avancer l'état de l'art de la recherche sur les notifications et en ingénierie:

D'abord, elle explore l'utilisation de concepts et méthodologies de recherche en psychologie, et du traitement des signaux physiologiques pour la collecte d'informations sur l'état interne et la perception des notifications par les utilisateurs. Pour ce faire, elle documente l'impact de la perception des notifications sur certains signaux physiologiques ainsi que le mouvement du poignet des utilisateurs de téléphones intélligents. Sur la base de ces changements physiologiques et mouvements, un nouveau système de prédiction de la perception des notifications est présenté et évalué.

Par la suite, cette thèse présente deux méthodes innovantes permettant d'augmenter les capacités de collecte de données physiologiques et l'utilisation de questionnaires en dehors du laboratoire. Ceci est rendu possible par la présentation et la validation d'une nouvelle interface gestuelle d'auto-administration de micro-enquêtes, réduisant le caractère intrusif de la collecte de données subjectives, et d'un système portable d'estimation de la force de contact entre un capteur de rythme cardiaque et la peau, permettant de mieux contrôller les conditions dans lesquelles les signaux physiologiques sont acquis.

Glossary

Authentication: Process through which a user identifies themselves securely to access a system (e.g., their smartphone). On smartphones, this is typically achieved using a personal identification number (PIN), pattern, password, fingerprint sensor. More modern approaches based on facial identification and iris scanning exist.

Electrodermal activity: A physiological measurement that captures changes in electrical properties of a user's skin (e.g., resistance, conductance, admittance, impedance).

Experience Sampling Methods (ESM): A research methodology that relies on the frequent presentation of questionnaires as participants go about their daily activities. The presentation of a questionnaire can be time triggered (e.g., every hour) or event triggered (e.g., based on geofences, smartphone notification).

Haptic: Refers to the sense of touch.

Heart rate variability (HRV): Physiological measurement interested in the variability of the time between consecutive heart beats.

In situ: Refers to research conducted within a limited subset of ecologically valid conditions.

In-the-wild: Refers to research conducted in *any and all* contexts in which a system is intended to be used.

Notification volume: Expression used in the notification research literature that refers to the number of notifications received during a given period of time (e.g., daily notification volume).

Photoplethysmograph (PPG): Sensing approach that allows the measurement of blood volume changes using light transmitted and reflected through human tissues. This is typically used in optical heart rate and pulse oximetry sensors.

Physiological response: Change observed in one or more physiological signals (e.g., electrodermal activity, heart rate) immediately following and typically resulting from the perception of a stimulus.

Psychophysiology: Refers to the a user's physiological and psychological state, and how they influence one another.

Unlock Journaling: A branch of experience sampling methods (ESM) that exclusively focuses on the collection of subjective data from users during the smartphone unlocking and authentication process.

Chapter 1

Introduction

With the number of smartphones exceeding that of personal computers in many North American households, people have never been as continuously interconnected as they are today. Working from home, finding a colleague you have not spoken to in years or messaging a distant relative can all be done from a hand-sized device, within minutes, 24/7. During the Covid-19 pandemic more than ever, push notifications play a crucial role in ensuring that important communications and virtual social events are brought to users' attention in a timely manner.

While there is little doubt that notifications facilitate effective human-computer and computer-mediated human-human interactions, their current pervasiveness is also known to have significant negative effects on smartphone users' workplace performance and mental well-being. Indeed, the frequent interruptions have been shown to negatively impact productivity, attention and the ability to focus [1]. In addition, the unpredictable outcome and presentation times of notifications have been theorized to contribute to the reinforcement of problematic smartphone usage patterns, frequently associated with significant disruption of users' ability to engage in their usual daily activities and social interactions [2, 3].

The ubiquitous nature of notifications and the early investigations of their impact on users have motivated researchers to further study the factors that shape users' notification experience. To do so, two main methodologies are typically employed. The first relies on the background observation of notification interactions to study participants' behaviour, while the second employs self-reporting interfaces to gather insights into attitudinal components of users' notification experience. Even though both of these methodologies enable the generation of meaningful findings on usage habits and the impact of notifications on users' mental state, they suffer from limitations that constrain the breadth and depth of the research questions they allow to explore. Most notably, by focusing exclusively on users' behavior, the first does not provide insights into whether or how a notification was perceived. The second does provide insights into subjective components of the notification experience, but does so using supplementary notifications associated with questionnaires, which may further introduce biases in the collected data.

1.1 Scope

This thesis acknowledges that the experience of a notification is multifaceted and argues that with a more thorough understanding of their users' context and internal states, smartphone manufacturers and researchers could design communication technologies that are better adapted to their users' device usage. This should hopefully lead to more effective, less disruptive technologies that contribute to a balanced digital life. Towards this objective, the current notification research methodologies need to be expanded to consider users' experience from different, more nuanced perspectives than is currently possible using exclusively behavioral- and/or questionnaire-based approaches. This thesis therefore proposes the application of concepts from psychology, advanced wearable sensing and most importantly, physiological signal processing, to the field of human-computer interaction, and reports on original work that expands existing notification research methodologies and aims to respond to the following broad questions:

- What, if any, are the effects of smartphone notifications on users' physiological signals?
- Are these responses sufficiently reliable and granular to allow practical inferences to be made about users' perception of notifications?
- How can current technical challenges with regards to the collection of high quality physiological signals in the wild be overcome?
- Considering the performance of psychophysiological inferences, what alternative methods are available for our devices to reliably collect subjective information from a user? How can such methods be adapted to minimize their intrusiveness?

1.2 Overview

This thesis presents answers to the above-mentioned questions through the introduction of original findings enabling, or enabled by the use of novel systems and interfaces. This overview summarizes the content of each chapter.

Chapter 2 presents an overview of the existing relevant literature on the topics of notifications, notification perception, notification research methodologies and psychophysiology.

Chapter 3 introduces Sweatsponse, a novel smartphone notification perception prediction system based on electrodermal activity. In addition to being the first method that allows the post-stimulation confirmation of notification perception without user interaction, results from this laboratory study demonstrate for the first time that smartphone notifications reliably induce skin conductance responses, supporting years of psychological research on the topic.

Chapter 4 extends the work presented in Chapter 3 by evaluating the impact of notification perception on heart rate, heart rate variability and wrist motion *outside* of laboratory conditions. Beyond documenting additional physiological and behavioral responses *in situ*, their benefits to notification perception prediction performance are quantified.

Chapter 5 presents a proof of concept for TightHR, a technique that allows estimating the force applied between an optical heart rate sensor (PPG) and a user's skin using PPG signal properties. Such a contact force estimation enables the collection of data under repeatable and reliable coupling conditions, which are anticipated to result in higher quality physiological signals and richer psychophysiological inferences.

Chapter 6 extends the existing unlock journaling field with a new self-reporting mechanism based on fingerprint sensor gestures with the intent to reduce the intrusiveness of existing self-reporting interfaces. By comparing this technique with state of the art unlock journaling methods, we demonstrate our approach's reporting performance. It is the first method to adapt unlock journaling to the increasingly used fingerprint authentication mechanism, reducing the intrusiveness of self-reporting interfaces for users of this authentication method.

Chapter 7 discusses how the systems and findings introduced in this thesis advance the state of the art in notification research and are envisioned to be impactful in other application domains. A brief presentation of practical challenges associated with conducting *in situ* physiologically informed notification research and how they were approached within the scope of this thesis is made to serve as practical recommendations for researchers interested in following this promising new research direction.

Chapter 8 concludes the thesis by reiterating its main findings and presenting promising future research directions enabled by its findings and systems.

Note: Chapters 3 to 7 consist of published or soon to be published manuscripts. A statement on each co-author's contribution is presented in their respective prefaces.

1.3 Summary of Contributions

All elements of this thesis are original scholarship and contribute to the advancement of the state of the art in the fields of notification research, human-computer interaction and engineering.

This thesis makes the following methodological, fundamental and technical contributions:

- A novel notification research methodology combining physiological signal monitoring with passive mobile interaction logging, allowing for the investigation of novel research questions related to notifications' impact on smartphone users' psychophysiological state in and outside of laboratory contexts.

- Evidence that the perception of a notification *in situ* has a direct impact on smartphone users' electrodermal activity, heart rate, heart rate variability and wrist motion, demonstrating the promise of the proposed methodology and partially supporting existing research on the arousing and interruptive nature of notifications.

- The introduction and evaluation of a technique that harnesses these newly discovered changes in physiological signals to confirm whether a notification was perceived by a user after its presentation, without the need for users to engage with the notification, notification tray, application, smartphone or any other device (e.g., personal computer).

- The introduction and proof of concept of a novel technique that employs raw optical heart rate signal properties to estimate the contact force between the sensor and a user's skin. Thanks to its reliance on a sensor that is already integrated in millions of consumer and medical devices, this technique offers a unique opportunity to acquire coupling information crucial to the collection of high quality physiological measurements without the need for dedicated force sensing equipment.

- An extension of unlock journaling techniques adapted for fingerprint sensor authentication users. A demonstration that the proposed interface performs at least as well as current state of the art unlock journaling methods while being perceived as significantly less intrusive, highlighting the benefits of data collection instruments that are consistent with participants' smartphone usage habits.

References

[1] K. Kushlev, J. Proulx, and E. 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*, ser. CHI '16. New York, NY, USA: ACM, 2016, pp. 1011–1020. [Online]. Available: http://doi.acm.org/10.1145/2858036.2858359

[2] S. P. L. Veissière and M. Stendel, "Hypernatural Monitoring: A Social Rehearsal Account of Smartphone Addiction," *Frontiers in Psychology*, vol. 9, p. 141, 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2018.00141

[3] Y.-H. Lin, C.-L. Chiang, P.-H. Lin, L.-R. Chang, C.-H. Ko, Y.-H. Lee, and S.-H. Lin, "Proposed Diagnostic Criteria for Smartphone Addiction," *PloS one*, vol. 11, no. 11, pp. e0 163 010–e0 163 010, Nov. 2016. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/27846211https://www.ncbi.nlm.nih.gov/pmc/PMC5112893/

Chapter 2

Background

This review chapter lays the foundation for this thesis by presenting prior research on notifications, notification perception, notification research methodologies and psychophysiology. Additional related work is introduced in later chapters when directly relevant to the topics discussed.

2.1 Notifications

2.1.1 Overview of Notifications

A push notification consists of a usually short stimulus and is intended to signal the presence of potentially relevant information to the user of a system. On modern mobile technologies, a notification consists of a vibrotactile, auditory and/or a visual alert delivered from the user's smartphone or other wearable system to attempt to capture the user's attention. The content varies based on its originating application, but can range anywhere from an urgent instant message or email, to a game reminding its players that they have not played in the last week. As of the writing of this thesis, smartphone operating systems do not include an intelligent notification filtering system by default, implying that given the right system permissions, any application can generate as many notifications as it wishes, at any time and for any reason. Users are left with the responsibility to manually modify configurations of each application to prevent it from delivering notifications, or to use their device's ringer mode setting to globally select whether they wish to be notified, and using what modality. Westermann et al. have found that only approximately 10% of users will actually modify application-specific notification settings [1]. This suggests that most smartphone users are either receiving notifications from all installed applications, or no notifications at all due to the selection of “silent” as their device’s global ringer mode.

In a study on smartphone notifications involving 278 mobile phone users, Pielot et al. reported that participants received a median number of 56 notifications per day [2]. This is aligned with prior literature that reported participants being presented with 45 to 63 notifications per day on average [3, 4] and a survey-based study that found that the majority of participants reported receiving 20-50 or 50-100 notifications per day [5]. Contrary to popular belief, as of 2018, the number of notifications delivered per day for average users did not vary significantly over the previous five years [2]. It would be particularly interesting to see whether the shift to working from home has modified that trend in the last year.

Beyond the daily number of alerts, the time at which they are presented and the application that generated them might have a significant impact on users’ perception and receptivity. Pielot et al. reported that the majority of notifications were presented between 6:00 and 24:00, with very few alerts being delivered between 3:00 and 6:00 [4]. More than three quarters of their participants’ notifications were coming from direct communication applications (personal and group conversations, and emails), with the remainder originating from social media and other applications (e.g., news, games). In their study, while approximately 65% of instant message notifications were consumed within 30 minutes, those associated with emails, social networks and other applications were attended in 15.47, 26.55 and 16.19% of cases respectively. This is in agreement with findings presented by Shirazi et al. who noted that participants attribute greater importance to notifications originating from messaging applications, events (e.g., calendar) or that provide information about their context and social circle [3] than to other sources of information.

All smartphone users do not interact with their notifications in the same manner. In a study based on 3953 Android users, Weber et al. identified three main approaches to notification handling [6]. The *frequent cleaners* attend to all notifications as soon as possible to keep the notification tray empty. The *notification regulators* tolerate a higher number of notifications in their notification tray than *frequent cleaners*, but do not let the number get too high. Finally, the *notification hoarders* let notifications accumulate to more than two to three times that of *notification regulators*. They remove each item once it has been attended to or clear all notifications in bulk after an extended period of time before starting a new accumulation cycle. The authors note that most smartphone users’ behavior is consistent with the first two categories and hypothesize that *notification hoarders* have given up on taking control of their notifications.

2.1.2 Notifications and Well-being

Interruptions caused by notifications are numerous and have been shown to have detrimental effects on mental well-being [7, 8], which impacts workplace performance (e.g., attention, productivity) [7]. Indeed, Iqbal and Bailey have shown that interruptions caused by emails result in significant increases in users' self-reported frustration levels [9]. Similarly, task interruptions have been found to significantly increase the participants' stress, perceived efforts and the time pressure experienced while finishing a task [4, 10]. In a user study based on a controlled laboratory task, Stothart et al. have shown that the perception of a smartphone notification increased risks of making mistakes as much as actually reading the message or engaging in a phone call [11]. Similarly and unsurprisingly, Bailey and Konstan have reported that the interruption of a primary task would result in a longer completion time and more errors being committed by participants [12]. Kushlev et al. reported results from a study in which 221 subjects were asked to turn on all possible notifications from their devices for a week, followed by a week where notifications were set to silent and phones placed out of direct access (e.g., pocket, bag) [7]. Their results suggested that turning on all notifications resulted in significantly higher levels of inattention and hyperactivity accompanied by lower levels of perceived productivity, social connectedness, mastery over their environment, meaning in life and choice over their actions than during the week of limited access. Interested in what specific properties of notifications were responsible for some of these negative consequences, Yoon et al. identified three main categories of notification-related stress: physical notification, message content and responsiveness stress. Of significant interest to this thesis, physical notification stress, related to the notification stimuli themselves, the context in which they are presented and the frequency of their presentation, was identified as one of the most important stressors for their participants [13].

In addition to these immediate negative effects on well-being and task performance, smartphone notifications are thought to be playing a significant role in the reinforcement of problematic smartphone usage patterns, a condition sharing many symptoms with traditional behavioral addictions such as gambling, a condition with long-lasting consequences. While smartphone addiction is not yet part of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), Lin et al. proposed a set of criteria and evaluated their diagnostic accuracy. The behavioral criteria that were found to allow the greatest diagnostic accuracy are: a continued inability to resist the impulse to use the smartphone, symptoms of dysphoria, anxiety or irritability after a period of withdrawal, using the smartphone for a period longer than intended, and persistent desire and/or unsuccessful attempts to quit or reduce smartphone use, heightened attention to using or quitting smartphone use, persistent smartphone use despite recurrent physical or psychological consequences. They also identified four functional criteria used in the diagnostic: excessive use resulting in persistent or recurrent physical or psychological problems, use in a physically hazardous situation (e.g., driving) or situations that have other negative impacts on daily life, use that impairs social relationships or performance at school or work, and finally, use that is very time-consuming or causes significant distress [14].

2.2 Notification Research

Due to their ubiquity and serious impact on mental well-being and workplace performance, notifications have started receiving significant research attention. Current notification research typically relies on one or a combination of the two main notification research methodologies: the passive approach, which relies on instrumenting participants' smartphones and observing their notification interactions and behavior, and the active approach, which instead probes participants for their perception of their notification using self-reporting instruments. For a full discussion of the two methodologies' advantages and limitations, the reader is referred to Section 4.1.1 and 4.1.2 of Chapter 4.

2.3 Context-Aware Notification Systems

Concerns about the negative impact of push notifications have been shared within the research community since their introduction in commercial devices. As such, a significant amount of literature on techniques to attempt to reduce the interruption burden of notifications exists.This section does not serve as an exhaustive review of existing context-aware notification systems. It instead introduces the reader to representative examples of context-aware notification frameworks relying on different sensing approaches and boasting various levels of system complexity.

Existing context-aware notification systems attempt to tackle the problem from two main angles. The first focuses on delaying the presentation of smartphone notifications until a time when the interruption burden will be minimized. Mehrotra et al. expanded on existing methods relying solely on contextual data [15, 16], by integrating characteristics of the notification itself (e.g., source application, application category, sender's relationship to the recipient) to the notification management system's prediction of whether it is an opportune moment to interrupt the user [17]. The five features that were found to be the strongest predictor of interruptibility were the source application, notification category (e.g., messaging, email, social network), phone status, user location and time of arrival. However, the reported performance of their predictors could be improved, with a sensitivity varying between approximately 40 and 85% and specificity ranging from 65 to 90%. Taking an approach less computationally heavy, and more robust, Fischer et al. used the endings of calls and SMS as indicators of opportune moments to deliver notifications [18]. The rationale is that the ending of phone activities inherently consists of breakpoints in the user's action sequence. Those breakpoints have been shown to be moments when the interruption burden is the lowest [12, 19–22]. Their results suggested that delivering notifications at the end of a phone call or submission of an SMS, rather than randomly presenting the alerts, resulted in significantly higher probability that they would be addressed immediately by users. Nevertheless, the analysis of self-reported data on the perceived appropriateness and burden imposed by either approach revealed significant differences, outlining the frequent contradictions between quantitative and qualitative measures.

The second angle on the use of intelligent notification systems employs contextual cues to modify the device's ringer mode and intensity with the objective to maximize perception, and minimize the risks of causing unnecessary disruption. By doing so, instead of delaying the delivery of an alert, the system attempts to modify the properties of the notification such that it is adapted to the user's current context. Blum et al. proposed to predict perception of a haptic stimulus by quantifying the amount of mechanical noise at the site of delivery using the built-in accelerometer of a smartwatch [23]. They demonstrated that activities involving higher vibration exposure (e.g., biking, running, etc.), required the presentation of more intense vibrations to be perceptually equivalent to that of low vibration exposure activities. They proposed using these results in a system that would automatically adapt a vibrotactile signal's intensity such that it would be delivered at the minimal strength at which it would be perceived. Achieving similar ends for auditory notifications and ringtones, *Smart Volume*,[1] an Android application available on Google Play store, allows the automatic adaptation of the device ringer's volume based on ambient sound. A number of applications already widely available allow simple contextual adaptation of devices' ringer mode and volume based on smartphone position and ambient data. *Smart Ring Control*[2] allows the control of the ringer mode based on device orientation (e.g., vertical, horizontal, face

[1] https://play.google.com/store/apps/details?id=com.gmail.at.mhassegawa.smartvolume
[2] https://play.google.com/store/apps/details?id=com.shumoapp.smartringcontrol

up, face down). Other more general applications such as *IFTTT*[3] (If this then that), *Automate*[4], and *Tasker*[5] allow users to automate actions or change settings based on contextual triggers. For example, a user can configure these applications such that upon their arrival at their workplace, their ringer mode is set to "vibration only" and automatically changes to "do not disturb" during meetings based on their calendar.

Of significant importance to this thesis, existing adaptive notification frameworks and applications have limited their input space to the immediate environment of the user, as quantified by their device's numerous sensors, including its device-specific user configuration and properties of the notifications themselves. These inputs only offer indirect insights into the user's internal state. In addition, by focusing exclusively on these contextual cues *before* the presentation of a notification or signal, they do not have the ability to confirm whether the signal was perceived. With recent advances in wearable physiological sensing technologies, it is anticipated that including physiological signals, and inferred psychological states made from those signals will significantly improve adaptive systems' performance by not only accounting for the user's external context and the nature of the communication, but also the user's internal state.

2.4 Psychophysiology

Psychophysiology is a broad discipline that investigates the complex relationship between people's psychological and physiological states. This field of research not only studies how users' psychology influences their physiology, but also how changes in their physiology reflect themselves on their users' internal state. A unique possibility offered by the psychophysiological approach is the acquisition of information about users' cognitive (e.g., attention, engagement, perception) and affective state via the observation of changes in their physiological signals, as opposed to relying on self-reports. When doing so is possible, researchers can deploy significantly less intrusive data collection protocols which reduce risks of biases introduced by questionnaires and self-reporting interfaces.

In engineering and human-computer interactions, the psychophysiological approach is typically applied to the development of techniques and frameworks allowing for the adaptation of systems based on their users' physiological and ultimately psychological state. For example, the

[3]https://play.google.com/store/apps/details?id=com.ifttt.ifttt
[4]https://play.google.com/store/apps/details?id=com.llamalab.automate
[5]https://play.google.com/store/apps/details?id=net.dinglisch.android.taskerm

field of affective computing is specifically interested in systems whose behavior is influenced, influences or is otherwise cognizant of participants' affective state. In its simplest form, this can consist of the development of new techniques to recognize users' emotions based on changes in their physiology. However, more complex systems proposing the real time adaptation of video games based on participants' emotional experience have also been proposed [24].

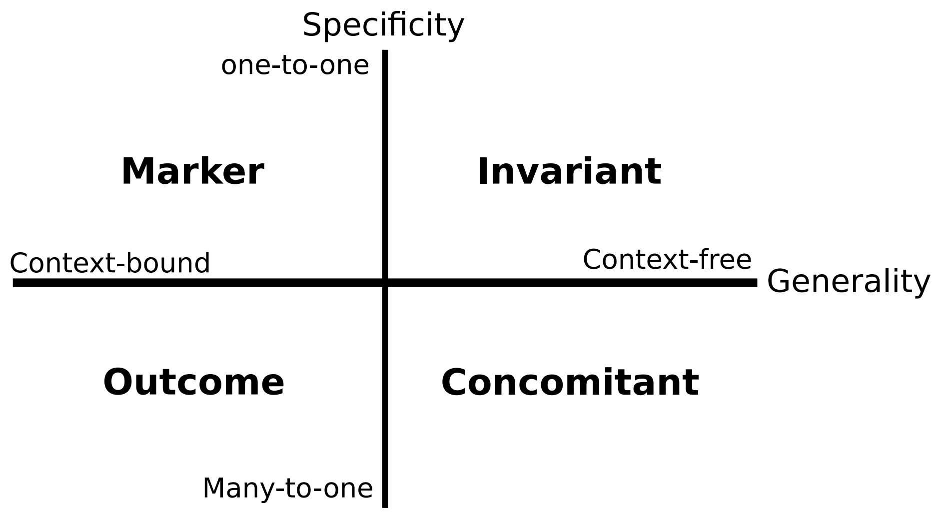


Fig. 2.1 A recreation of Cacioppo's major dimensions of psychophysiological relations and main classes.

While these systems have been relatively successful in controlled laboratory environments and to some extent *in situ*, they remain affected by significant issues that limit their applicability in real world settings [25]. One of the main challenges that this field of research faces is the difficulty in making practically meaningful psychophysiological inferences in ecologically valid scenarios [26]. Indeed, since the vast majority of the psychophysiological literature is based on measurements achieved in controlled laboratory conditions, little is known about how participants' living environment, social interactions and activities interact with psychological constructs of interest. Further exacerbating these issues is the absence of clear guidelines or tools to assist non-expert users in adequately positioning sensors on their body. This results in inconsistent signal quality that negatively impacts system performance.

Assuming perfect data collection conditions, establishing practically meaningful psychophysiological inferences remains challenging. Indeed, changes in biosignals are often interrelated via different physiological processes (e.g., respiratory sinus arrhythmia) which increases the complexity of their interpretations. In addition, a given psychological or cognitive state may impact more than one physiological signal and vice versa. Figure 2.1 presents the major dimensions and classes of psychophysiological relations [27]. The psychophysiological relation class that is the most sought after in practical systems is the *invariant*, where one value, variation or pattern of a signal corresponds to, and only to, a single psychological construct, in any context [26]. However, in practice, given a thorough knowledge of the context in which a subject is evolving, *outcome* relations can also result in practically meaningful psychological inferences [28–30].

This thesis argues that affective and physiological computing approaches could be used to meaningfully expand the breadth and depth of notification research questions by offering insights into participants' internal state with minimal reliance on self-reporting interfaces. It is anticipated that knowing exactly when a notification is being presented will allow us to exclude contextual confounds and focus our analysis on notification-induced physiological responses.

References

[1] T. Westermann, S. Möller, and I. 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*, ser. MobileHCI '15. New York, NY, USA: Association for Computing Machinery, 2015, pp. 652–659. [Online]. Available: https://doi.org/10.1145/2786567.2793684

[2] M. Pielot, A. Vradi, and S. Park, "Dismissed!: A Detailed Exploration of How Mobile Phone Users Handle Push Notifications," in *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '18. New York, NY, USA: ACM, 2018, pp. 3:1—-3:11. [Online]. Available: http://doi.acm.org/10.1145/3229434.3229445

[3] A. Sahami Shirazi, N. Henze, T. Dingler, M. Pielot, D. Weber, and A. Schmidt, "Large-scale Assessment of Mobile Notifications," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: ACM, 2014, pp. 3055–3064. [Online]. Available: http://doi.acm.org/10.1145/2556288.2557189

[4] M. Pielot, K. Church, and R. 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*, ser. MobileHCI '14. New York, NY, USA: ACM, 2014, pp. 233–242. [Online]. Available: http://doi.acm.org/10.1145/2628363.2628364

[5] J. A. Gallud and R. Tesoriero, "Smartphone Notifications: A Study on the Sound to Soundless Tendency," in *Proceedings of the 17th International Conference on*

Human-Computer Interaction with Mobile Devices and Services Adjunct, ser. MobileHCI '15. New York, NY, USA: Association for Computing Machinery, 2015, pp. 819–824. [Online]. Available: https://doi.org/10.1145/2786567.2793706

[6] D. Weber, A. Voit, and N. Henze, "Clear All: A Large-Scale Observational Study on Mobile Notification Drawers," in *Proceedings of Mensch Und Computer 2019*, ser. MuC'19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 361–372. [Online]. Available: https://doi.org/10.1145/3340764.3340765

[7] K. Kushlev, J. Proulx, and E. 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*, ser. CHI '16. New York, NY, USA: ACM, 2016, pp. 1011–1020. [Online]. Available: http://doi.acm.org/10.1145/2858036.2858359

[8] E. Kanjo, D. J. Kuss, and C. S. Ang, "NotiMind: Utilizing Responses to Smart Phone Notifications as Affective Sensors," *IEEE Access*, vol. 5, pp. 22 023–22 035, 2017.

[9] S. T. Iqbal and B. P. Bailey, "Effects of Intelligent Notification Management on Users and Their Tasks," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '08. New York, NY, USA: ACM, 2008, pp. 93–102. [Online]. Available: http://doi.acm.org/10.1145/1357054.1357070

[10] G. Mark, D. Gudith, and U. Klocke, "The Cost of Interrupted Work: More Speed and Stress," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '08. New York, NY, USA: ACM, 2008, pp. 107–110. [Online]. Available: http://doi.acm.org/10.1145/1357054.1357072

[11] C. Stothart, A. Mitchum, and C. Yehnert, "The attentional cost of receiving a cell phone notification." Stothart, Cary: Department of Psychology, Florida State University, 1107 West Call Street, Tallahassee, FL, US, 32306-4301, stothartpsy.fsu.edu, pp. 893–897, 2015.

[12] B. P. Bailey and J. 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*, vol. 22, no. 4, pp. 685–708, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S074756320500107X

[13] S. Yoon and K.-p. Lee, "A Study on Notification System Design of Smartphone Messenger Considering the User ' s Stress," *Archives of design research*, vol. 28, no. 2, pp. 77–91, 2015.

[14] Y.-H. Lin, C.-L. Chiang, P.-H. Lin, L.-R. Chang, C.-H. Ko, Y.-H. Lee, and S.-H. Lin, "Proposed Diagnostic Criteria for Smartphone Addiction," *PloS one*, vol. 11, no. 11,

pp. e0 163 010–e0 163 010, Nov. 2016. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/27846211https://www.ncbi.nlm.nih.gov/pmc/PMC5112893/

[15] V. Pejovic and M. Musolesi, “InterruptMe: Designing Intelligent Prompting Mechanisms for Pervasive Applications,” in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp ’14. New York, NY, USA: ACM, 2014, pp. 897–908. [Online]. Available: http://doi.acm.org/10.1145/2632048.2632062

[16] J. Ho and S. S. Intille, “Using Context-aware Computing to Reduce the Perceived Burden of Interruptions from Mobile Devices,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’05. New York, NY, USA: ACM, 2005, pp. 909–918. [Online]. Available: http://doi.acm.org/10.1145/1054972.1055100

[17] A. Mehrotra, M. Musolesi, R. Hendley, and V. Pejovic, “Designing content-driven intelligent notification mechanisms for mobile applications,” in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp ’15)*, ser. UbiComp ’15. New York, NY, USA: ACM, 2015, pp. 813–824. [Online]. Available: http://doi.acm.org/10.1145/2750858.2807544

[18] J. E. Fischer, C. Greenhalgh, and S. 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*, ser. MobileHCI ’11. New York, NY, USA: ACM, 2011, pp. 181–190. [Online]. Available: http://doi.acm.org/10.1145/2037373.2037402

[19] P. D. Adamczyk and B. 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*, ser. CHI ’04. New York, NY, USA: ACM, 2004, pp. 271–278. [Online]. Available: http://doi.acm.org/10.1145/985692.985727

[20] V. Kostov, T. Tajima, E. Naito, and J. Ozawa, “Analysis of appropriate timing for information notification based on indoor user’s location transition,” in *Fourth Annual IEEE International Conference on Pervasive Computing and Communications (PERCOM’06)*, 2006, pp. 6 pp.–250.

[21] S. T. Iqbal and B. P. Bailey, “Leveraging Characteristics of Task Structure to Predict the Cost of Interruption,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’06. New York, NY, USA: ACM, 2006, pp. 741–750. [Online]. Available: http://doi.acm.org/10.1145/1124772.1124882

[22] T. Tanaka and K. Fujita, “Study of User Interruptibility Estimation Based on Focused Application Switching,” in *Proceedings of the ACM 2011 Conference on Computer*

Supported Cooperative Work, ser. CSCW ’11. New York, NY, USA: ACM, 2011, pp. 721–724. [Online]. Available: http://doi.acm.org/10.1145/1958824.1958954

[23] J. R. Blum, I. Frissen, and J. R. Cooperstock, “Improving Haptic Feedback on Wearable Devices through Accelerometer Measurements,” in *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST ’15)*. New York, New York, USA: ACM Press, 2015, pp. 31–36. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2807442.2807474

[24] A. Fortin-Côté, N. Beaudoin-Gagnon, C. Chamberland, F. Desbiens, L. Lefebvre, J. Bergeron, A. Campeau-Lecours, S. Tremblay, and P. L. Jackson, “FUNii: The Physio-Behavioural Adaptive Video Game BT - Augmented Cognition,” D. D. Schmorrow and C. M. Fidopiastis, Eds. Cham: Springer International Publishing, 2019, pp. 14–28.

[25] F. Larradet, R. Niewiadomski, G. Barresi, D. G. Caldwell, and L. S. Mattos, “Toward Emotion Recognition From Physiological Signals in the Wild: Approaching the Methodological Issues in Real-Life Data Collection,” *Frontiers in Psychology*, vol. 11, p. 1111, 2020. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2020.01111

[26] S. H. Fairclough, “Fundamentals of physiological computing,” *Interacting with Computers*, vol. 21, no. 1-2, pp. 133–145, 2009. [Online]. Available: http://dx.doi.org/10.1016/j.intcom.2008.10.011

[27] J. T. Cacioppo and L. G. Tassinary, “Inferring psychological significance from physiological signals.” *American psychologist*, vol. 45, no. 1, p. 16, 1990.

[28] A. T. Pope, E. H. Bogart, and D. S. Bartolome, “Biocybernetic system evaluates indices of operator engagement in automated task,” *Biological Psychology*, vol. 40, no. 1-2, pp. 187–195, 1995.

[29] R. W. Picard, S. Papert, W. Bender, B. Blumberg, C. Breazeal, D. Cavallo, T. Machover, M. Resnick, D. Roy, and C. Strohecker, “Affective Learning – A Manifesto,” *BT Technology Journal*, vol. 22, no. 4, pp. 253–269, 2004. [Online]. Available: https://doi.org/10.1023/B:BTTJ.0000047603.37042.33

[30] K. M. Gilleade and A. Dix, “Using Frustration in the Design of Adaptive Videogames,” in *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, ser. ACE ’04. New York, NY, USA: ACM, 2004, pp. 228–232. [Online]. Available: http://doi.acm.org/10.1145/1067343.1067372

Chapter 3

Detecting Perception of Smartphone Notifications using Skin Conductance Responses

Fortin, P. E., Sulmont, E., & Cooperstock, J. R. (2019). Detecting Perception of Smartphone Notifications using Skin Conductance Responses. In Conference on Human Factors in Computing Systems, CHI ' 2019 (p. 9). New-Orleans, LA: ACM.

Preface to Chapter 3:

This chapter introduces SweatSponse, a system that uses a wearable skin conductance sensor to make predictions as to whether an auditory or vibrotactile notification that was just delivered was perceived by smartphone users, without the need for them to explicitly engage with their device. The objective of this system is to grant notification researchers and electronic devices with a better awareness of their user's perception, which in turn should allow for a more nuanced analysis of participants' experience. In addition, the perception feedback afforded by SweatSponse could lead to a reduction in disruptiveness of smartphone notifications through the automatic adaptation of alert presentation behavior.

More broadly, the user study presented not only shows the feasibility of the proposed physiologically adaptive notification system in laboratory, but also reports the first evidence of physiological responses to smartphone notifications. While notifications reliably induced skin conductance responses, the responses were found to be significantly larger for vibrotactile than for auditory notifications. In addition, unlike arbitrary stimuli employed in prior work, participants' responses to their own notifications do not seem to, or minimally, habituate over time. This is hypothesized to be due to the social relevance of the signal.

Contributions of Authors:

Pascal E. Fortin was the primary author and contributor to the TechAccelR innovation award partially funding this work, was responsible for the ideation, conceptual and technical framework that was required for data collection and analysis. He wrote all sections of the paper. Elisabeth Sulmont edited the manuscript and contributed to the implementation of the data parsing and time synchronization code. Prof. Jeremy R. Cooperstock edited the manuscript, contributed to the TechAccelR innovation award and supervised the research at a high level.

This work was awarded a "Best of CHI Honorable Mention" (top 5% of accepted papers).

Abstract

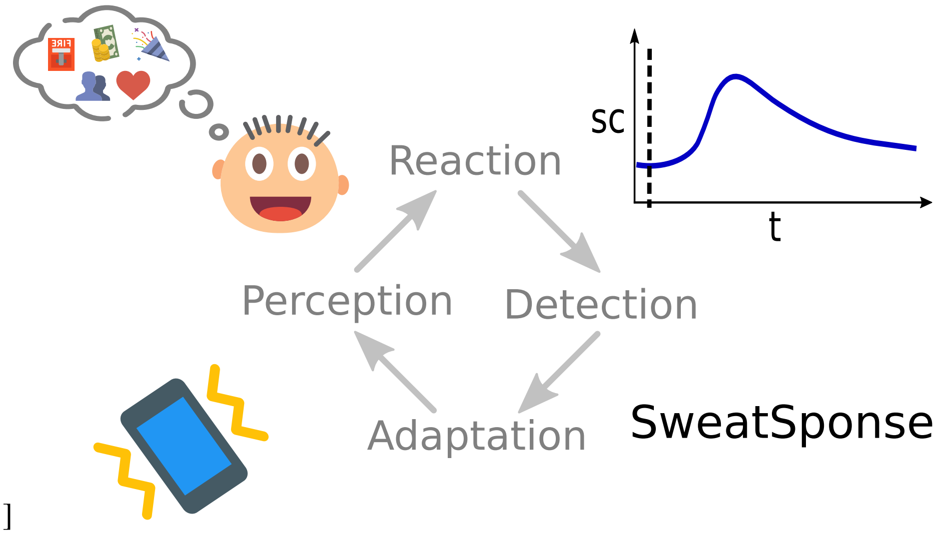
Today's smartphone notification systems are incapable of determining whether a notification has been successfully perceived without explicit interaction from the user. If the system incorrectly assumes that a notification has not been perceived, it may repeat it redundantly, disrupting the user and others (e.g., phone ringing). Or, if it incorrectly assumes that a notification was perceived, and therefore fails to repeat it, the notification will be missed altogether (e.g., text message). Results from a laboratory study confirm, for the first time, that both vibrotactile and auditory smartphone notifications induce skin conductance responses (SCR), that the induced responses differ from that of arbitrary stimuli, and that they could be employed to predict perception of smartphone notifications after their presentation using wearable sensors.

3.1 Introduction

Although intelligent devices are increasingly embedded into our daily lives, in most cases, their delivery of notifications operates in an open loop framework. This leads to inefficient and potentially disruptive communication approaches, as seen in both synchronous and asynchronous contexts. In the former (e.g., phone calls and videoconferencing), alerts are repeated until acknowledged by being explicitly addressed or silenced. During the interval between the user's initial perception of the alert and then reaching their device, the continuous ringing has the potential to cause unnecessary disruption of colleagues or nearby strangers. In the context of asynchronous interactions (e.g., text messaging, email and instant messaging applications), a single alert is delivered. The user is often never reminded of the event, which can delay the response to potentially critical messages.

We believe that a notification should be presented at the minimal volume or vibration intensity necessary for perception, and only be repeated as needed to reduce disruption to a user's environment, whether social or professional. With recent advances in wearable technologies, especially in the domain of wellness and physiological sensing, systems now have access to information about their users' internal states and context that we anticipate enables the possibility of closing the loop on notification delivery.

In this paper, we report the first evidence of physiological responses to smartphone notifications and present supporting evidence that those responses could be used to improve the notification experience if carefully integrated into a perception prediction system. Drawing from these



]

Fig. 3.1 Overview of the proposed feedback loop. A notification is perceived by a user. The user anticipates the potential rewarding social interaction, which induces an SCR. SweatSponse captures the SCR using a wearable sensor, predicts whether it was perceived and feeds the information back to the notification system.

findings, we introduce SweatSponse, a feedback loop relying on skin conductance responses (SCR) that could allow systems to infer a user's perception of a vibrotactile or auditory notification following its presentation, without explicit intervention from the user. While it is still an early prototype, we envision that in the future this feedback channel could allow a notification system to adapt its communication behavior approach by silencing, repeating, or otherwise modifying the sensory characteristics of a notification based on the user's perception (See Figure 3.1).

3.2 Related work

3.2.1 Determining Notification Perception

Two main approaches exist in determining a user's perception of a notification or other vibrotactile or auditory stimulus. The first approach is the standard for today's devices and relies on explicit user interaction. Perception is assumed after a user manually acknowledges the item in the notification tray or opens the application and/or conversation that generated the notification [1–3]. In this active approach, depending on system-specific implementations, failure from a user to acknowledge a notification in a timely manner can lead to the repeated rendering of an alert (e.g., the continuous ringing of an incoming call). Their limitation is evident when a user *perceives* a notification, but is unable, or does not wish, to respond at that moment. In this case, the system falsely interprets the lack of direct interaction with the application or notification tray as a failure in alerting the user of a notification.

The second approach attempts to predict if a signal will be perceived prior to its delivery based on user context attributes and properties of the stimulus. For example, Andersen et al. successfully used users' age, current activity and vibration intensity to predict the probability that a vibrotactile signal would be perceived [4]. One of the limitations outlined by the authors was the impractical reliance on a discrete activity recognition system. Blum et al. addressed this issue by using aggregated continuous accelerometer measurements to represent the amount of haptic noise prior to the delivery of the vibrotactile stimulus, instead of using discrete activity, to predict the likelihood of perceiving the stimulus [5]. While these examples theoretically allow the adjustment of stimulus properties to maximize perception, they are employed prior to the presentation of the notification and as such cannot confirm that the stimulus was actually perceived.

To the best of the authors' knowledge there currently exists no method that can automatically confirm perception *after* the delivery of a stimulus that does not require users to interact with their devices.

3.2.2 Skin Conductance Responses to Notifications

Since the middle of the twentieth century, electrodermal activity has been employed as a robust indicator of a subject's perception of novel, startling, aversive or otherwise significant stimuli [6]. In these contexts, a change in skin conductance in response to a specific event or stimulus is called an event-related skin conductance responses (eSCR). An eSCR is characterized by a sharp increase in skin conductance beginning one to four seconds after the presentation of a stimulus [7], followed by a slow decrease until the baseline is reached.

Physiological responses to arbitrary auditory, visual, and vibrotractile stimuli have been studied extensively and these signals' capacity to induce identifiable eSCR has been demonstrated on hundreds of occasions. However, the relation between smartphone notifications and physiological signals remains largely unexplored. We argue that notifications differ from arbitrary stimuli since in addition to their sensory component, they are used to announce a social interaction. Prior work has demonstrated a causal relationship between digital social interactions (i.e. subject of notifications) and activation of dopaminergic reward circuits [8]. This kind of activation has been correlated with heightened arousal states [9] that are known to influence electrodermal activity [10]. This makes notifications an extremely promising candidate in robustly and repeatably inducing eSCR.

In addition, smartphone notifications follow a variable-ratio reinforcement schedule [9], i.e., a variable delivery rate and uncertain outcome (e.g., positive message from a friend versus a work-related email). This reinforcement schedule, also observed in gambling, is known for its addictive behavior reinforcement. The uncertainty and unpredictability in delivery time and content, combined with our innate desire for social interactions, induces strong arousal states [9] that are less likely to be subject to habituation. In the case of smartphone notifications, the hypothesized relationship between the anticipated rewarding social interaction and the stimulus of the notification is reinforced dozens of times per day [1]. As such, the difference between notifications and arbitrary stimuli should be outlined in their more complex habituation and conditioning characteristics.

3.3 SweatSponse

The aim of Sweatsponse is to improve the notification experience by creating a perception feedback loop that enables a device to efficiently adapt its communication based on user perception without requiring any explicit intervention (see Figure 3.1). This relies on Sweatsponse's ability to infer a user's perception of a notification from the occurrence of an eSCR, or the lack thereof, using a wearable skin conductance sensor. The proposed method is based on the tight temporal coupling between the delivery of a known stimulus (in this case, a notification) and an anticipated eSCR (1-4 s post-stimulus [7]) to avoid responses that could be induced by external non-relevant stimuli. We believe accurate measurements are possible with recent wearable physiological sensor technologies such as Empatica's E4[1] and Thought Technology's Triple Point Sensor (TPS)[2] that offer long-term electrodermal activity recordings. From these measurements, we anticipate perception can be inferred using features extracted by existing effective eSCR response modeling and detection tools [11].

[1] Empatica E4
[2] Thought Technology TPS

3.4 User study

The central objective of this project is to investigate whether skin conductance measurement could be used to infer the perception of notifications without requiring explicit user interaction. To achieve this goal, the following research questions first need to be answered:

- **Q1** Can smartphone notifications provoke measurable event-related skin conductance responses (eSCR)?
- **Q2** Are those eSCR correlated with the participants fear of missing out (FoMO)?
- **Q3** Is prediction performance affected by the modality through which a stimulus was perceived? E.g., does a notification delivered using an auditory alert provoke the same response as a vibrotactile notification?
- **Q4** Knowing that a notification was delivered, is it possible to predict, with meaningful performance, whether it was perceived based on properties of the potentially induced eSCR?

3.4.1 Method

Subjects were greeted with an explanation of the experiment's objectives and asked to read and sign an institutionally approved consent form (REB# 83-0814). A pre-test questionnaire was used to collect standard demographic information, the users' usual notification settings and which of their applications usually generated notifications. Participants were asked to complete the fear of missing out (FoMO) scale, which attempts to quantify one's anxiety in response to missing a potentially rewarding social experience [12].

A TPS was attached to the participants' non-dominant hand following the manufacturer's recommended placement instructions. The sensor streamed the participant's skin conductance measurements to an Android tablet for logging. For the purpose of the study, a notification logging application was developed and installed on the participant's Android smartphone. The experiment application uses notification access permissions to log the time at which a notification was presented and the application responsible for generating it. As observed in prior work, certain Android packages spam the notification channel by continuously updating the notification tray's content without presenting a stimulus to the users [1]. To attenuate the impact of those events on the results, consecutive events that were logged less than one second following an initial

(a) (b)

Fig. 3.2 (a) Overall experiment setup (b) Smartwatch graphical interface

notification by the same application were not considered for analysis.

Participants were instructed to use the buttons on a Pebble smartwatch, placed on the table, to report perceived notifications, and indicate the modality through which each notification was perceived(see Figure 3.2b). For example, if a notification was perceived because of the sound of the device's vibrations on the table, or because of an auditory notification, they would press the "Sound" button. Due to the difficulty of matching the presentation time of a visual notification (e.g., screen lighting up or blinking LED) and the often delayed perception of such an event while engaging in non-smartphone based visual activities, the visual modality was not considered for this study. As such, participants were instructed to not report visual notifications. For the duration of the experiment, the phone's ringer mode was set to the first non-silent mode (e.g., vibration, vibration and sound, etc.) that the participant reported in the pre-test questionnaire as most likely to be used during a normal day. To minimize risks of heightened stress states due to smartphone separation [13] and avoid interference with sensor measurements, participants were allowed to respond to incoming messages and look at notifications using their dominant hand only, but were told to decline incoming calls.

To investigate the influence of user activity on the measured signals, measurements were made under two experimental conditions:

- **Inactive (IC):** Participants were asked to watch a wildlife documentary [14]. The volume was adjusted to ensure it was sufficient for the participants to comfortably understand the documentary's narration through the integrated speakers of a Lenovo G40 laptop. To minimize perceived workload, participants were explicitly told that they would not be questioned about the documentary after the session. This condition was designed to allow for the collection of skin conductance measurements with a minimum amount of motion artifacts and noise introduced by psychological processes. Furthermore, the task took the subjects' attention away from their smartphone and incoming notifications.

- **Active (AC):** Participant were asked to complete a collection of hexagonal paper mazes[3] using a pen. The maze set contains 40 hexagonal mazes of increasing difficulty and was assembled to ensure no possible completion within the duration of the experiment. Participants were told that they had to complete as many mazes as possible during the session, and that they could only move to the next maze once they completed the previous one. This condition aimed at increasing the amount of motion artifacts as well as noise in the physiological signals induced by the hypothesized higher mental demands of the task.

The presentation order of the two conditions was balanced across participants where each condition was presented for 40 minutes. Following completion of each task, perceived workload was sampled using a standard NASA-TLX pen and paper instrument [15].

Although it reduces the ecological validity of the findings, in addition to naturally occurring notifications, an experimenter sent a message to the participant every 120 ± 20 seconds to ensure sufficient data collection during the experiment. Messages were sent using each participant's favorite messaging application, e.g., Whatsapp, Signal, or text message, and did not require a response.

The notification perception rate was anticipated to be artificially high in the quiet environment of the lab. Since most machine learning approaches require a representative amount of negative and positive samples, synthetic "missed" notifications were introduced in the log file 15 seconds before each notification perceived by the user (see Figure4.2). The introduction of "missed" notifications is based on the assumption that if a notification was not perceived, it is impossible for it to induce an eSCR, and is therefore equivalent to sampling the skin conductance signal's noise. "Missed" notifications were not introduced within 15 seconds of perception of a real notification in order to avoid polluting the response to the synthetic ("missed") notification with that of actual notifications.

[3] SRL Maze Task

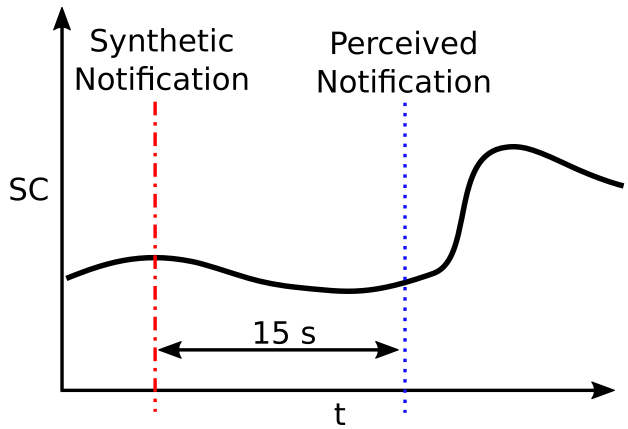


Fig. 3.3 Sketch presenting a notification labeled as perceived by the user, preceded by a synthetic "missed" notification that was never presented to the user.

3.4.2 Skin conductance analysis

All skin conductance signals were processed and analyzed post-experiment using Ledalab[4] in a Matlab environment. Traditionally, the skin conductance signal is decomposed into its tonic component, a low frequency oscillation independent of specific events, and its phasic component, characterized by abrupt changes in skin conductance level associated with discrete events [10]. Continuous decomposition analysis (CDA) was used to extract the phasic activity from the raw skin conductance signal [11]. The maximum of the phasic activity (PhasicMax) was extracted within a response window of one second after notification presentation to an additional six seconds. Two seconds were added to Lockhart's suggested 1-4 s average onset delay to include the peak of the responses [7].

3.5 Hypotheses

Based on the prior literature on skin conductance responses, notifications and their social components, the following hypotheses were made:

- **H1** It is anticipated that despite the usually non-startling sensory properties of a smartphone notification, the anticipation of a potentially rewarding social interaction [9] will be sufficient to trigger a measurable eSCR.

[4] Ledalab

- **H2** The difference in maximum phasic activity between perceived and missed notifications will be correlated with the subject's score on the Fear of Missing Out scale (FoMO) due to the value associated with their social interactions.
- **H3** Since the eSCR response to notifications is assumed to be more heavily influenced by its conditioned social component than its sensory characteristics, vibrotactile and auditory presentation of notifications will not have a significantly different impact on the maximum phasic activity of the skin conductance signals.
- **H4** Assuming an eSCR in response to a notification can be measured, a classifier will be able to predict whether a notification was perceived from skin conductance measurements.

Predictor	Estimated Coefficient	Std. Error	z value	p-value
Intercept	0.015251	0.075648	0.202	0.840221
$log2$(PhasicMax)	0.375126	0.143005	2.623	**0.008712**
$log2$(PhasicMax):Age	-0.023024	0.006689	-3.442	**0.000577**
$log2$(PhasicMax):Gender	0.035841	0.047386	0.756	0.449432
$log2$(PhasicMax):FoMO	0.114687	0.046617	2.460	**0.013886**

Null deviance 1624.2 on 1177 degrees of freedom.
Residual deviance 1588.8.8 on 1173 degrees of freedom

Table 3.1 Logistic regression analysis summary and Wald's test output.

3.6 Results

3.6.1 Participants

A total of 17 subjects aged between 19 and 29 years ($\bar{x}$ = 24, 8 identified as females) participated in the study and received CAD$ 15 per hour as compensation for their time for a total of CAD$ 30. Participants were recruited from university mailing lists, classified ads, and the university community's social network groups. Only Android users older than 18 years of age, who reported receiving at least 50 notifications per day, and who had not participated in a previous notification perception study, were recruited. Considering that the physiological sensor has to be worn on the non-dominant hand, and that smartwatches are almost always worn on the non-dominant wrist, only participants used to receiving notifications through their smartphone were eligible to participate, minimizing risks of artificially modifying subjects' notification perception habits by changing the location of their wearable(s) during the study.

3.6.2 Experimental tasks

The two experimental conditions were initially selected to offer a controlled and semi-controlled context for evaluation. Since they are not standardized tasks, a paired t-test was used to confirm our working hypothesis that the mean aggregated NASA task load index in the AC was significantly greater than in the IC ($\overline{tlx}_{IC} = 32.11$, $\sigma_{IC} = 17.06$, $\overline{tlx}_{AC} = 60.41$, $\sigma_{AC} = 17.03$, t(16)=-5.9554, p< .0001). Cohen's effect size (d=1.66) suggests a very high practical difference between the two tasks' aggregated perceived workload.

3.6.3 Notifications

On average, participants received 26.2 ($\sigma_{notif} = 8.3$) notifications per experimental block, of which 20 were initiated by the experimenter. Out of the 17 participants, 11 chose to set their phone's ringer mode to vibrations only, 5 to sound only and 1 to sound and vibration. Based on the registered button presses, on average 93.4% ($\sigma_{perc} = 5.48\%$) of notifications presented were perceived when they were first delivered to the user, which supports the initial decision to introduce synthetic "missed" notifications in the log file to balance the dataset. In the absence of more true missed notifications to do a formal comparison, skin conductance measurements following both true and synthetic missed notifications seemed to follow random patterns, i.e., there was no evidence of repeatable event-related responses in either case.

3.6.4 Skin Conductance Measurements

Building on the data analysis approach used by Andersen et al. and Blum et al. [4, 5], a logistic regression analysis was conducted to investigate the contribution of the maximum phasic activity following the reception of a notification to the prediction of its perception. Based on prior work on skin conductance activity, interactions between PhasicMax and the participants' age and gender were included as predictor variables [16]. To address Q2, interaction with self-reported fear of missing out (FoMO) was also included in the model.

To attenuate the influence of inter-subject PhasicMax variations on the model's coefficients, the base-2 logarithm of the raw PhasicMax values was used [10]. Table 3.1 presents the logistic regression model employed for the analysis, its estimated regression coefficients and the output of Wald's test investigating the contribution of each predictor variable to the model's fitness.

Q1 - Smartphone Notifications on Electrodermal Activity

Based on these results, we can conclude that the contribution of log_2PhasicMax to the model's fitness is statistically significant. The estimated coefficient of 0.375 shows that a two-fold increase in the measured PhasicMax following the delivery of a notification increases the probability that the alert was perceived by exp(0.375) = 1.45 times.

A second test on the residual deviance is used to evaluate how well the proposed model fits the collected data [5]. Considering the probability that a χ^2 test with 1173 degrees of freedom would be greater than 1588.8 is $< .00001\%$ ($p < .05$), we *must reject* the null hypothesis that our logistic regression model provides an adequate fit of the data. Although this test reveals poor model fitness, it does not invalidate the significant contribution of PhasicMax, but rather, suggests that other factors, not accounted for in the current model, could explain the variance of the data. For example, the model could be enhanced by including stimulus-related properties, as proposed by Andersen et al. and Blum et al. [4, 5].

A Spearman correlation test between the delivery time of a notification and its corresponding PhasicMax was employed to validate our working hypothesis that from an electrodermal perspective, smartphone notifications are different from arbitrary auditory and vibrotactile stimuli (see Section 3.2.2). As previously mentioned, we suggest that this is due to the social nature of smartphone notifications, for which we would not expect habituation, or would only observe habituation at a much slower rate than that of arbitrary stimuli (see Section 3.2.2). A very weak negative monotonic relationship was observed between the two variables (r=-.06, $p < .05$), indicating that the amplitude of responses showed a weak downward trend over the two-hour duration of the experiment, thus providing support for our working hypothesis. Even though a very small, yet significant, habituation was observed, the scale at which it was occurring far exceeds the habituation time observed in the cases of arbitrary stimuli presentations [10]. In addition to the relevance of the stimuli, there is a possibility that eSCR to smartphone notifications behaves more similarly to that of defensive responses, which were shown to exhibit very little habituation over time, than to orienting responses, which usually have fast habituation [10]. This defensive interpretation would also be aligned with prior work that outlined the negative perception of notifications and their properties as physical and psychological stressors [17]. However, a different experiment design explicitly comparing habituation to smartphone notifications and arbitrary stimuli, delivered in the same time period, would be required to conclude that this hypothesis holds.

The logistic regression analysis supports H1, stating that the perception of smartphone notifications provoke event-related skin conductance responses, as reflected by the increased phasic electrodermal activity.

Q2 - Fear of Missing Out on Physiological Response

The logistic regression analysis showed a statistically significant contribution of the interaction between PhasicMax and FoMO on the model's fitness. Further model analysis revealed that FoMO scores below those observed during the study would cause a slope inversion. This inversion could be interpreted as an illogical decrease in the probability that a notification was perceived when larger skin conductance responses are observed. Based on the model's lack of fit to the data, its observed behavior and the reported statistically significant contribution of the interaction term, we must conclude that the sample size used in this study was insufficient to allow for analysis of the influence of fear of missing out on eSCR. Similar observations and conclusions can be made regarding the statistically significant interaction between age and PhasicMax, as the fitted model contradicts prior work that has repeatedly shown a negative monotonic correlation between age and electrodermal activity [10, 16].

A moderate positive correlation was observed for the difference between the mean PhasicMax of perceived and missed notifications, and subjects' FoMO scores ($r=0.604$, $t(15)=2.5114$, $p < .05$). This suggests that participants with high self-reported FoMO scores generally showed larger PhasicMax difference between perceived and missed smartphone notifications than those with lower scores.

It is conceivable that the model used the interaction terms to identify participants' unique response amplitude, which statistically had a significant positive effect on the model's fitness. As such, even though a positive correlation was observed between FoMO scores and the range of amplitude of eSCR, the significant interaction term from the logistic regression analysis prevents us from drawing any conclusion with regards to our second research question.

Q3 - Perception Modality on Physiological Response

The finding that notifications indeed provoke eSCR allows us to consider our third research question: how does perception modality influence these responses? Out of all the alerts that were perceived, 73.0% were heard and 27.0% were felt tactually, either through direct contact with the phone or by conduction of the vibrations through the table.

A Wilcoxon rank sum test revealed statistically significant difference between the PhasicMax distribution location of notifications reported as "Sound‘" and "Vibration" (estimated difference in distribution location of 0.25, W = 39245, p< .0001). Notifications perceived due to their tactual properties provoked greater skin conductance responses than their auditory counterparts (medium effect size, Cohen's d = 0.509). To examine this difference in greater detail, a comparison between notifications perceived via sound was conducted for the cases where the auditory stimulus is an unintentional consequence of the device's vibration instead of a normal audio alert. A Wilcoxon rank sum test shows that there exists a significant PhasicMax difference between these two cases (estimated difference in distribution location of 0.207, W = 22602, p< .0001). Interestingly, alerts perceived due to the sound of vibrations were accompanied by greater phasic activity than their purely auditory counterparts (medium effect size, Cohen's d=0.578). Since participants were asked to indicate the modality that they believed allowed them to perceive the notification, even in the cases where a vibrotactile stimulus was detected because of its sound, it is possible that the haptic component contributed to an increase in phasic skin conductance activity.

Similarly, a comparison of the maximum phasic activity following the perception of a vibrotactile notification via its tactual or auditory component shows that vibrotactile notifications perceived because of their haptic properties elicited greater responses than those reported to be caused by the auditory artifacts of the vibrations (estimated difference in distribution location of 0.211, W = 20077, p< .0001, small effect size, Cohen's d=0.38).

These results do *not* support our third hypothesis (H3) stating that the phasic component of the skin conductance signal would not be significantly different between the cases where notifications were perceived through the auditory and haptic channels. Instead, significant differences in the amplitude of responses were observed, with notifications presented in vibration mode generally eliciting larger eSCR than their auditory counterparts. Even though these results contradict our research hypothesis, this presents an additional opportunity: the differences between modalities suggest that a perception prediction system performance could benefit from knowing a device's current ringer mode when attempting to make perception predictions.

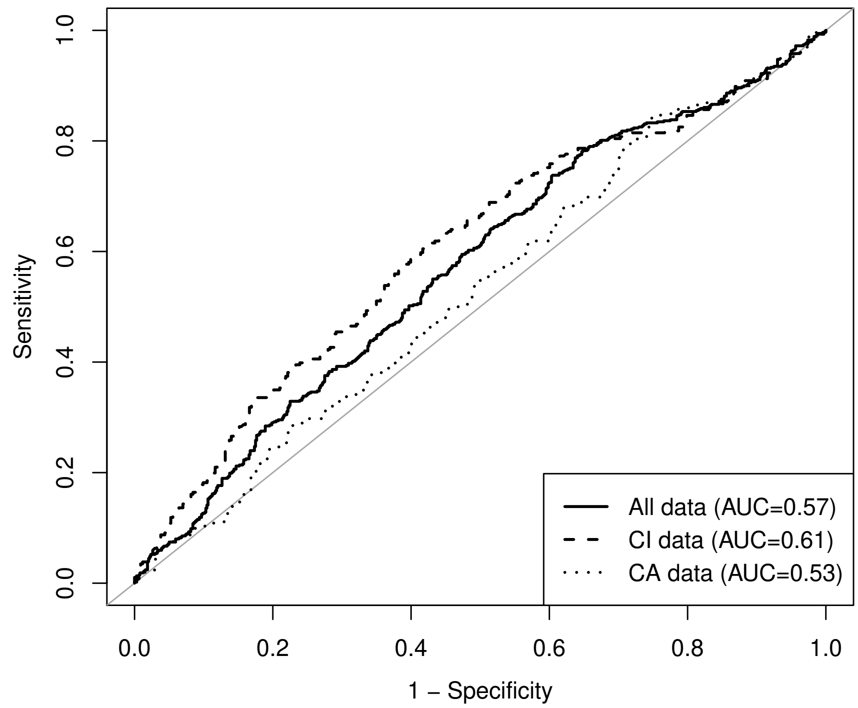


Fig. 3.4 ROC curves generated from the proposed model for the inactive condition (IC), active condition (AC) and aggregated (all data).

Q4 - Perception Prediction Performance

To investigate whether a perception prediction system based on skin conductance could assist in a notification scenario, the device's current ringer mode was included as a predictor variable to the logistic regression model, and receiver operating characteristic (ROC) curves were generated using leave one subject out cross-validation (see Figure 4.3).

To compare the effect of the user's activity and perceived workload on the model's performance, three ROC curves are presented. The first represents the general SweatSponse performance and was created using all of the held-out subject's data as the test set independently from the experimental condition. The second and third ROC curves were generated using only the held-out test data collected during the inactive (IC) and active (AC) experimental conditions, respectively.

The difference between the area under the curve (AUC) for combined experimental conditions and that of a random binary predictor is statistically significant as revealed by pROC's

”Bootstrap” method (AUC_{ALL} = .0.573, AUC_{RDM} = .50, D=4.37, n.boot=2000, boot.stratified=1, $p < .0001$). Hence, the overall performance of SweatSponse is statistically significantly better than randomly predicting perception.

The observable difference in AUC between the data from the IC and AC conditions was found to be statistically significant (AUC_{IC} = .61, AUC_{AC} = .53, D=2.1965, n.boot=2000, boot.stratified=1, $p < .05$). This suggests that there is a significant drop in performance as the perceived workload and user activity increases.

When selecting a general threshold that minimizes the distance between the ROC curve and the [0,1] coordinate on the ROC plot, an accuracy of 0.61, recall value of 0.75 and specificity of 0.38 are obtained. By interpreting these results and the ROC curves, one can conclude that the prediction performance of the system performs better than a random binary classifier, but has room for improvement. It correctly inferred perception in 75% of the cases where a notification was indeed perceived by participants, at the expense of only correctly identifying 38% of missed notifications as such. In its current state and using this threshold, it is anticipated that the proposed system could potentially enhance users' notification experience by automating non-critical actions such as reducing the intensity of a notification once it was perceived. By automatically reducing notification intensity, as opposed to completely silencing it, SweatSponse could reduce the risks of negatively impacting the effectiveness of the notification system due to false predictions.

The comparison of ROC curves and logistic regression analysis suggest that the perception of smartphone notifications do indeed produce measurable event-related skin conductance responses, as quantified by the maximum of skin conductance phasic activity (H1), and that those responses could allow a system to obtain perception feedback. However promising these results, more empirical data, specifically considering the user experience with such a system, would be required to conclude that it offers a superior notification experience in practice (H4).

3.7 Limitations

3.7.1 Experiment

While the results of this study are promising, it must be noted that the participants were drawn from a young adult population, who were all Android users, and who lived in a North American cultural context at the time of the experiment.

This study was conducted in a laboratory setting over a two-hour period. It is possible that the time of the day at which participants completed the procedure influenced their physiological responses, due to prior expectations of a message's origins. For example, if a subject usually receives work-related emails during the time frame in which they participated in the study, they may not experience the same response as if they were participating at a later time when they usually expect their partner's call.

Furthermore, it is possible that the high notification frequency induced frustration or other negative emotional states that may have influenced physiological measurements [10]. We hypothesize that the observed weak habituation could be a consequence of this phenomena, combined with the knowledge that most notifications were in fact originating from an experimenter. It is expected that under regular, non-experimental conditions, the effect of smartphone notifications on eSCR would, if anything, be even more pronounced.

To maximize performance and external validity of SweatSponse, more data should be collected in the wild, allowing the measurement of eSCR in response to naturally occurring notifications, as opposed to those acquired under laboratory conditions. Nevertheless, the chosen tasks, i.e., watching a documentary and completing a set of paper mazes, are representative of everyday activities such as watching television, engaging in desk work and attending a meeting or presentation.

3.7.2 System

SweatSponse cannot avoid limitations inherent to the skin conductance measurements. Due to the 1-4 second latency in the response following a stimulus presentation [7], we would advise against using eSCR for the detection of frequent events (less than 5 seconds inter-stimulus intervals) and especially to detect time- and safety-critical events. Further work would be needed to validate whether other behavioral or physiological signals could reliably be used for the cases where multiple stimuli are being delivered in rapid sequence. Indeed, prior work on smartphone separation showed that participants exhibited significantly greater stress levels, reflected by an increase in stress-specific gestures, when they could not access their own smartphone or could only use a stranger's system than when they were allowed to quickly access their own device [13]. It is possible that stress-related markers in motion patterns could be used as indicators of smartphone notification perception even when users are not separated from their device.

In addition, some environmental contexts could interfere with the function of SweatSponse by not satisfying minimal conditions required for the electrical measurement of skin conductance. For example, even though they should be tightly coupled to the skin, existing SCR sensing wearables often expose the electrodes to the elements (e.g., rain or snow), which can negatively impact measurement accuracy. Furthermore, while this experiment was conducted in a static environment, certain physical activities could introduce a greater level of motion artifacts to the measurements than existing automated signal correction techniques can process reliably. This is particularly problematic, since the contexts in which the quality of measurements are the worst are the same as those in which notifications are most likely to be missed [5].

3.8 Future Applications

In the future, we expect that physiological measurement devices will increase in accuracy and be embedded in mainstream wearable technologies, which will improve SweatSponse's perception detection and viability for day-to-day use.

We now present a possible application of SweatSponse in notification systems that we anticipate could be realized with further development. Instead of users manually adjusting the volume or intensity of their notifications, Sweatsponse could allow the introduction of ”scaling” notifications. A ”scaling” notification would start at the minimum intensity at which it is likely to be perceived, using methods such as the one proposed by Blum et al. [5]. It would then gradually ramp up in volume and/or vibration intensity until SweatSponse reports a high probability of perception at which point, the alert intensity could be reduced or stopped. Scaling could also involve a change in modality. For example, if the notification is being delivered initially through vibration, the intensity could be increased until its maximum value, and then switched to more salient auditory alerts in the prolonged absence of responses. This would be particularly valuable when the user is attempting to locate their phone that was last left in vibration or silent mode.

3.9 Conclusion

From the results of the presented laboratory study, we conclude for the first time that smartphone notifications reliably induce skin conductance responses. Furthermore, given their strong association with potential social interaction, smartphone notifications differ from arbitrary vibrations, sound, and light feedback used in prior studies employing electrodermal measurements. This is reflected in the comparatively marginal habituation observed over the duration of the study. Based on these results, this work introduces SweatSponse, the first method that allows for prediction of perception of a notification after its presentation, without the need for users to engage with their device. Preliminary performance assessments indicate that the system presents a promising approach to perception prediction of smartphone notifications.

3.10 Acknowledgements

The authors wish to thank all colleagues from the Shared Reality Lab for their valuable help, and Prof. Stefanie Blain-Moraes for lending us Thought Technology TPS sensors without which this study would not have been possible.

This work was funded by a McGill Engineering Doctoral Award (MEDA), and a McGill TechAccelR innovation Grant. Icons used in Figure 1 were made by Pixel perfect, Smashicons and Freepik from Flaticon.

References

[1] M. Pielot, R. de Oliveira, H. Kwak, and N. Oliver, "Didn'T You See My Message?: Predicting Attentiveness to Mobile Instant Messages," in *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: ACM, 2014, pp. 3319–3328. [Online]. Available: http://doi.acm.org/10.1145/2556288.2556973

[2] Y.-J. Chang and J. C. Tang, "Investigating Mobile Users' Ringer Mode Usage and Attentiveness and Responsiveness to Communication," in *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '15. New York, NY, USA: ACM, 2015, pp. 6–15. [Online]. Available: http://doi.acm.org/10.1145/2785830.2785852

[3] A. Sahami Shirazi, N. Henze, T. Dingler, M. Pielot, D. Weber, and A. Schmidt, "Large-scale Assessment of Mobile Notifications," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: ACM, 2014, pp. 3055–3064. [Online]. Available: http://doi.acm.org/10.1145/2556288.2557189

[4] H. J. Andersen, A. Morrison, and L. Knudsen, "Modeling Vibrotactile Detection by Logistic Regression," in *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, ser. NordiCHI '12. New York, NY, USA: ACM, 2012, pp. 500–503. [Online]. Available: http://doi.acm.org/10.1145/2399016.2399092

[5] J. R. Blum, I. Frissen, and J. R. Cooperstock, “Improving Haptic Feedback on Wearable Devices through Accelerometer Measurements,” in *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST ’15)*. New York, New York, USA: ACM Press, 2015, pp. 31–36. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2807442.2807474

[6] J. T. Cacioppo, L. G. Tassinary, and G. Berntson, *Handbook of Psychophysiology*. Cambridge University Press, 2007. [Online]. Available: http://ebooks.cambridge.org/ref/id/CBO9780511546396

[7] R. A. Lockhart, “Interrelations Between Amplitude, Latency, Rise Time, and the Edelberg Recovery Measure of the Galvanic Skin Response,” *Psychophysiology*, vol. 9, no. 4, pp. 437–442, 1972. [Online]. Available: http://dx.doi.org/10.1111/j.1469-8986.1972.tb01791.x

[8] S. Krach, F. M. Paulus, M. Bodden, and T. Kircher, “The Rewarding Nature of Social Interactions,” *Frontiers in Behavioral Neuroscience*, vol. 4, p. 22, May 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2889690/

[9] S. P. L. Veissière and M. Stendel, “Hypernatural Monitoring: A Social Rehearsal Account of Smartphone Addiction,” *Frontiers in Psychology*, vol. 9, p. 141, 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2018.00141

[10] W. Boucsein, “Electrodermal activity,” New York, 2012. [Online]. Available: http://site.ebrary.com/id/10517958

[11] M. Benedek and C. Kaernbach, “A continuous measure of phasic electrodermal activity,” *Journal of Neuroscience Methods*, vol. 190, no. 1, pp. 80–91, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165027010002335

[12] A. K. Przybylski, K. Murayama, C. R. DeHaan, and V. Gladwell, “Motivational, emotional, and behavioral correlates of fear of missing out,” *Computers in Human Behavior*, vol. 29, no. 4, pp. 1841–1848, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0747563213000800

[13] V. Konok, Á. Pogány, and Á. Miklósi, “Mobile attachment: Separation from the mobile phone induces physiological and behavioural stress and attentional bias to separation-related stimuli,” *Computers in Human Behavior*, vol. 71, pp. 228–239, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0747563217300754

[14] H. M. Mix, “Wild Russia - The secret forest,” 2009.

[15] S. G. Hart and L. E. Staveland, “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research,” in *Human Mental Workload*, P. A. Hancock and N. B. T. A. i. P. Meshkati, Eds. North-Holland, 1988, vol. 52, pp. 139–183. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0166411508623869

[16] C. Eisdorfer, H. O. Doerr, and W. Follette, "Electrodermal Reactivity: An Analysis by Age and Sex," *Journal of Human Stress*, vol. 6, no. 4, pp. 39–42, Dec. 1980. [Online]. Available: https://doi.org/10.1080/0097840X.1980.9936107

[17] S. Yoon and K.-p. Lee, "A Study on Notification System Design of Smartphone Messenger Considering the User ' s Stress," *Archives of design research*, vol. 28, no. 2, pp. 77–91, 2015.

Chapter 4

In Situ Electrodermal, Heart Rate, Heart Rate Variability and Wrist-Motion Responses to the Presentation of Mobile Notifications

Fortin, P. E., & Cooperstock, J. R. (2021). In Situ Electrodermal, Heart Rate, Heart Rate Variability and Wrist-Motion Responses to the Presentation of Mobile Notifications. To be submitted to the Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)

Preface to Chapter 4:

Chapter 3 demonstrated the feasibility of using skin conductance measurements as a means to better understand users' perception of smartphone notifications. In this chapter, we expand on that work by reporting on an *in situ* user study investigating the impact of notification perception on skin conductance, heart rate, heart rate variability and wrist-motion. In addition, we explore how the different sensing channels contribute to the notification perception prediction system's classification performance.

This chapter presents the first evidence of physiological responses to smartphone notifications outside of laboratory conditions. More specifically, perceiving a notification resulted in a decrease in heart rate and heart rate variability. Changes in wrist-motion patterns were observed, even in cases where the notification was not addressed within the accelerometer measurement period, outlining the profoundly disruptive nature of notifications on users' primary tasks. Finally, the inclusion of supplementary physiological signals allowed for a significant increase in perception classification performance, with heart rate and heart variability features being identified as the strongest predictors. Together, these different contributions are proof that the proposed physiologically informed notification research approach can, and does reveal insights that were beyond the scope of traditional notification research methodologies.

Contributions of Authors:

Pascal E. Fortin was the primary author and contributor to the TechAccelR innovation award partially funding this work, was responsible for the ideation, design and implementation of the technical framework required for data collection, data analysis and paper writing. Prof. Jeremy R. Cooperstock edited the manuscript, contributed to the TechAccelR innovation award and supervised the research.

Abstract

Current smartphone notification research relies on direct notification interactions and subjective self-reports as means to study users' experience and behavior. In this work, we instead employ a research methodology based on interaction-less physiological and behavioral monitoring before and after the presentation of each notification. We illustrate the richness of this approach by presenting the first reports of the impact of notification presentation on users' heart rate, heart rate variability and wrist motion *in situ*. Results from a prior laboratory study reporting an increase in electrodermal activity following the presentation of a notification were successfully replicated. Beyond documenting physiological changes induced by notifications *in situ*, we demonstrate that the physiological responses collected allow for a device to infer with up to 81.8% accuracy whether a notification was perceived by its user without the need for user action (e.g., notification tray or application interactions), a source of information that is unavailable using traditional notification research methodologies. Based on classification models' performance, we further claim that such perception confirmation technology could be deployed on today's commercially available smartwatches and fitness trackers.

4.1 Introduction

Interruptions caused by smartphone notifications are not only numerous, but were also shown to have detrimental effects on mental well-being [1, 2], which projects on workplace performance (e.g., attention, productivity) [1]. Indeed, Stothart et al. have shown that the perception of a notification increased risks of making errors as much as actually reading a message or engaging in a phone call [3]. Kushlev et al. reported on a study during which 221 subjects were asked to turn on all possible notifications from their devices for a week, followed by a week where notifications were set to silent and phones placed out of direct access (e.g., pocket, bag) [1]. Their results suggested that turning on all notifications resulted in significantly higher levels of inattention and hyperactivity accompanied by lower levels of perceived productivity, social connectedness, mastery over their environment, meaning in life and choice over their actions than in the week of reduced access. Interested in what specific properties of notifications were responsible for some of these negative consequences, Yoon et al. identified three main categories of notification-related stress: the physical notification, itself, the associated message content and finally the social expectations associated with responsiveness. Of significant interest to this research, physical notification stress related to the notification stimuli themselves, the context in which they are presented and the frequency of their presentation, was identified as one of the most important stressors for their participants [4].

The prior literature typically relied on one or a combination of the two main notification research methodological trends: the passive approach, which relies on instrumenting participants' smartphones and observing their notification interactions and behavior, and the active approach, which instead probes participants for their perception of notification using self-reporting instruments. The following sections discuss examples of these two methodological approaches by bringing to light their advantages and limitations.

4.1.1 Passive Notification Research

Passive notification research methods rely on the background observation of participants' notification interactions. Typically, research based on this approach uses a mobile application to log notification events (i.e., presentation, update, removal), their content, contact(s) when applicable, and the application that generated the alert. Additional information can also be collected to explore specific topics (e.g., wifi network, screen lock state, user activity). Using only notification logging limits the type of research questions that can be explored to purely objective and behavioral ones, e.g., what is a typical daily notification volume? [5], how do people interact with notifications in the notification tray? [6], what types of applications are generating the largest amount of notifications? [7], or what is the best moment to present a notification if trying to maximize engagement time? [8].

Excellent examples of this technique were presented by Weber et al. and Okoshi et al., who deployed their data collection instruments to 3,953 and 687,840 participants respectively [6, 8]. Harnessing the scale of their user base, Weber et al. were able to identify notification handling "personas" that showed distinct notification accumulation and interaction patterns [6]. Okoshi et al., on the other hand, studied when the presentation of a notification had the highest probability of leading to deep user engagement, i.e., not only make the user open the application, but also keep them on the platform. To do so, they manipulated the point at which they notified users using an attention- and engagement-aware system that delivered the signals based on sensor readings, device state and user actions [8]. While not all projects employing passive notification logging decide to do so (e.g., [5, 7, 9]), making the data collection application freely available on public application repositories [6] or as part of an existing application [8] significantly broadens the reach and potential scale of the research. In addition to the possibility of reaching to a large audience, the passive approach does not require participants to explicitly engage with subjective data collection instruments, which reduces participation burden and risks of introducing biases in the data. That is further strengthened by the long data collection period afforded by the passive methodology, which may even lead participants to forget that they are taking part in a study.

This approach has demonstrated its benefits. For example, following this research methodology, Weber et al. were able to demonstrate that three main user types exist with regards to notification handling and notification tray interactions, and were able to make literature-informed hypotheses on the factors that drive participants' behavior [6]. However, this methodology also suffers from a serious limitation. Indeed, a purely passive approach offers no insights into attitudinal components of the notification experience and the reasons behind the observed behaviors. Referring again to Weber et al., the authors lacked the means to probe the participants of their study to determine the validity of their hypotheses.

4.1.2 Active Notification Research

Filling this methodological gap, active notification research relies on the presentation of time- or event-triggered questionnaires pertaining to the participants' experience of one or more notifications. Questionnaires can be presented as frequently as after each notification to reduce retrospective bias, or at fixed time intervals to reduce the experiment participation burden. In the case where questionnaires are presented at a later time, they can be accompanied by a reminder of notifications' content to make sure the participant recalls the context of the interaction [6]. One distinctive advantage of this methodology over the passive alternative is that it allows the exploration of attitudinal research questions such as the subjective workload imposed by notifications [10], their perceived disruption [11] and participants' interruptibility [12]. It can also be useful in cases where a research question's variables of interest cannot be measured automatically in a reliable manner, e.g., number of people around the participant, whether they consider their current location a public or private space or the activity in which they are currently engaged [13].

A significant limitation of the active notification research methodology is the supplementary burden it imposes on participants. Indeed, the frequency at which participants are polled about their experience adds to the preexisting interruption burden by increasing the total number of notifications they are presented with. Reducing the questionnaire presentation frequency reduces the bias introduced by the measurement tools themselves. However, this also reduces the granularity of the data and risks an increase in retrospective bias, which further impacts the quality of the data. Another component to consider is the self-reporting interface itself. Considering a participant may have to interact frequently with the interface, how optimized it is for such reporting and where it is embedded in the user's mobile interaction can impact the quantity of collected data as well as the degree to which it will be perceived as intrusive [14, 15]. Poor instrument design and embedding risks the introduction of biases by inducing negative affect and annoyance.

Both passive and active notification research approaches can take advantage of large scale application deployments. However, designing, deploying and supporting applications at that scale requires significant mobile development experience and resources. One difficulty researchers interested in using this approach may encounter is ensuring cross-manufacturer compatibility of their Android application. In the authors' experience, certain manufacturers (i.e., Xiaomi, OPPO and VIVO) employ aggressive power saving features that can interrupt foreground services, required for ongoing data collection, from operating smoothly. Android One and Google Pixel devices have been found to be the most reliable for such data collection. While the authors do not have experience designing notification logging applications for iOS, it seems to be possible [16] and we hypothesize that with sufficient development experience, the hardware and software homogeneity should make the task easier.

4.1.3 Notifications Psychophysiology

Smartphone notifications are a signal that frequently precede digital social interactions. Prior work has identified a causal relationship between digital social interactions and heightened arousal states [17] that are known to significantly influence electrodermal activity, heart rate and heart rate variability [18–20]. In addition, smartphone notifications follow a variable-ratio reinforcement schedule [17], i.e., a variable delivery rate and uncertain outcome. This reinforcement schedule, also observed in gambling, is known for its addictive behavior reinforcement. The uncertainty and unpredictability in delivery time and content, combined with our innate desire for social interactions, induces strong arousal states that are less likely to be subject to habituation [9]. In the case of smartphone notifications, the hypothesized relationship between the anticipated rewarding social interaction and the stimulus of the notification is reinforced dozens of times per day. Based on this prior literature, it is anticipated that perceiving a notification will provoke significant changes in physiological signals.

Changes in heart rate (HR) following the presentation of stimuli were extensively studied in the psychology community, but never in response to the participants' perception of their own notifications outside of the laboratory. Increases in HR are frequently associated with high psychological arousal states [19, 21], rejection of the environment, painful [22], and stressful or otherwise unpleasant stimuli [23]. On the other hand, a deceleration of HR is thought to be correlated with increased sensory sensitivity [24], e.g., to a stimulus that draws one's attention, and less arousing experiences [19]. The impact of participants' perception of their own notifications on heart rate remains unclear considering the marked contrast between the positive anticipation of potentially rewarding social interactions and their alerts' negative, stressful and interrupting nature [4]. Further increasing analysis complexity, theories supporting the possible simultaneous observation of activational and inhibitory heart rate responses to stimulation were presented [24], suggesting that they could cancel out each other's physiological impact resulting in a null heart rate change.

Similarly, changes in heart rate variability (HRV) have been measured in response to a variety of arbitrary stimuli, but never to participants' perception of their own notifications. Of particular interest, HRV is known to offer unique insights into participants' autonomic nervous system activity, comprising both the sympathetic nervous system, responsible for the well known fight or flight response, and the parasympathetic nervous system, associated with the digestion, reproduction and resting functions [25]. Decreases in HRV have been reported in response to increases in task demand [26], cognitive load [27], psychological stress level [20, 28] and emotional arousal [19]. Considering the wealth of information it can reveal on participants' internal state, HRV is a promising measurement to consider in the analysis of participants' responses to smartphone notifications.

Skin conductance measurements are frequently used to index emotional arousal and stress in laboratory and *in situ* conditions [29, 30]. In addition, the perception of a relevant stimulus is frequently accompanied by an event-related skin conductance response (ER-SCR) [18]. An ER-SCR can be described as an abrupt increase in skin conductance starting one to four seconds after the presentation of a stimulus, followed by a comparatively slower return to baseline levels [31]. In a laboratory study, Fortin et al. demonstrated that participants perceiving *their own smartphone notifications* exhibited significantly larger maximum phasic activity after the presentation of an alert, than before its presentation [9]. The authors argued that the social component and relevance of notifications significantly differentiated them from the impersonal "notifications" (e.g., arbitrary sounds, vibrations, electric shock, flashes of lights or fake notifications) frequently used in the psychology and HCI literature. Even though the authors claim that their work was the first report of physiological responses to participants' own notifications, major limitations are that the study took place in laboratory context and under specific experimental task conditions, significantly reducing the ecological validity of their findings. Furthermore, participants were asked to press a button to confirm notification perception, which may have introduced physiological responses associated with demand characteristics. As such, it is uncertain whether their results would hold *in situ*, where electrodermal activity measurements are known to be extremely noisy.

4.2 Proposed Approach

In this work, we propose a novel notification research methodology that extends the passive notification research approach with *in situ* physiological sensing. Indeed, quantifying changes in physiological signals resulting from the perception of a notification would offer a more complete picture of the user's internal state, a component that has not yet been explored in the notification research literature. We anticipate that the proposed approach will open the door to the investigation of new behavioral and attitudinal research questions, while minimally burdening participants with questionnaires during data collection. Inspired by the work of Fortin et al., we propose an *in situ* investigation of notification perception which should significantly enhance the ecological validity of the findings. Furthermore, whereas Fortin et al. employed only skin conductance measurements and limited their investigation to post-notification maximum phasic activity, we expand the sensing modalities and report on changes in heart rate and heart rate variability. In addition to physiological signals, we also assess whether the perception of a smartphone notification has a significant impact on how users are physically carrying out their main activity as reflected by changes in their wrist motion patterns.

This paper makes the following contributions:

1. A successful *in situ* replication and expansion of the laboratory results presented by Fortin et al. [9] reporting increased electrodermal activity following the presentation of a notification.

2. Evidence that heart rate and some of the most popular heart rate variability features significantly decrease after the perception of a smartphone notification.

3. Evidence that the perception of a notification disrupts activity execution to an extent that significantly alters the user's motion patterns while conducting a task, even when the user does not immediately attend to the notification.

4. An extension of the interaction-less notification perception prediction system presented by Fortin et al., integrating the supplementary physiological and motion signals to achieve a perception prediction accuracy of up to 81.8%.

5. Among the physiological and behavioral sensing channels considered in this work, the identification of heart rate and heart rate variability as the strongest predictors of notification perception, followed by wrist-motion and electrodermal activity.

4.3 Experiment

The central objective of this work is to investigate *in situ* the impact of notification perception on users' physiology and wrist motion. A secondary objective consists of the refinement of the interaction-less notification perception prediction framework presented by Fortin et al. [9]. To achieve these goals, the following research questions need to be answered:

- **RQ1** How does the presentation of a notification influences a user's
  - **RQ1.1** heart rate?
  - **RQ1.2** heart rate variability?
  - **RQ1.3** electrodermal activity (replication and extension of [9])?
- **RQ2** Does perceiving a notification disrupt users' activity to an extent that can be detected from wrist-based accelerometer measurements?
- **RQ3** Can the collected physiological and motion signals be used to improve the current state of the art in terms of interaction-less notification perception prediction?

4.3.1 Protocol

Preparation

Due to constraints imposed by the Covid-19 pandemic, an equipment package containing participant instructions and a Shimmer3 GSR+ physiological sensor was delivered to the participants' residence on the day before data collection. A video call was held to walk them through the process of putting on the sensor, as well as downloading, installing and using the data collection application. The participants therefore had remotely supervised practical experience in wearing the sensor appropriately, operating and troubleshooting the data collection application. During that first meeting, they were also asked to read and digitally sign an institutionally approved consent form (REB# 83-0814), and to complete the pre-experiment questionnaire (Section 4.3.2).

Data Collection

Participants were instructed to use the data collection system for a duration of 8 waking hours, spent in their residence, during which they would be engaging in their regular activities. These hours did not have to be consecutive, but the data collection had to be performed during a single calendar day. Furthermore, participants were told that they had to have their smartphone with them at all time to prevent connection interruptions with the sensor. During data collection, their device's ringer mode had to be set to a non-silent mode, i.e., sound, vibration or sound and vibration, to ensure they would perceive their incoming notifications. As they had their smartphone with them at all time, and its ringer was non-silent, all presented notifications were assumed to be perceived by participants. Participants had to pause data collection and remove the sensor when engaging in activities that involved liquids or that could otherwise damage the device, e.g., hand washing, laundry, cooking, eating. Beyond these requirements, participants were instructed to use their smartphone as they would on any regular day. In addition to wearing the physiological sensor, participants were asked to keep a log of activities they were engaging in during the day, and when they were in social settings, e.g., in a video call, with a roommate, parent, child or partner. The log was completed by pen with entries every half hour from 7 am to 11 pm. At the end of the data collection period, participants used the application to compress and upload their data to a cloud storage service. A post-experiment questionnaire was then self-administered (Section 4.3.2).

The equipment and daily activity log were retrieved from the participants' residence by an experimenter on the following day.

4.3.2 Data Collection and Measurements

Mobile Application

Inspired by prior notification research [5–7, 9, 32], an Android application was designed to collect information on notification presentation, interactions and to log physiological sensor data. The application saved the data to the device's internal memory to avoid excessive data plan usage.

Table 4.1 Features collected or extracted for each notification presented. Refer to Section 4.3.2 for details on how features were extracted and aggregated.

Information Source	Features
Notification	Presentation time, removal time, active time, reason for removal, source application, presentation screen-lock state, removal screen-lock state
Photoplethysmograph	Heart rate, Inter-beat Interval (IBI), Root mean square of the NN intervals standard deviation (RMSSD), NN interval standard deviation (SDNN), percentage of successive NN intervals that differ by more than 50 and 20 ms (PNN50, PNN20), mean absolute deviation of heart rate (MAD_{HR}), Standard deviation along the two principal axes of the ellipsis fit on the poincaré plot of the NN intervals (SD_1, SD_2), surface of the ellipsis (S), SD_1/SD_2 (SD_{12})
Electrodermal Activity	Number of SCR (nSCR), SCR latency, Sum of SCR amplitudes (AmpSum), mean SCR amplitude (SCR), Integral of SCRs (ISCR), Maximum Phasic Activity (MaxPhasic), Tonic level
Accelerometer	Mean, standard deviation (STD), mean crossing rate (MCR), area under the curve (AUC), skewness, kurtosis, root mean square (RMS), maximum (MAX), entropy

Since Android's notification channel is used by a number of processes that do *not* produce alerts, notification events were filtered [5, 7, 9]. A notification was considered valid if:

1. it originated from an application that produces alert-generating notifications, as defined by Iqbal and Bailey [33], thus excluding, e.g., Google maps, Spotify, Android Downloads.
2. a notification from the same application and with the same unique key was *not* presented in the last 500 ms.

Table 4.1 enumerates the features that were extracted from incoming notifications.

Sensor Data Acquisition and Processing

Participants were instructed to wear the Shimmer3 GSR+ wearable sensor[1] on their non-dominant wrist, and to slide it up their arm so as to avoid collision with the electrode connectors when moving their wrist. Manufacturer-provided velcro-mounted reusable Ag/AgCl dry electrodes were installed on the inside of the proximal phalanxes of the middle and ring fingers. The PPG sensor was secured on the inside of the proximal phalanx of the index finger. The proximal location was chosen to ensure participants were able to engage in their regular activities with minimal physical encumbrance [34]. All raw sensor data were collected, streamed to the participant's Android device and logged to its internal memory at a fixed rate of 64 Hz.

PPG signals were first band-pass filtered (zero lag filtering, 4th order Chebyshev II, 0.5-4 Hz [35]) to remove signal offset and high-frequency artefacts. Heart rate and the most frequently encountered time domain and non-linear heart rate variability features (Table 4.1) were extracted from a 60 s window before and after each notification using Heartpy's built-in processing function [25, 36]. Frequency domain features (e.g., VLF, LF, HF) were not considered in this study due to their questionable validity in measurements sessions that are shorter than one minute [25].

Traditionally, the skin conductance signal is decomposed into its tonic component, a low frequency oscillation independent of specific events, and its phasic component, characterized by abrupt changes in skin conductance level [18]. The phasic and tonic components of the skin conductance signal were extracted using Ledalab's Continuous Decomposition Analysis (CDA) in a Matlab R2020b environment [37] after being downsampled to 16 Hz and segmented in one

[1]Shimmer3 GSR+, Shimmer Engineering

hour segments for computational efficiency reasons. All CDA-specific electrodermal activity features available in Ledalab (Table 4.1) were extracted from a 4 s window starting 1 s after notification presentation [31]. To serve as a notification-specific baseline, features were also extracted from a 4 s window beginning 5 s before the presentation of each notification.

Accelerometer measurements were first low-pass filtered (zero lag filtering, 8th order Butterworth, 10Hz) to only retain information pertinent to the participants' wrist motion [38]. The three measured axes were then aggregated by computing the norm of the acceleration vector. A set of simple descriptive features frequently used in activity change detection was extracted from the acceleration magnitude signal [39] (Table 4.1) within a 15 s window before and after each notification presentation.

To ensure comparability across notifications between participants and highlight relative changes caused by the presentation of notifications, the percentage change between all pre- and post-notification feature values was computed and was used for subsequent reporting and analysis.

Questionnaires

A pre-experiment questionnaire was used to collect demographic information (i.e., age, gender, occupation) and participants' notification setting preferences. The questionnaire used the short version of the smartphone addiction scale (SAS-SV) to quantify the participants' experience of problematic smartphone usage symptoms [40]. In addition, the revised Self-Consciousness Scale (SCS-R) was employed to evaluate the participants' public and private self-consciousness as well as their experience of social anxiety [41].

The post-experiment questionnaire focused on how the data collection day compared to a typical pre-Covid and Covid day with regards to the number of notifications received and amount of physical activity.

4.3.3 Participants and Compensation

Participants were recruited from university classified ads and social media groups. They were given a base compensation of CAD$ 5 for their time. A bonus CAD$ 30 was offered if they complied with the full data collection protocol, i.e., pre-experiment questionnaire, 8 hours of sensor data, post-experiment questionnaire and the activity log.

4.3.4 Hypotheses

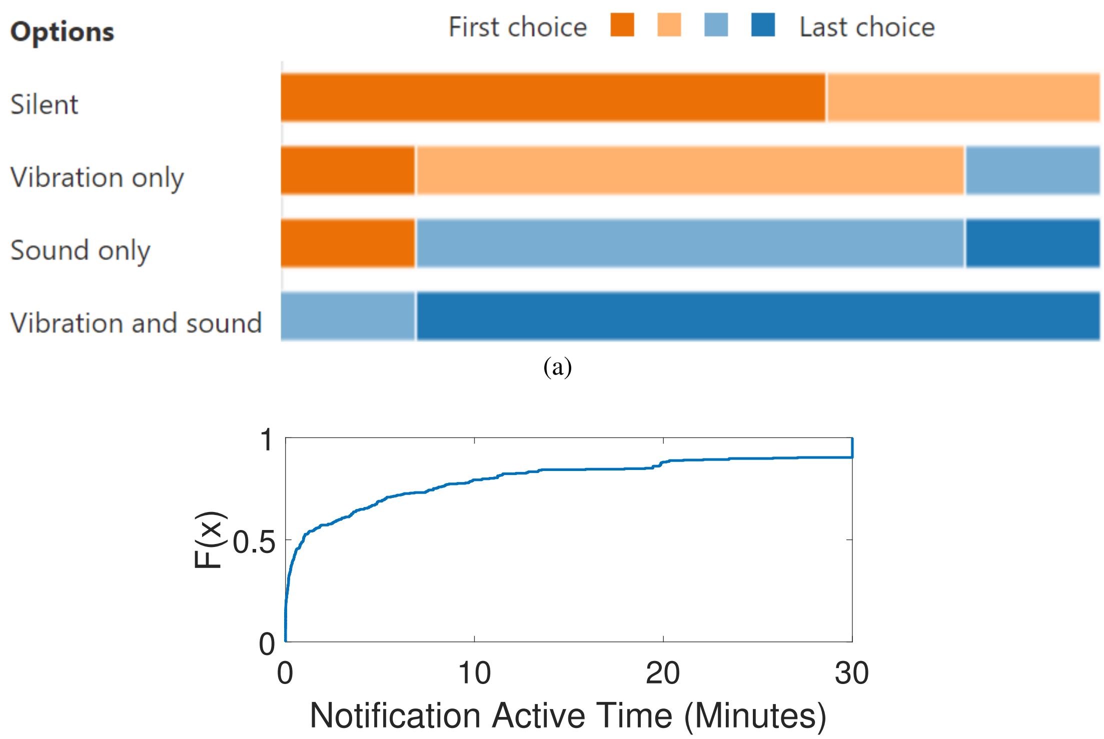
Based on the prior literature on physiological responses to the perception of stimuli and notifications, the following hypotheses were made:

- **H1.1** In response to a notification, the participant's heart rate is expected to increase [19, 21], consistent with a heightened arousal state associated with the anticipated social interaction announced by a notification [9, 17].
- **H1.2** Participants' heart rate variability should decrease in response to the perception of a notification, consistent with the experience of a physiologically stressful or arousing stimulus [19, 20, 28].
- **H1.3** Notification perception should result in more phasic electrodermal activity (nSCR) and larger responses (AmpSum, SCR, ISCR, MaxPhasic), aligned with prior laboratory investigations [9]. In addition, the response should occur within the physiologically valid time window (1 and 4 s post-notification [31]).
- **H2** It is anticipated that participants' wrist motion patterns will significantly vary after the presentation of an alert, even in cases where participants do not immediately interact with their device. This effect is hypothesized to be co-occurring with the reduced primary task performance reported by prior work based on laboratory studies [3].
- **H3** Since successful interaction-less notification perception prediction was previously achieved using electrodermal activity alone [9], we anticipate that the inclusion of supplementary physiological and behavioral signals will significantly increase perception classification performance.

4.4 Results and Discussion

4.4.1 Participants and Notifications

Despite initially recruiting 9 participants, the data from 3 had to be rejected due to unforeseen incompatibility between their smartphone, the data collection application and the Shimmer3 GSR+ sensor. Of the remaining 6 participants, 4 identified as female and 2 as male. Participants reported being between 21 and 28 years old ($\overline{x}$ = 24.6, σ = 2.5). Their primary occupations were: researcher at a local university (1), undergraduate or graduate student (4) and retail worker (1).



(a)

(b)

Fig. 4.1 (a) Participants' self reported most frequently used ringer modes. (b) Notification active time cumulative frequency distribution

As per Fig. 4.1a, participants reported most frequently using their smartphone in silent mode (4/6), followed by sound (1/6) or vibration only (1/6) during their daily life. The least frequently used ringer mode was vibration *and* sound. As such, participants who most often use their device in silent mode had to modify their usual usage pattern for the study, negatively impacting the external validity of the findings presented in this work.

Participants received a median volume of 82 notifications during the 8 hours of data collection (min=40, max=406), significantly above the level reported in prior work, which ranged from 45 to 63 daily notifications [5, 7, 42]. It should be noted that one participant received a particularly high number of notifications (406). To attenuate the impact of that participant's data on the overall data set, we made the decision to uniformly randomly downsample that participant's data to the median daily notification volume of 82. After this manipulation, a total of 441 notifications were considered for analysis.

Based on the difference between presentation and removal times, notifications were attended to at a rate similar to that reported in the literature [5, 7, 42]. Indeed, 50% of presented notifications were active for slightly less than 1 minute, 75% less than 7.8 minutes and a total of 90.2% of notifications spent less than 30 minutes in the notification tray (Figure 4.1b) [7]. The remaining 9.8% of notifications were cleared after 30 minutes, or remained active until the end of the study. Unsurprisingly, a Mann-Whitney U test revealed that notifications presented while the screen was unlocked (60.5% of notifications) were removed significantly faster than those presented from a locked screen (z=-2.86, ranksum=35912.5, p=.004), with median active times of 31.4 seconds and 61.7 seconds respectively.

Considering the limited number of participants, questionnaire elements pertaining to smartphone addiction symptoms (SAS-SV), self-consciousness (SCS-R) and how the data collection days compared to others are not reported in this paper.

4.4.2 RQ1-Physiological Impact of notification perception

The histogram of each physiological feature was visually inspected and approximately matched that of a normal distribution. A series of one-tailed t-tests were used to compare the distribution of the percentage change for each feature against a mean value of zero. The direction of each test's tail was determined such as to match the research hypotheses presented in Section 4.3.4. Considering the number of statistical tests used, all p-values were corrected using the Bonferroni-Holm method for multiple comparisons within each research question to attenuate the probability of Type I errors. To facilitate interpretation of results, the 95% confidence interval on the true distribution mean is presented for each feature's percentage change along with Cohen's d measure of effect size.

RQ1.1 - Heart rate

The observed data did not support our hypothesis that perception of notifications would result in increased heart rate, despite the rationale that such perception would result in increased arousal levels [9, 17], which, in turn, are known to be correlated with cardiac output [19, 21, 29] (t(407)=-2.446885, p=0.992584, ci=[-4.68, -0.51], Cohen's d=-0.12).

Contrary to that assumption, the negative upper and lower bounds of the confidence interval of the true distribution mean suggest that heart rate may have significantly decreased. A one-tailed t-test allowed us to confirm that theory (t(407)=-2.446885, p=0.007, Cohen's d=-0.12). The observed small yet significant decrease in heart rate would be consistent with inhibitional cardiovascular responses that typically accompany increased sensory sensitivity [21, 24], and less arousing experiences [19]. While the first seems fairly intuitive considering the relevance of notifications for smartphone users, the second contradicts findings that had predicted and shown increases in arousal after the presentation of a notification [9, 17]. Based on prior work that studied heart rate changes as a sign of orienting and defense reflexes, it is possible that this disagreement could be a consequence of co-occurring opposed activational and inhibitional cardiovascular responses. Such response pattern could be caused by conflicting negative and positive notification experience components, the context in which the signal is presented as well as differences in participants' coping mechanisms [21, 24]. Furthermore, the discrepancy between the expected and measured changes might be due to the fact that most of the literature on the topic was based on evidence acquired in controlled laboratory studies versus in-the-wild measurements. Both environments present participants with significantly different sensory, social and activity contexts that may introduce variations in physiological baseline levels and response patterns. Further discussion of the observed heart rate deceleration, considering changes in other physiological signals is presented in Section 4.4.2.

Independently of its interpretation, the fact that a small yet significant heart rate deceleration was observed in response to the presentation of a notification warrants further investigation and shows that heart rate measurements could potentially be used to further understand users' perception of notifications. It is also indicative of a promising predictor of notification perception (Section 4.4.4).

RQ1.2 - Heart Rate Variability

With the body of literature documenting the negative impact of notifications, we hypothesized that heart rate variability features would respond to notifications similarly to physiologically stressful and arousing stimuli. This would typically result in a decrease in heart variability (RMSSD, SDNN, MAD_{HR}, PNN20, PNN50, SD_1, SD_2, S and SD_{12}). Percentage changes for MAD_{HR}, SDNN and PNN20 were found to be statistically significantly lower than 0, partially supporting our research hypothesis. It should be noted that with the exception of the inter-beat interval (IBI) and S, the lower and upper bounds of the 95% confidence interval of the true distribution mean for all features were found to be inferior to 0 with small to medium effect sizes. This hints at the possibility that statistical significance might be attained with a larger sample size.

These changes in heart rate variability points towards a decrease in parasympathetic nervous system activity, commonly associated with digestion, energy conservation and the promotion of

Table 4.2 Results from the Bonferroni-Holm-corrected one-tailed t-tests investigating whether mean percentage change in heart rate variability features between before and after the presentation of a notification differ from zero. Confidence intervals are the lower and upper bounds of the true distribution mean at alpha .05 (%Δ).

Feature	95% CI Mean		Cohen's d	df	t stat	p	Adj.p
%Δ MAD_{HR}	-198	-46.2	-0.16	407	-3.161129	.001	**.008**
%Δ SDNN	-266	-42.0	-0.13	407	-2.718062	.003	**.031**
%Δ PNN20	-2.81	-0.34	-0.13	388	-2.515147	.006	**.049**
%Δ SD_1	-195	-20.5	-0.12	388	-2.427633	.007	.054
%Δ SD_{12}	-6.45	-0.56	-0.12	385	-2.341825	.010	.059
%Δ RMSSD	-250	-6.78	-0.11	388	-2.075531	.019	.135
%Δ SD_2	-182	-1.16	-0.10	388	-1.991300	.024	.141
%Δ S	-58e+3	2.9e+3	-0.09	388	-1.773434	.038	.192
%Δ IBI	-3.84	0.22	-0.09	407	-1.749911	.040	.162
%Δ PNN50	-2.21	-0.75	-0.05	388	-0.972425	.166	.497

rest [43]. In addition, while it is impossible to isolate the underlying reason(s) of the observed decrease in heart rate variability outside of laboratory conditions, similar changes have been observed in response to increases in task demand [26], cognitive load [27], psychological stress level [20] and emotional arousal [19]. Without further contextual control, these observations partially support the existing literature on the arousing nature of notifications [9, 17], their negative impact on attention [3], and their identification as significant stressors [4]. However, considering the lack of global agreement between the considered features, further research would be required to confirm any such relationship outside of laboratory conditions with certainty.

RQ1.3 - Electrodermal Activity

As previously mentioned, Fortin et al. already demonstrated that the perception of their own smartphone notification caused participants to exhibit significantly larger maximum phasic skin conductance activity [9]. Based on that observation, they suggested that the perception of a notification results in a heightened arousal state. However, their work was conducted in laboratory conditions, which significantly limits the generalizability of their results. As such, we attempted to replicate their findings under more realistic conditions and were interested in expanding the investigation to additional SCR characteristics beyond the maximum phasic activity.

Considering adjusted p-values, all features except SCR and the Tonic skin conductance level were found to statistically significantly increase between the pre- and post-stimulation measurements (Table 4.3). Our results successfully replicate the findings presented by Fortin et al. by showing that all phasic electrodermal activity features, with the exception of the mean SCR amplitude (SCR), statistically significantly increase after the presentation of a notification. Indeed, the considered features increased by 0.6 to 18% of their pre-stimulation levels on average. Such an increase in the number of SCR (nSCR) and their amplitude (PhasicMax, ISCR, AmpSum) is typically associated with sympathetic nervous system activation and a heightened arousal state in the psychophysiology and human-computer interaction literature [18, 21, 30]. However, due to the uncontrolled nature of the study, the observed increase in phasic EDA could potentially be attributed to other events occurring at approximately the same time as the presentation of a notification (e.g., door slamming, change in temperature). While it is important to acknowledge that possibility, we argue that it is highly improbable that such an event took place sufficiently frequently, and at exactly the right time, to significantly shift the distribution means. We therefore claim that the skin conductance features considered in this work support the hypothesis that notifications induce high arousal states [9, 17].

Table 4.3 Results from statistical analysis investigating whether mean percentage change in skin conductance features between before and after the presentation of a notification are greater than zero. Confidence intervals are the lower and upper bounds of the true distribution mean in percentage change (%Δ).

Feature	95% CI	Mean	Cohen's d	df	t stat	p	Adj.p
%Δ Latency	866	925	3.41	308	59.85732	.001	**<.001**
%Δ ISCR	1.99	11.7	0.15	363	2.772188	.003	**.018**
%Δ nSCR	2.49	17.9	0.14	363	2.602344	.005	**.024**
%Δ PhasicMax	1.09	9.45	0.13	363	2.480880	.007	**.027**
%Δ AmpSum	0.62	7.16	0.12	363	2.340251	.010	**.030**
%Δ SCR	-0.08	0.22	0.05	363	0.912958	.181	.362
%Δ Tonic	-0.70	1.66	0.04	363	0.790695	.215	.215

In addition to successfully replicating the findings presented by Fortin et al. outside of the laboratory, these results reinforce the notion that features extracted from skin conductance measurements could constitute reliable notification perception predictors.

Physiological Signals Reconciliation

The most impactful benefit of measuring and analysing multiple physiological signal channels is the researcher's ability to combine the findings and identify conflicting and reinforcing response patterns.

For example, in the current study, only certain HRV features were found to significantly decrease, consistent with the hypothesized increase in arousal and stress [17, 19, 20]. Similarly, phasic electrodermal activity features, typically used to assess the experience of stress and arousal [29, 30] in and outside of the laboratory were found to significantly increase post-notification. On the other hand, by decelerating, the heart rate went completely against that initial arousal hypothesis. Considering the lack of general agreement between the different physiological channels, our results do not allow us to claim that the perception of a notification causes an increase in arousal outside of the laboratory. That being said, given the partial support of that hypothesis by two out of the three signals and prior laboratory evidence, we argue that the effect remains extremely plausible and should be further investigated.

This exercise also applies to the reported heart rate deceleration. The observed decrease in heart rate variability features is generally consistent with a reduction in parasympathetic (PNS) tone [25]. Without consideration for sympathetic nervous system (SNS) activity, this would typically result in an increase in heart rate [21]. In addition, SD_{12}, often employed as an indicator of balance between PNS and SNS activity decreased, suggesting a relative increase in sympathetic tone [25]. That heightened SNS activity is further supported by a significant increase in phasic skin conductance activity [18]. Together, the decrease in parasympathetic and increase in sympathetic tone should have resulted in an elevated heart rate.

A possible explanation for the heart rate deceleration could be that the perception of a notification has an effect on respiratory rate (e.g., breath holding, sudden exhalation), a physiological signal that was not controlled for in this work. In such a case, the altered respiration patterns would have a direct impact on heart rate, a phenomenon known as respiratory sinus arrhythmia. Indeed, enforced by this mechanism, heart rate accelerates during inhalation and decelerates during exhalation [44]. Holding one's breath after the perception of a notification would typically result in a deceleration of heart rate [44]. However, it is impossible to validate this hypothesis without accurate respiratory rate measurements. Finally, the observed heart rate deceleration could simply be part of the orienting reflex [24]. In such a case, the deceleration associated with the orienting reflex could overtake a weak heart rate increase that would have been associated with a heightened arousal state. In absence of a conclusive interpretation, the signal can be used most reliably as an indicator that a notification was detected, without offering further insights on how it impacted the participant's internal state.

4.4.3 RQ2-Activity disruption

Prior literature has shown that the presentation of a notification significantly disrupts users while they are engaged in a primary activity. Researchers typically demonstrate such effects by closely monitoring participants' behavior while they are engaged in controlled tasks [3], or by using experience sampling methods to determine the degree to which notifications were interrupting. As per our second research question, we were instead interested in objectively measuring how a notification's interruption manifests itself through changes in users' wrist motion following the presentation of the notification, independently of whether the alert is acted on.

The normality of all features' distribution was first confirmed by visually inspecting their histograms. A series of two-tailed t-tests were used to investigate whether the mean percentage changes of each feature differed from zero. The two-tailed version of the test was chosen as it is impossible to predict, without knowing what activities participants are engaged in, the direction in which each feature would vary. All p-values were corrected using the Bonferroni-Holm method for multiple comparisons to attenuate the probability of false positives. To facilitate interpretation of the results, the confidence interval of the true distribution mean for each feature's percentage change is presented alongside Cohen's d standardized measure of effect size.

After correcting p-values, the percentage change of the mean crossing rate, kurtosis and standard deviation of the wrist's acceleration magnitude were found to have statistically significantly varied after the presentation of a notification (Table 4.4). The observed medium negative effect sizes, as well as the negative confidence interval bounds for the mean of most features, suggest a decrease in overall amount of wrist motion following the presentation of a notification. These observations would be consistent with a disruption of users' primary activity to reallocate their attention towards the notification or hold their device in a static pose while attending to the alert.

By considering only accelerometer measurements occurring around notification presentation, it is impossible to determine whether the observed changes in motions were caused by users attending to their notification or more subtle variations in motion introduced by a disruption of users' primary activity. We therefore apply the same statistical testing approach only for notifications that were cleared *after* the 15 s period during which accelerometer measurements were

Table 4.4 Results from statistical analysis investigating whether percentage change in accelerometer features between before and after the presentation of a notification differ from zero. Confidence intervals are the lower and upper bounds of the true distribution mean at alpha .05 (%Δ).

Feature	95% CI Mean		Cohen's d	df	t stat	p	Adj.p
%Δ mean crossing rate	-113	-56.3	-0.28	440	-5.877802	.0001	**<.0001**
%Δ kurtosis	-68.6	-32.9	-0.27	440	-5.578155	.0001	**<.0001**
%Δ standard deviation	-9.65	-3.16	-0.18	440	-3.883376	.0001	**.0008**
%Δ max	-3.15	-0.437	-0.12	440	-2.597415	.010	.058
%Δ area under the curve	-1.57	-0.093	-0.11	440	-2.212221	.027	.137
%Δ root mean square	-1.38	-0.062	-0.10	440	-2.149240	.032	.129
%Δ mean	-1.38	-0.045	-0.10	440	-2.098835	.036	.096
%Δ entropy	-0.028	-0.000	-0.07	440	-1.454448	.147	.293
%Δ skewness	-79.4	120	-0.07	440	0.400298	.689	.689

acquired. The rationale is that if a notification was cleared beyond the measurement period, the accelerometer signal is unlikely to capture the motion associated with the consumption of the alert. In that second analysis, the percentage change of the mean crossing rate (t(263)=4.866539, Adj. p=<.0001, ci = [58.4, 138], Cohen's d=0.30) and kurtosis (t(263)=4.285884, Adj. p<.0001, ci = [30.3, 81.8], Cohen's d=0.26) of the wrist's acceleration magnitude were also found to statistically significantly vary. Interestingly, even though almost the same features were found to change significantly after the presentation of a notification, the direction in which they varied is opposite to previous observations. This indicates an increase in extreme acceleration values and a *different* change in motion patterns following the presentation of a notification, instead of a decrease as was observed when notifications were immediately attended to.

Knowing changes in wrist motions occur even when notifications are not attended to within the measurement period, we investigate the reasons behind the previously observed decrease in wrist activity. The same statistical testing pipeline was applied to the comparison of wrist motion in the 15 seconds leading to and following the *removal* of each notification, as opposed to its presentation. By focusing on the removal of notifications, these measurements are representative of actions associated with attending to a notification. As expected, the acceleration's magnitude mean crossing rate (t(221)=-4.107649, Adj. p<.0001, ci = [-218, -76.8], Cohen's d=-0.28), and kurtosis (t(221)=-4.745734, Adj. p<.0001, ci = [-95.2, -39.3], Cohen's d=-0.32) were found to significantly decrease after the removal of a notification. This supports the theory that the decrease in wrist activity observed after the perception of a notification can most likely be attributed to actions associated with attending to a notification, the smartphone or the same application accessed from another device.

Confirming our research hypothesis, we successfully identified two features: kurtosis and mean crossing rate, that were found to be associated with changes in wrist motion patterns after the presentation of a notification and that are independent of whether participants attended to their notification immediately after its presentation. While it is difficult to interpret the meaning of these changes without knowing more about the activities during which they were observed, the fact that they exist is indicative of a disruption of how the participant was executing their primary task after the presentation of a notification. Knowledge that the perception of a notification interrupts smartphone users is not new. However, these are the first reports of the manifestation of this disruption on users' motion *in situ*, without control over participants' activities or reliance on subjective self-reports.

4.4.4 RQ3 - Notification Perception Prediction

Knowledge that a notification was perceived by a user can be used to better contextualize their notification experience, but also to provide technology with the means to adapt its information presentation strategies based on users' perception [9]. Extending the notification perception classifier presented by Fortin et al. relying solely on the maximum phasic activity collected in laboratory conditions, we evaluate which of the sensing channels (i.e., EDA, PPG or accelerometer) offers the best predictors to determine whether a notification was perceived *in situ*.

Leave one subject out (LOSO) cross-validation would typically be used to evaluate user-independent perception prediction model performance. However, the current data set is not sufficiently large to apply this technique. We instead opt to hold out 10% of the data as a test set (uniformly randomly sampled), and use 10-fold cross-validation within the remaining 90% to train and optimize hyperparameters of the following notification perception prediction models:

- M_{EDA}: Skin conductance features only
- $M_{HR/V}$: Heart rate and heart rate variability features only
- M_{ACC}: Accelerometer-based features only
- M_{ALL}: All extracted features

- $M_{0 \notin MeanCI}$: Only features where the 95% confidence interval on the true distribution mean *excludes* 0.

Each model was trained individually using Matlab's optimizable ensemble method, allowing for the automatic determination of the best performing ensemble method among Bagging, GentleBoost, LogitBoost, AdaBoost and RUSBoost. Hyperparameter optimization was achieved for each model using Matlab's Bayesian optimization functions.

The implementation of any machine learning solution requires a representative number of positive (perceived) and negative (missed) notifications to adequately model the data. Since participants were instructed to put their device in a non-silent ringer mode *and* keep it physically close to them during data collection, all presented notifications are assumed to be perceived. Following the approach proposed by Fortin et al., synthetic "missed" notifications were introduced in the log file 60 s before each actual notification presentation (Figure 4.2). The introduction of "missed" notifications is based on the assumption that if a notification was never presented (i.e., not perceived), it is impossible for it to induce physiological and behavioral changes, and is therefore equivalent to sampling the noise of the different signals.The 60 s interval was chosen as it is the smallest period that ensures no post-notification measurement overlap between the synthetic and real notification, yet is sufficiently close to ensure contextual similarity (Figure 4.2). It is possible that the measurement periods of a synthetic and the previous real notification overlapped. However, we believe this scenario could also occur with consecutive notifications in real life and therefore did not reject or otherwise modify synthetic notifications when these occurred.

We first consider the models' overall performance by inspecting their receiver operating characteristic (ROC) curves (Figure 4.3). ROC curves are effective at representing a model's performance by considering a wide range of classification thresholds. A model that performs perfectly

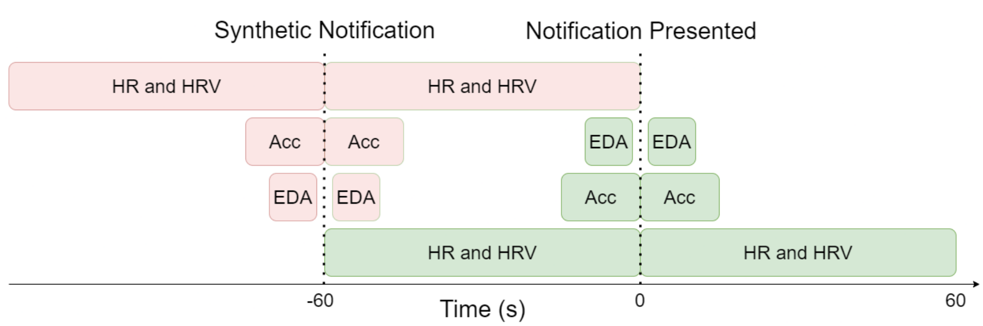


Fig. 4.2 Schematic representation of time windows in which features and their baselines were computed for presented (green) and synthetic notifications (red).

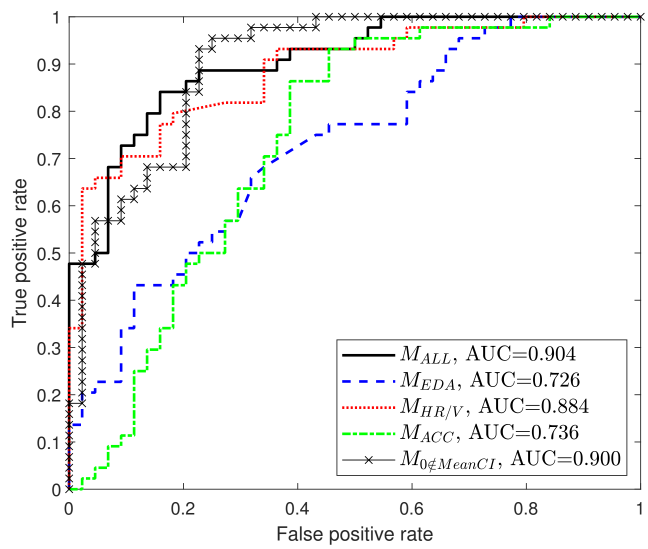


Fig. 4.3 Receiver Operating Characteristic (ROC) curve obtained by predicting data from the test set for the five trained models

would reach the upper left corner with a true positive rate of 1 and a false positive rate of zero. On the other hand, a random binary predictor would have a ROC curve that traverses the plot on its diagonal. Visually, all models seem to perform well above chance level. However, M_{EDA} and M_{ACC} are significantly closer to the plot's center line than $M_{0 \notin MeanCI}$, M_{ALL} and $M_{HR/V}$ with areas under the curve of 0.726 and 0.736 respectively. A statistical comparison of the ROC curves' AUC was achieved using the bootstrap method (R 4.0.4, pROC 1.17.0.1, n.boot=2000, boot.stratified=1) (Table 4.5). These results highlighted the absence of meaningful difference in performance between $M_{0 \notin MeanCI}$, M_{ALL} and $M_{HR/V}$, as well as between M_{EDA} and M_{ACC}. However, visually and statistically significant differences were found between all elements of these two groups of models. These results suggest that PPG-based heart rate and heart rate variability features offer the strongest predictors of notification perception among the considered sensing channels. Interestingly, the absence of a difference in AUC between $M_{0 \notin MeanCI}$, M_{ALL} and $M_{HR/V}$ leads to the conclusion that EDA- and accelerometer-based features provide little information that is not already present in HR and HRV features.

To further explore and compare models, an optimal cutpoint was selected so as to minimize the distance between the ROC curve and the (0,1) coordinate. It is important to mention that these classification thresholds were selected to depict the general system performance. Specific applications may benefit from using different thresholds, for example to minimize the false positive

Table 4.5 Results from the statistical comparison of ROC curves' area under the curve using the bootstrap method (n.boot=2000, boot.stratified=1).

Models Compared	D-stat	p
M_{ALL} - $M_{0\notin meanCI}$	0.1501	.881
M_{ALL} - $M_{HR/V}$	0.6563	.512
M_{ALL} - M_{ACC}	3.0510	**.002**
M_{ALL} - M_{EDA}	3.7020	**<.001**
$M_{0\notin meanCI}$ - $M_{HR/V}$	0.4825	.670
$M_{0\notin meanCI}$ - M_{ACC}	3.0949	**.002**
$M_{0\notin meanCI}$ - M_{EDA}	3.3014	**<.001**
$M_{HR/V}$ - M_{ACC}	2.4176	**.015**
$M_{HR/V}$ - M_{EDA}	2.6379	**.008**
M_{ACC} - M_{EDA}	0.1424	.887

rate in safety critical scenarios.

Models' accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) support the observations made from the ROC curves and their AUC (Table 4.6). PPG-extracted heart rate and heart rate variability features were found to be the strongest notification perception predictors, yielding an overall model accuracy of 76.1%, 4.5% and 10.2% above accelerometer- and EDA-based features respectively. Electrodermal activity and wrist-motions were again found to be the least informative signals when attempting to classify notification perception.

Table 4.6 Trained models classification performance.

Model	Accuracy	Sensitivity	Specificity	PPV	NPV	AUC
$M_{0\notin MeanCI}$	**0.818**	**0.841**	**0.796**	**0.804**	**0.833**	0.900
M_{ALL}	0.773	0.886	0.659	0.722	0.853	**0.904**
$M_{HR/V}$	0.761	0.818	0.705	0.735	0.795	0.884
M_{ACC}	0.716	0.818	0.614	0.679	0.771	0.736
M_{EDA}	0.659	0.750	0.568	0.635	0.694	0.726

Fortin et al. previously reported an overall classification accuracy of 61% using a notification perception prediction model based exclusively on the maximum phasic activity (PhasicMax) of the skin conductance signal. Through the inclusion and combination of additional physiological and behavioral signals (M_{ALL} and $M_{0\notin meanCI}$), we report overall model accuracies of 77.3 and 81.2% respectively. While the models presented in this work seem to significantly outperform the one presented by Fortin et al., it is important to note that significantly different data collection conditions, sample size and cross validation techniques were used to produce these results. As such, we refrain from drawing conclusions from the comparison of their respective classification performance metrics.

Using only EDA features, our model reached a classification accuracy of 65.9%, marginally above that reported in the literature. It must be noted that the two models were trained from data collected in significantly different conditions and were not evaluated using the same cross-validation technique (LOSO CV vs 10% hold out). On the one hand, an increase in performance is expected when a model is trained and evaluated from samples originating from a same subject. On the other hand, a decrease in classification accuracy is typically observed when taking a system outside of laboratory conditions. This is typically due to an increase in noise and motion artefacts contaminating the sensor data. Considering the possible influence of these methodological discrepancies, we believe the difference in model performance is not sufficiently large to claim that the inclusion of more EDA-related features is beneficial to the notification perception prediction problem. We instead find it extremely promising for the proposed technique that above chance classification performance was observed both in and outside of laboratory conditions.

The results presented in this work have serious implications for future applications of the proposed notification perception confirmation technique in both notification research and end-user applications. The fact that electrodermal activity was found to be the worst signal to predict notification perception is particularly interesting, as it is also the only signal to require the use of a dedicated physiological sensor such as the Shimmer3 GSR+. Indeed, most commercially available smartwatches and fitness trackers are already equipped with an optical heart rate sensor and accelerometer. We therefore trained a supplementary model, $M_{HR|V+ACC}$, using only HR, HRV and accelerometer-based features, to explore how such notification perception classification technique could perform given today's device constraints. An overall classification accuracy of 0.773, sensitivity of 0.818, specificity of 0.727 and area under the ROC curve of 0.909 were observed. Promisingly, this last model performs practically as well as M_{ALL}, with a slightly lower sensitivity and higher specificity.

Through the evaluation and comparison of the different models' performance, we demonstrated the value of expanding sensing channels beyond electrodermal activity and confirmed our research hypothesis stating that the inclusion of heart rate, heart rate variability and accelerometer features would be beneficial to notification perception classification performance. In fact, we discovered that EDA measurements are not required at all to reach optimal classification performance. This suggests that it could technically be possible to deploy a notification perception classification system using optical heart rate sensors and accelerometers already available in commercially available smartwatches and fitness trackers.

Despite discovering that the EDA signal is not required to reach the best perception classification performance, researchers should not refrain from collecting it if they have the opportunity to do so. Indeed, the notification experience goes far beyond binary perception classification. Collecting EDA in addition to heart rate, heart rate variability and wrist-motion measurements provides complementary insights as they are known to capture different perspectives of the participants' autonomic nervous system, behavior and psychophysiological state (Section 4.4.2).

4.5 Limitations

While this paper presents novel findings on the physiological impact of notifications, it suffers from limitations that need to be acknowledged and addressed in future work.

Data collection for this study was executed amidst a global pandemic. As such, even though the measurements were made *in situ*, the fact that participants spent most of their day at home is not representative of a "normal", as in pre-COVID, daily amount of physical activity. This may have allowed the collection of cleaner signals than otherwise would have been possible. Public health measures and the negative affects caused by the pandemic itself may have also significantly impacted participants' psychological state. Indeed, a recent study by Varma et al. reported that a large proportion of young adults were experiencing severe stress, anxiety, depression and general psychological distress since the beginning of the pandemic [45]. This may have changed baseline levels of physiological signals as well as participants' response patterns and behaviors. Beyond its impact on the generalizability of the findings, the pandemic and the public health measures in place also significantly impacted the participant recruitment and experiment execution process. While this is beyond the scope of this work, we hypothesize that prospective participants were much less likely to inquire about participating in the study than for similar experiments in previous years, potentially due to an increase in perceived risks of virus propagation. In addition, the city where the study took place was under an 8 pm to 5 am curfew, which extended the end-to-end study duration to 3 days, with only one day used for data collection. This may have negatively influenced the perceived value of the monetary compensation for participants, further impeding the recruitment process. As such, this paper reports on physiological data collected from only 6 individuals. Even though we believe we have collected a sufficient amount of notifications to conduct meaningful statistical analyses, the results cannot be generalized to the global population. This is particularly true given the limited age range of participants.

Measurements were collected over a period of approximately 8 hours. The decision to collect data over a single versus multiple days or weeks stemmed from the complexity for participants to repeatedly adequately position the physiological sensor, as well as limitations of the sensor's battery life and smartphones' storage capacity. The short duration of the study means that the results cannot capture physiological changes happening across all of representative daily activities or the impact of psychological states experienced over longer time frames (i.e., weeks, months, years). Different activities come with varying levels of user engagement and cognitive load that directly impact participants' physiological state [18]. In this work user activities were collected using a daily activity log, but were not analyzed because breaking down physiological responses between different activities would have been slicing the data too thin to conduct a meaningful analysis. Despite the small sample size, we believe the amount of notifications collected is sufficient to demonstrate the potential benefits brought forward by the proposed notification research methodology.

Even though they were instructed to set their smartphone to a non-silent ringer mode and to keep it close to them at all time, it is possible that some notifications were not perceived by participants. For example, they may have been fully immersed in a primary activity and not feel or hear a notification. Alternatively, listening to music, watching a video or engaging in other sensory activities might have masked the signals. Regardless of the reason, false positive notifications are a possibility and they directly reduce the power of our analyses and classification models by introducing data points where no practically meaningful or even completely opposite changes in physiological changes occurred. That being said, we believe that this is a small price to pay in comparison to the risk of contaminating physiological signals with experimentally induced responses, i.e., press of a button when a notification is perceived.

Finally, being the first researchers to study the physiological impact of notifications *in situ*, we made the conscious decision to report on a large number of physiological features with the intent to document the observed responses as broadly as possible and to offer a reference point for future physiologicals-based notification research. Typically, psychology, affective and physiological computing researchers would only measure and report on features that are known to be strong indices of the psychological construct they are investigating. Due to the large numbers of statistical tests used, we applied the Bonferroni-Holm method to attenuate risks of type I errors. In doing so, it is possible that type II errors were introduced. We therefore strongly encourage readers interested in employing these findings to inform their own research to carefully consider the 95% confidence interval on the true distribution mean and Cohen's d for each feature to better understand the magnitude of the effect and its variability.

4.6 Conclusion

This paper introduced a novel notification research methodology combining application-based notification logging with the passive collection of physiological signals. In an full day *in situ* study, this methodology was employed to demonstrate for the first time that the perception of smartphone notifications causes a deceleration of heart rate and a decrease in heart rate variability features. In addition, we were able to fully replicate the previously reported increase in phasic electrodermal activity following the presentation of a notification [9] outside of laboratory conditions. Beyond physiological signals, we reported the first evidence of the impact of notification disruption on users' wrist motion, without control over participants' activities or reliance on subjective self-reports typically used in the literature.

The notification perception classification system previously introduced by Fortin et al. was extended with heart rate, heart rate variability and wrist-motion measurements. In doing so, the performance benefits of supplementary physiological and behavioral channels were demonstrated and PPG-based features were identified as the single best channel to consider for this classification problem. Even though it has not yet been validated, these findings suggest that such perception confirmation system could potentially be deployed using today's commercially available smartwatches and fitness trackers, which are largely equipped with optical heart rate sensors and accelerometers.

4.7 Acknowledgments

This research was financially supported by the Fonds de Recherche Québec Nature et Technologie (FRQNT B2X-211767), the National Sciences and Engineering Council of Canada (RGPIN-

2017-05013) and a McGill Engine TechAccelR Innovation grant.

References

[1] K. Kushlev, J. Proulx, and E. 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*, ser. CHI ’16. New York, NY, USA: ACM, 2016, pp. 1011–1020. [Online]. Available: http://doi.acm.org/10.1145/2858036.2858359

[2] E. Kanjo, D. J. Kuss, and C. S. Ang, “NotiMind: Utilizing Responses to Smart Phone Notifications as Affective Sensors,” *IEEE Access*, vol. 5, pp. 22 023–22 035, 2017.

[3] C. Stothart, A. Mitchum, and C. Yehnert, “The attentional cost of receiving a cell phone notification.” Stothart, Cary: Department of Psychology, Florida State University, 1107 West Call Street, Tallahassee, FL, US, 32306-4301, stothartpsy.fsu.edu, pp. 893–897, 2015.

[4] S. Yoon and K.-p. Lee, “A Study on Notification System Design of Smartphone Messenger Considering the User ’ s Stress,” *Archives of design research*, vol. 28, no. 2, pp. 77–91, 2015.

[5] M. Pielot, A. Vradi, and S. Park, “Dismissed!: A Detailed Exploration of How Mobile Phone Users Handle Push Notifications,” in *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, ser. MobileHCI ’18. New York, NY, USA: ACM, 2018, pp. 3:1—-3:11. [Online]. Available: http://doi.acm.org/10.1145/3229434.3229445

[6] D. Weber, A. Voit, and N. Henze, “Clear All: A Large-Scale Observational Study on Mobile Notification Drawers,” in *Proceedings of Mensch Und Computer 2019*, ser. MuC’19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 361–372. [Online]. Available: https://doi.org/10.1145/3340764.3340765

[7] M. Pielot, K. Church, and R. 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*, ser. MobileHCI ’14. New York, NY, USA: ACM, 2014, pp. 233–242. [Online]. Available: http://doi.acm.org/10.1145/2628363.2628364

[8] T. Okoshi, K. Tsubouchi, M. Taji, T. Ichikawa, and H. 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 2017*, 2017, pp. 100–110.

[9] P. E. Fortin, E. Sulmont, and J. Cooperstock, “Detecting Perception of Smartphone Notifications using Skin Conductance Responses,” in *Conference on Human Factors in Computing Systems - Proceedings*. New-Orleans, LA: ACM, 2019, p. 9.

[10] T. Okoshi, J. Ramos, H. Nozaki, J. Nakazawa, A. K. Dey, and H. Tokuda, “Attelia: Reducing user’s cognitive load due to interruptive notifications on smart phones,” in *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, mar 2015, pp. 96–104.

[11] A. Mehrotra, M. Musolesi, R. Hendley, and V. Pejovic, “Designing content-driven intelligent notification mechanisms for mobile applications,” in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp ’15)*, ser. UbiComp ’15. New York, NY, USA: ACM, 2015, pp. 813–824. [Online]. Available: http://doi.acm.org/10.1145/2750858.2807544

[12] F. Yuan, X. Gao, and J. Lindqvist, “How busy are you? Predicting the interruptibility intensity of mobile users,” in *Conference on Human Factors in Computing Systems - Proceedings*, ser. CHI ’17, vol. 2017-May. New York, NY, USA: ACM, 2017, pp. 5346–5360. [Online]. Available: http://doi.acm.org/10.1145/3025453.3025946

[13] E. Toch, H. Chassidim, and T. Hatuka, “Can you Turn it Off?” *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW2, pp. 1–18, oct 2020. [Online]. Available: https://doi.org/10.1145/3415162

[14] K. N. Truong, T. Shihipar, and D. J. Wigdor, “Slide to X: Unlocking the Potential of Smartphone Unlocking,” *Proceedings of the 32nd annual ACM conference on Human factors in computing systems.*, pp. 3635–3644, 2014.

[15] P. E. Fortin, J. R. Blum, and J. R. Cooperstock, “Towards consistent haptic coupling with HaptiStrap: Doing better than ”Tight yet comfortable”,” in *UIST 2019 Adjunct - Adjunct Publication of the 32nd Annual ACM Symposium on User Interface Software and Technology*. New Orleans, USA: ACM, oct 2019, pp. 69–71.

[16] T. Westermann, S. Möller, and I. 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*, ser. MobileHCI ’15. New York, NY, USA: Association for Computing Machinery, 2015, pp. 652–659. [Online]. Available: https://doi.org/10.1145/2786567.2793684

[17] S. P. L. Veissière and M. Stendel, “Hypernatural Monitoring: A Social Rehearsal Account of Smartphone Addiction,” *Frontiers in Psychology*, vol. 9, p. 141, 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2018.00141

[18] W. Boucsein, "Electrodermal activity," New York, 2012. [Online]. Available: http://site.ebrary.com/id/10517958

[19] K. Schaaff and M. T. Adam, "Measuring emotional arousal for online applications: Evaluation of ultra-short term heart rate variability measures," in *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, 2013, pp. 362–368.

[20] J. P. Delaney and D. A. Brodie, "Effects of short-term psychological stress on the time and frequency domains of heart-rate variability," *Perceptual and Motor Skills*, vol. 91, no. 2, pp. 515–524, oct 2000. [Online]. Available: https://doi.org/10.2466/pms.2000.91.2.515

[21] J. T. Cacioppo, L. G. Tassinary, and G. Berntson, *Handbook of Psychophysiology*. Cambridge University Press, 2007. [Online]. Available: http://ebooks.cambridge.org/ref/id/CBO9780511546396

[22] M. L. Loggia, M. Juneau, and M. C. Bushnell, "Autonomic responses to heat pain: Heart rate, skin conductance, and their relation to verbal ratings and stimulus intensity," *PAIN®*, vol. 152, no. 3, pp. 592–598, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0304395910007177

[23] M. Nardelli, G. Valenza, A. Greco, A. Lanatá, E. P. Scilingo, and R. Bailón, "Quantifying the lagged Poincaré plot geometry of ultrashort heart rate variability series: automatic recognition of odor hedonic tone," *Medical and Biological Engineering and Computing*, vol. 58, no. 5, pp. 1099–1112, 2020. [Online]. Available: https://doi.org/10.1007/s11517-019-02095-7

[24] F. K. Graham and R. K. Clifton, "Heart-rate change as a component of the orienting response." US, pp. 305–320, 1966.

[25] F. Shaffer and J. P. Ginsberg, "An Overview of Heart Rate Variability Metrics and Norms," *Frontiers in Public Health*, vol. 5, p. 258, sep 2017. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/29034226https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5624990/

[26] A. Luque-Casado, J. C. Perales, D. Cárdenas, and D. Sanabria, "Heart rate variability and cognitive processing: The autonomic response to task demands," *Biological Psychology*, vol. 113, pp. 83–90, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301051115300880

[27] S. Solhjoo, M. C. Haigney, E. McBee, J. J. van Merrienboer, L. Schuwirth, A. R. Artino, A. Battista, T. A. Ratcliffe, H. D. Lee, and S. J. Durning, "Heart Rate and Heart Rate Variability Correlate with Clinical Reasoning Performance and Self-Reported Measures of Cognitive Load," *Scientific Reports*, vol. 9, no. 1, p. 14668, 2019. [Online]. Available: https://doi.org/10.1038/s41598-019-50280-3

[28] L. Salahuddin, J. Cho, M. G. Jeong, and D. Kim, "Ultra Short Term Analysis of Heart Rate Variability for Monitoring Mental Stress in Mobile Settings," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007, pp. 4656–4659.

[29] C. A. Wang, T. Baird, J. Huang, J. D. Coutinho, D. C. Brien, and D. P. Munoz, "Arousal Effects on Pupil Size, Heart Rate, and Skin Conductance in an Emotional Face Task," *Frontiers in Neurology*, vol. 9, p. 1029, 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/fneur.2018.01029

[30] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," pp. 156–166, 2005.

[31] R. A. Lockhart, "Interrelations Between Amplitude, Latency, Rise Time, and the Edelberg Recovery Measure of the Galvanic Skin Response," *Psychophysiology*, vol. 9, no. 4, pp. 437–442, 1972. [Online]. Available: http://dx.doi.org/10.1111/j.1469-8986.1972.tb01791.x

[32] D. Weber, A. Voit, G. Kollotzek, and N. Henze, "Annotif: A system for annotating mobile notifications in user studies," in *ACM International Conference Proceeding Series*, ser. MUM '19. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: https://doi.org/10.1145/3365610.3365611

[33] S. T. Iqbal and B. P. Bailey, "Investigating the Effectiveness of Mental Workload As a Predictor of Opportune Moments for Interruption," in *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '05. New York, NY, USA: ACM, 2005, pp. 1489–1492. [Online]. Available: http://doi.acm.org/10.1145/1056808.1056948

[34] M. van Dooren, J. J. G. G.-J. de Vries, and J. H. Janssen, "Emotional sweating across the body: Comparing 16 different skin conductance measurement locations," *Physiology & Behavior*, vol. 106, no. 2, pp. 298–304, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031938412000613

[35] Y. Liang, M. Elgendi, Z. Chen, and R. Ward, "An optimal filter for short photoplethysmogram signals," *Scientific Data*, vol. 5, no. 1, p. 180076, 2018. [Online]. Available: https://doi.org/10.1038/sdata.2018.76

[36] P. van Gent, H. Farah, N. van Nes, and B. van Arem, "HeartPy: A novel heart rate algorithm for the analysis of noisy signals," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 66, pp. 368–378, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1369847818306740

[37] M. Benedek and C. Kaernbach, "A continuous measure of phasic electrodermal activity," *Journal of Neuroscience Methods*, vol. 190, no. 1, pp. 80–91, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165027010002335

[38] S. Schreven, P. J. Beek, and J. B. Smeets, "Optimising filtering parameters for a 3D motion analysis system," *Journal of Electromyography and Kinesiology*, vol. 25, no. 5, pp. 808–814, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1050641115001182

[39] A. Akbari, J. Martinez, and R. Jafari, "Facilitating Human Activity Data Annotation via Context-Aware Change Detection on Smartwatches," *ACM Transactions on Embedded Computing Systems*, vol. 20, no. 2, pp. 1–20, jan 2021. [Online]. Available: https://doi.org/10.1145/3431503

[40] M. Kwon, D. J. Kim, H. Cho, and S. Yang, "The smartphone addiction scale: Development and validation of a short version for adolescents," *PLoS ONE*, vol. 8, no. 12, p. e83558, dec 2013. [Online]. Available: https://doi.org/10.1371/journal.pone.0083558

[41] A. J. Martin and R. L. Debus, "Alternative Factor Structure for the Revised Self-Consciousness Scale," *Journal of Personality Assessment*, vol. 72, no. 2, pp. 266–281, apr 1999. [Online]. Available: https://doi.org/10.1207/S15327752JP720211

[42] A. Sahami Shirazi, N. Henze, T. Dingler, M. Pielot, D. Weber, and A. Schmidt, "Large-scale Assessment of Mobile Notifications," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: ACM, 2014, pp. 3055–3064. [Online]. Available: http://doi.acm.org/10.1145/2556288.2557189

[43] T. L. Powley, "Central Control of Autonomic Functions: Organization of the Autonomic Nervous System," in *Fundamental Neuroscience: Fourth Edition*, L. R. Squire, D. Berg, F. E. Bloom, S. du Lac, A. Ghosh, and N. C. B. T. F. N. F. E. Spitzer, Eds. San Diego: Academic Press, 2013, pp. 729–747. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780123858702000342

[44] F. Yasuma and J.-I. Hayano, "Respiratory sinus arrhythmia: why does the heartbeat synchronize with respiratory rhythm?" *Chest*, vol. 125, no. 2, pp. 683–690, feb 2004.

[45] P. Varma, M. Junge, H. Meaklim, and M. L. Jackson, "Younger people are more vulnerable to stress, anxiety and depression during COVID-19 pandemic: A global cross-sectional survey," *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 109, p. 110236, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0278584620305522

Chapter 5

Contact Force Estimation from Raw Photoplethysmogram Signal

Fortin, P. E., Blum, J. R., Weill–Duflos, A. & Cooperstock, J. R. (2020). "Contact Force Estimation from Raw Photoplethysmogram Signal". IEEE SENSORS, Rotterdam, Netherlands, pp. 1-4, doi:10.1109/SENSORS47125.2020.9278658.

Preface to Chapter 5:

As outlined in previous chapters, a major impediment to the collection of physiological signals *in situ* and the deployment of the systems based on physiological signals, such as SweatSponse, is the complexity of collecting high quality biosignals in uncontrolled environments. Indeed, many factors are known to influence the quality of contact-based physiological measurements including how tightly a sensor is pressed against the skin, and its position on the body. For optimal performance, such devices need to be firmly coupled, yet also remain comfortable when worn for extended periods of time.

This chapter presents design considerations and a proof of concept for a novel technique that enables wearables equipped with an optical heart rate sensor to estimate the contact force between the sensor and its user's skin. Initial tests indicate that the proposed method can estimate contact force as accurately as a force sensitive resistor (FSR), without the increased system complexity that typically accompanies discrete force sensing hardware.

Considering the wide availability of optical heart rate sensors in consumer and medical devices, the proposed force estimation technique has the potential to significantly enhance physiological signal quality by tightly monitoring coupling properties known to negatively influence measurements.

Contributions of Authors:

Pascal E. Fortin was the primary author of this paper, was responsible for the ideation, design and execution of the data collection apparatus, experiment and data analysis. Jeffrey R. Blum was co-author, he contributed to the ideation, experiment design and paper editing. Antoine Weill–Duflos edited the manuscript. Prof. Jeremy R. Cooperstock was co-author and supervised the research.

Abstract

Commercial smartwatches and fitness trackers are integrating increasingly advanced physiological sensors. For optimal performance, such devices need to be firmly coupled to the body, yet also remain comfortable when worn for extended periods of time. Existing solutions for measuring the contact force in order to ensure it is in an optimal tightness range typically depend on direct force measurement, but this adds hardware, and therefore cost, to the devices. This paper presents a novel method for estimating contact force by using only an optical heart rate sensor, as already found in many wearable devices. Initial tests indicate that the proposed method can estimate contact force with a mean absolute error of 0.36N, on par with FSRs. This new approach has the potential to expand the utility of existing sensors for both researchers and end-users, with anticipated applications not only in optimizing physiological sensing, but also in haptic information delivery.

5.1 Introduction

Modern wearables are equipped with an array of sensing hardware. Yet, if a device is not worn adequately tightly, sensor performance can deteriorate enough to become unusable. Both end-users, who desire the best value out of their devices' physiological sensing, as well as scientists, who need repeatable data collection conditions, stand to gain from consistent wearable tightness. However, without objective guidance, relying on either of these user groups to choose a subjective "reasonable" strap tightness for coupling is unlikely to yield optimal results.

Alternatives to such subjective coupling guidelines necessarily require a means to measure or estimate the contact force between a device and the user's skin. While we acknowledge the value of a dedicated force and/or pressure sensor in a wearable device [1, 2], these components increase a system's complexity and cost. This work therefore focuses on solutions that allow indirect estimation of contact force by using sensors that are already embedded in commercially available smartwatches and fitness trackers, and thus does not require additional electronics or hardware modifications.

There is little literature concerning such an approach. One relevant example of indirect force sensing proposed to use a smartphone's accelerometer in combination with its vibration actuator to estimate contact force by measuring the damping of a vibrotactile stimulus [3]. In this case, tighter coupling decreases vibrotactile amplitude and modifies the frequency spectrum. Despite the fact that this approach was proven effective and could be implemented on a wearable platform, activating the vibration motor solely to determine coupling may annoy the wearer and reduces battery life.

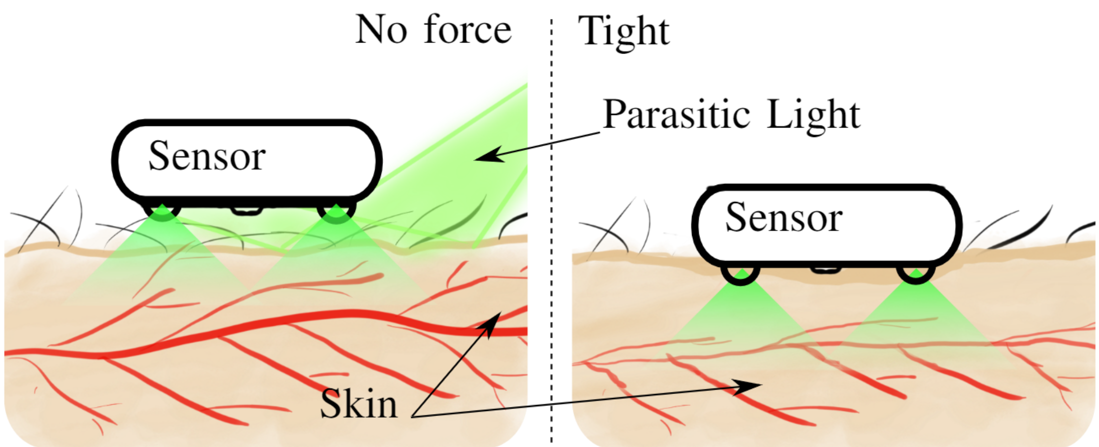


Fig. 5.1 Increased contact force compresses the tissues, blood vessels and correspondingly modifies the reflected light's properties.

5.2 Sensing Principles

As a less disruptive approach, we instead turn our attention to the optical heart rate sensor. Promisingly for our use case, the accuracy of optical heart rate sensors is known to be sensitive to how tightly they are pressed against the skin. Indeed, if too loosely coupled, the pulsatile component of the signal may become negligible in comparison with the attenuation caused by the soft tissues' thickness [4, 5]. In addition, parasitic light may saturate the light sensor, or otherwise degrade the signal. Alternatively, excessive contact force leads to vasoconstriction [4, 5], which also limits the PPG sensor's ability to accurately extract meaningful information from the raw signals (Figure 5.1). Further, [4] and [5] have both experimentally demonstrated that changes in contact force between a PPG sensor and the skin result in significant differences in the AC and DC amplitudes of the PPG waveform. Our approach therefore proposes to quantify these changes in raw PPG signal properties in order to estimate the normal force applied between the sensor and the body.

5.3 Data collection

Given the scarcity of the literature on the impact of contact force on PPG waveform properties, we conducted an experiment to explore PPG features beyond amplitude, and determine their potential for estimating skin coupling.

5.3.1 Apparatus

A custom data collection apparatus (Figure 5.2) was designed based on a publicly available 3D-printed vise model.[1] Two identical sensor modules measure PPG signals at the wrist and index fingertip (Pulse Sensor, World Famous Electronics llc). A 500 g rated miniature load cell (TAL221, HT Sensor Technology Co.), amplified by a HX711-based module (SEN-13879, Sparkfun Electronics), measures contact force. All sensor data are acquired using an Arduino Nano V3 microcontroller development board.

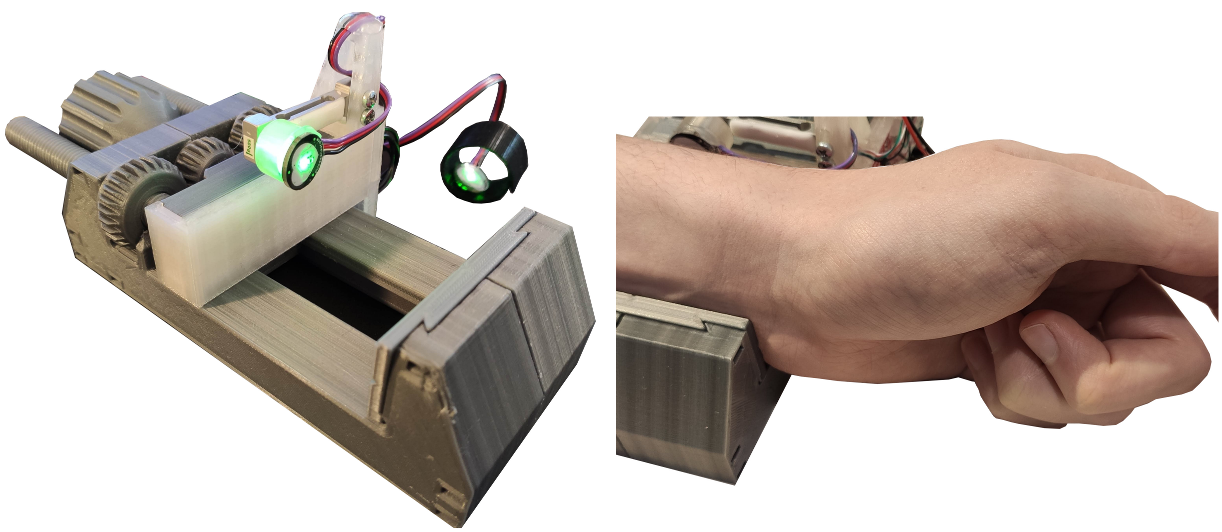


Fig. 5.2 Data collection apparatus and wrist placement.

5.3.2 Methodology

We expect that any factor that influences the ability to make accurate heart rate measurements, e.g., skin color, body fat, and skin hairiness [6] has an effect on raw PPG traces. Furthermore,

[1] ``Yet ANOTHER Machine Vise'' by TheGoofy, https://www.thingiverse.com/thing:2794662

we hypothesized that the analysis of the signal may be impacted by the heart rate itself. For example, when using a fixed window size during signal processing, different heart rates could result in different features being captured within each window, e.g., the number of heartbeat cycles in each frame modifies the distribution of the extracted features.

Typically, running multiple participants would address these potential factors as independent variables. Unfortunately, due to the enforcement of country-wide self-isolation measures,[2] the research team only received permission from the university's research ethics board for one of the co-authors (male, 28 years old, right handed, pale Caucasian skin tone) to self-administer the data collection protocol. Given these constraints, the protocol was modified to collect data at different heart rates and alternate between left and right wrists, as described in detail below. We fully acknowledge that data from a single participant limits the generalizability of the results. However, we contend that the findings would be sufficient to demonstrate the viability of the proposed sensing approach, with the caveat that we cannot say how the models may need to change on a per-user basis without further data collection.

The participant first secured the PPG sensor to his index finger. He then positioned his arm such that it was in contact with the fixed side of the apparatus and that his ulnar styloid process was pressed against the outer wall of the vise (Figure 5.2). An adjustment screw was used to set the sensor's position, and thus control the force applied.

Inspired by prior work from Teng and Zhang [4], PPG measurements were acquired at contact forces ranging from 0 (no contact but almost touching) to 2.6 N in increments of 0.2 N (14 levels). PPG measurements were sampled at each level for approximately 30 seconds. Sampling was performed in an ascending then descending contact force sequence for a total of 28 trials per session. Knowing that mean heart rate varies significantly throughout the day [7], seven measurement sessions were held at 7:30, 9:00, 11:30, 13:00, 15:30, 17:00 and 18:30 in order to introduce variability in the collected physiological signals. Each 11:30, 15:30 and 18:30 measurement session was immediately preceded by 2 minutes of jumping jacks to further modify the participant's heart rate.

5.3.3 Feature Extraction

Typically, the raw PPG signal would be pre-processed to attenuate artifacts and facilitate extraction of information of interest, e.g., heart rate. However, since our objective is to determine how

[2] `https://en.wikipedia.org/wiki/2019%E2%80%9320_coronavirus_pandemic`

Table 5.1 Features extracted from raw wrist PPG signal. Bold font indicates that the predictor is included in the final model.

Feature	Reference
Mean, Kurtosis	[8–11]
Variance, Skewness	[8–10]
Standard Deviation	[8, 9, 11]
Energy, Approximate entropy, Maximum slope, Singular value decomposition (SVD), Sym8 wavelet transform energy at levels 1-9 **(level 4)**	[8]
Maximum, Minimum	[11]
1st to 4th statistical moment **(3rd moment)**, 1st to 4th statistical moment (freq. domain)	[10]
Entropy, Median frequency	[9, 10]
Median, Root mean square, Inter-quartile interval, Total spectral power (0-10 Hz), **Relative spectral power (0-10 Hz)**, Peak amplitude (0-10 Hz), Mean of 1st derivative, Standard deviation of 1st derivative	[9]
Number of median crossings, Power spectral density at 1, 3, 5, 7, 9, 13, 17, 21 and 29 Hz **(7 Hz)**, Coefficients from 3rd order AR model, Number of median crossings of the instantaneous frequency	[12]

coupling affects the raw signal, we instead preserve the artifacts by not performing any such signal treatment.

A set of 55 feature candidates was taken from the PPG classification and automated artifact detection literature (Table 5.1). After downsampling the signal uniformly to 256 Hz, all features were extracted from a 0.5 s window that was moved in steps of 0.25 s over each trial's PPG recordings. Force measurements were averaged within each window to provide a single estimate of contact force.

5.4 Data set Description

A total of 50836 data points were collected from 14 measurement sessions. Measurements were generally evenly split across conditions, with 43.9% acquired following physical activity and 51.8% collected from the left wrist. Finger PPG signal clipping was observed in two sessions (18:30, left wrist following exercise, and 17:00, right wrist). Two replacement sessions were conducted under identical conditions on a subsequent day.

5.5 Contact Force Estimation

To estimate contact force from the wrist-based PPG signal, a bagged tree model was selected for its generally strong predictive ability (Matlab 2020a, "Optimizable ensemble" method). In an effort to reduce computational load while maintaining most of the predictive power, the minimal viable number of predictors was found by sequentially training models that included an increasingly large subset of features from Table 5.1. Step forward feature selection was used at each iteration to determine the set of predictor variables to use in each model. The process stopped when the RMSE dropped by less than 1% from that of the previous iteration. See Table 5.1 for the four selected features.

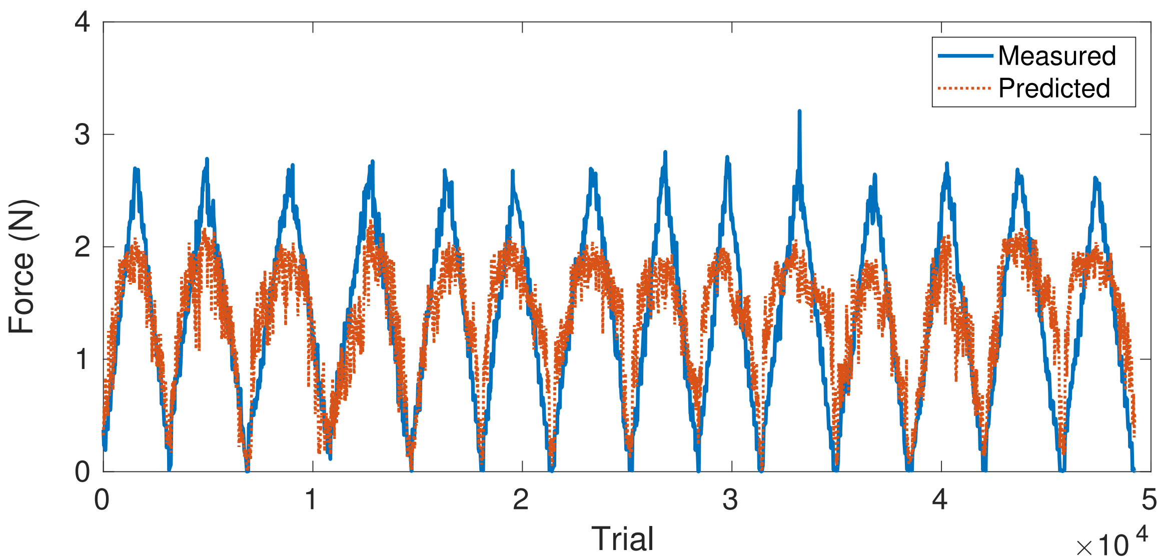


Fig. 5.3 Visual comparison of measured and predicted forces across the measurement sessions. Sensor saturation visibly begins at approximately 1.5 N.

As a reasonable compromise between noise reduction and latency of the output, a moving average was applied, using a 5 s window, to the force estimates. As indicators of our technique's performance, mean absolute errors (MAE) and root mean square errors (RMSE) of 0.36 N and 0.44 N were observed respectively. These two metrics, along with an r^2 of 0.67, suggest that the goodness of fit is moderate, and hints that other explanatory variables could be included in the model to more thoroughly explain the variance in the data. Inspection of the model's predictions (Figure 5.3 and 5.4) suggests that it most accurately predicts forces under 1.5 N, the point from the output begins to saturate.

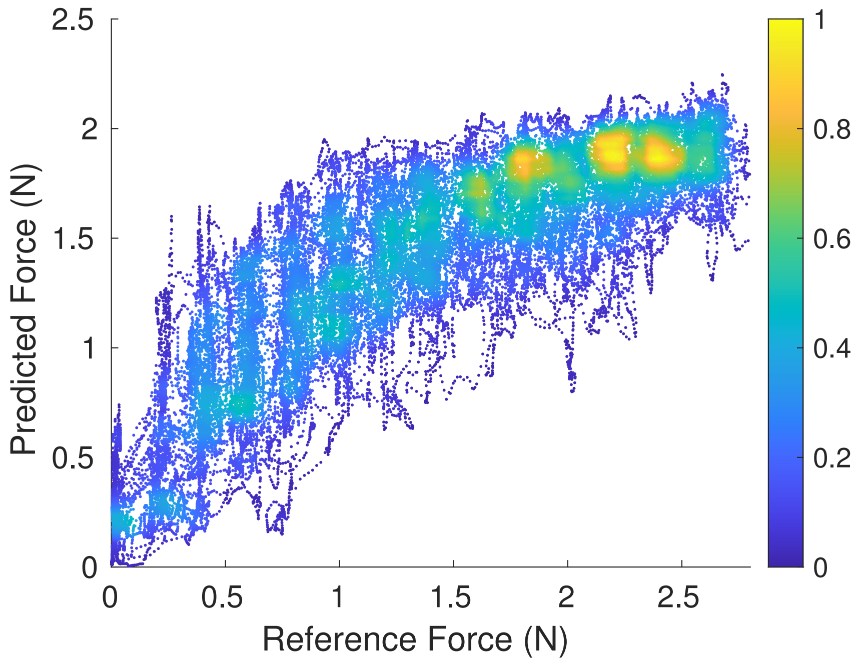


Fig. 5.4 Contact force estimation model characterization with relative density colormap

5.6 Discussion and Limitations

While the use of a sensor that is already available in commercial wearables makes the proposed approach an appealing alternative to load cells, its current MAE of 0.36 N cannot compete with the precision and accuracy of the latter. However, based on these initial tests, the performance of our technique is comparable to that of Force Sensitive Resistors (FSR), which can deviate by up to 25% from a reference force [3]. Thus, for studies conducted in laboratory settings that require accurate and repeatable contact force measurements, the authors advise against the use of the proposed approach. However, for cases where researchers have traditionally relied on a subjective "tight yet comfortable" coupling approach [13], or in-the-wild research where the use of custom wearable prototypes can pose significant technical challenges [14], our method provides a simple solution that allows better tightness consistency, increasing the likelihood of collecting high quality data.

The data used in this work were collected in static conditions using a vise mechanism and do

[3] Interlink Electronics, FSR Force Sensing Resistor Integration Guide and Evaluation Parts Catalog, https://www.interlinkelectronics.com

not account for motion artifacts that would be observed when using watch straps or wristbands. In a mobile setting, different features may better represent the noise in the signal, particularly when coupling is loose. Under such conditions, we hypothesize that the more significant effect of motion artifacts would increase our technique's prediction performance in the 0 to 1.5 N force range.

5.7 Conclusion

This work introduced a novel wearable contact force estimation system that employs raw optical heart rate measurements to estimate skin coupling. Promisingly, initial tests found the proposed method's force sensing accuracy to be comparable to that of FSRs. However, our solution has the advantage that it employs hardware that is already embedded in a large number of commercial wearable devices, ultimately reducing system complexity and overall costs.

References

[1] E. M. Young, A. H. Memar, P. Agarwal, and N. Colonnese, "Bellowband: A Pneumatic Wristband for Delivering Local Pressure and Vibration," in *2019 IEEE World Haptics Conference (WHC)*, 2019, pp. 55–60.

[2] E. Pezent, A. Israr, M. Samad, S. Robinson, P. Agarwal, H. Benko, and N. Colonnese, "Tasbi: Multisensory Squeeze and Vibrotactile Wrist Haptics for Augmented and Virtual Reality," in *2019 IEEE World Haptics Conference (WHC)*, 2019, pp. 1–6.

[3] S. Hwang, A. Bianchi, and K.-y. Wohn, "VibPress: estimating pressure input using vibration absorption on mobile devices," *Proc. 15th Int. Conf. Human-computer Interact. with Mob. devices Serv.*, pp. 31–34, 2013. [Online]. Available: http://dl.acm.org/citation.cfm?id=2493193

[4] X. F. Teng and Y. T. Zhang, "The effect of applied sensor contact force on pulse transit time," *Physiological Measurement*, vol. 27, no. 8, pp. 675–684, 2006. [Online]. Available: http://dx.doi.org/10.1088/0967-3334/27/8/002

[5] E. Geun, H. Heo, K. C. Nam, and Y. Huh, "Measurement site and applied pressure consideration in wrist photoplethysmography," in *ITC-CSCC: International Technical Conference on Circuits Systems, Computers and Communications*, 2008, pp. 1129–1132.

[6] B. Bent, B. A. Goldstein, W. A. Kibbe, and J. P. Dunn, “Investigating sources of inaccuracy in wearable optical heart rate sensors,” *npj Digital Medicine*, vol. 3, no. 1, p. 18, 2020. [Online]. Available: https://doi.org/10.1038/s41746-020-0226-6

[7] D. J. Ewing, J. M. Neilson, C. M. Shapiro, J. A. Stewart, and W. Reid, “Twenty four hour heart rate variability: effects of posture, sleep, and time of day in healthy controls and comparison with bedside tests of autonomic function in diabetic patients.” *Heart*, vol. 65, no. 5, pp. 239–244, 1991. [Online]. Available: https://heart.bmj.com/content/65/5/239

[8] D. Ramachandran, V. Ponnusamy Thangapandian, and H. Rajaguru, “Computerized approach for cardiovascular risk level detection using photoplethysmography signals,” *Measurement*, vol. 150, p. 107048, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0263224119309145

[9] T. Pereira, K. Gadhoumi, M. Ma, X. Liu, R. Xiao, R. A. Colorado, K. J. Keenan, K. Meisel, and X. Hu, “A Supervised Approach to Robust Photoplethysmography Quality Assessment,” *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 3, pp. 649–657, 2020.

[10] D. Álvarez, R. Hornero, J. V. Marcos, and F. del Campo, “Feature selection from nocturnal oximetry using genetic algorithms to assist in obstructive sleep apnoea diagnosis,” *Medical Engineering & Physics*, vol. 34, no. 8, pp. 1049–1057, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1350453311003006

[11] V. Jindal, J. Birjandtalab, M. B. Pouyan, and M. Nourani, “An adaptive deep learning approach for PPG-based identification,” in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 6401–6404.

[12] N. Pradhan, S. Rajan, A. Adler, and C. Redpath, “Classification of the quality of wristband-based photoplethysmography signals,” in *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2017, pp. 269–274.

[13] P. E. Fortin, J. R. Blum, and J. R. Cooperstock, “Towards consistent haptic coupling with HaptiStrap: Doing better than ”Tight yet comfortable”,” in *UIST 2019 Adjunct - Adjunct Publication of the 32nd Annual ACM Symposium on User Interface Software and Technology*. New Orleans, USA: ACM, oct 2019, pp. 69–71.

[14] J. R. Blum, P. E. Fortin, F. Al Taha, P. Alirezaee, M. Demers, A. Weill-Duflos, and J. R. Cooperstock, “Getting Your Hands Dirty Outside the Lab: A Practical Primer for Conducting Wearable Vibrotactile Haptics Research,” *IEEE Transactions on Haptics*, vol. 12, no. 3, pp. 232–246, 2019.

Chapter 6

Exploring the Use of Fingerprint Sensor Gestures for Unlock Journaling : A Comparison with Slide-to-X

Fortin, P. E., Huang, Y., & Cooperstock, J. R. (2019). Exploring the Use of Fingerprint Sensor Gestures for Unlock Journaling: A Comparison With Slide-to-X. In Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services (p. 24). ACM.

Preface to Chapter 6:

While techniques presented in previous chapters allow for an objective physiology-based assessment of notification perception, certain variables of interest (e.g., number of people around a user, whether they consider themselves in a public or private space) may not be captured without self-reporting. Mobile experience sampling methods (ESM) are an ensemble of techniques and tools that allow the collection of subjective information from participants throughout the day using prompts delivered from a smartphone.

The use of unlock journaling as an extension to existing mobile ESM has already demonstrated its potential by allowing the collection of more data per day and being perceived as less intrusive than traditional notification based approaches. This chapter introduces and evaluates a new unlock journaling mechanism based on fingerprint sensor gestures that is better adapted to modern authentication methods than existing touchscreen unlock journaling techniques. This novel self-reporting interface is implemented, alongside touchscreen approaches in an Android application, allowing the collection of subjective data from users as they unlock their smartphones.

This work extends the unlock journaling methods with the first mechanism that is consistent with fingerprint authentication. Results from a twelve-day user study show that the proposed method outperforms existing approaches in terms of response compliance, and offers comparable performance as quantified by its reporting time. Beyond showing the merit of the proposed input method in terms of raw performance, participants subjectively reported the fingerprint sensor gesture approach as being less intrusive and preferred it over the other interfaces. By increasing the accessibility of unlock journaling methods, this work will allow the collection of subjective data in ecological scenarios which, used alongside our notification perception system, will allow for a better understanding of the user's notification experience and context.

Contributions of Authors:

Pascal E. Fortin was the primary author of this paper, was responsible for the ideation and design of the application. He oversaw the technical implementation of the application, contributed to the data collection effort and data analysis. Daniel Huang contributed to the technical implementation of the application and to the data collection effort. Prof. Jeremy R. Cooperstock was co-author and supervised the research.

Abstract

Experience Sampling Methods (ESM) allow the timely collection of subjective self-reports that would otherwise be impossible to measure accurately in ecologically valid scenarios. Recent work suggests that unlock journaling allowed the collection of more data points per day, was faster and perceived as being less intrusive by participants than notification-based ESM. This work extends the unlock journaling field by introducing a novel lockscreen data collection mechanism harnessing an increasingly popular authentication mechanism: the fingerprint sensor. Results collected during a twelve-day user study with fingerprint sensor users show that fingerprint sensor gesture reporting compares favorably to Slide-to-X approaches. The proposed gestural interface was subjectively perceived as being the fastest, least intrusive, and overall most preferred interface, in addition to offering the highest response compliance. By offering a reporting mechanism better aligned with modern smartphone unlocking habits, this work encourages the deployment of unlock journaling in the wild.

6.1 Introduction

Experience sampling method (ESM) is a research methodology that uses questionnaires delivered at fixed times, fixed time intervals, or following specific events, to sample various aspects of the subject's experience throughout the day. ESM offers unique insights into the participants' internal state as they go about their daily activities, allowing for a better understanding of their context and motivations. It was adopted by a number of groups in the HCI community, as it was shown to be effective in the collection of subjective data related to the use and perception of novel interfaces and technologies [1]. In addition to painting a more externally valid picture of the user's experience of a technology than laboratory experiments, the subjective data provided by ESM, when put in combination with passive sensor monitoring, truly opens the door to practical machine learning systems that go beyond standardized datasets.

While paper and pen diaries have been used extensively by clinicians and researchers, they are limited by their introduction of retrospective biases and lack the temporal resolution necessary to observe behaviors in detail [2]. The spread of mobile technologies has greatly simplified the task for participants to provide in situ self-reports. This is primarily due to the fact that questionnaires can be completed faster on mobile interfaces, which are less cumbersome than a paper diary for users to carry all day [3]. Nevertheless, participant compliance with ESM protocols remains low and experiment dropout rates high, mainly because of the significant interruption burden imposed by having to stop one's task, complete a sometimes lengthy questionnaire, and return to the task at hand. [3].

In attempt to reduce the interruption burden experienced by users, Ponnada et al. introduced and evaluated the use of microinteractions in ESM [4]. Through a longitudinal user study, they demonstrated that the interruption burden of infrequent but long ESM prompts surpasses that of frequent, predictably short prompts that can be answered by a single touch interaction. They found this conclusion to be valid on both smartphone and smartwatch devices.

Tackling the issue from a different angle, a number of projects attempted to identify optimal delivery time for notifications and ESM prompts instead of altering the prompts themselves. Studies have shown interruption costs to be significantly reduced when delivering information during breakpoints, i.e., a change in context, task, action or application [5–9], than during high task engagement states [10]. This knowledge, in addition to the wealth of sensor data available in modern mobile platforms, was employed to develop context-aware notification systems that attempt to estimate the optimal time to deliver information in order to minimize interruption burden [11–14]. As promising as these proposed prediction systems are, they both require complex prediction algorithms in addition to large amounts of training data. Furthermore, their performance still needs improvement before they can be widely adopted.

Focusing exclusively on a breakpoint that can easily be recognized with minimal computing resources, researchers have investigated the use ESM prompts presented during the smartphone unlocking process, also known as unlock journaling. Indeed, to avoid accidentally triggering undesired actions, e.g., initiating a call or sending a message, most smartphones currently require a swiping gesture to be performed to initiate the unlocking process [15]. Beyond offering this rudimentary protection, this gesture does not have any practical use, offloading the lack of contextual awareness of the device to the users, by requiring them to waste a gesture. As such, prior literature has put this wasted gesture to work, by collecting data from subjects by requiring them to end their action on scales or other self-reporting instruments instead of in a random or arbitrary position [16–19]. Zhang et al. conducted a thorough analysis of response times, frequency of reports and perceived intrusiveness of unlock journaling techniques, comparing the Slide-to-X approach [16, 17] with traditional notification-based ESM. Their results suggested that lock-screen reporting led to a higher number of self-reports per day and was perceived as less intrusive than their implementation of a traditional notification-based ESM system. It is hypothesized that the performance gain associated with the use of unlock journaling can be attributed to:

1. Double-usage of swiping gesture: by replacing the wasted swiping gesture with one that lands on a reporting element, scale or grid, unlock journaling interfaces allow the collection of data, without introducing more touch interactions.

2. High interruptibility: unlock journaling integrates itself in a transition between two tasks or activities, during which interruptibility was shown to be high.

3. Co-location of the reporting-authentication interface: the user's finger is already close to, or on, the screen when they authenticate using their usual method, which reduces the time needed to complete a report.

Modern authentication techniques, however, do not rely on touch screen interactions to let users access their device, and thus, are ill-suited to integration with ESM. For example, Apple's Face ID and Samsung's iris scanner technologies only require the user to stare at their device to unlock it, while fingerprint authentication asks from users to press their finger on the sensor. Although fingerprint sensors are becoming commonplace, even among budget smartphones, the value and advantages offered by unlock journaling may no longer hold when they are used for authentication. Indeed, a user who uses a fingerprint sensor located at the back of the smartphone, as is increasingly frequent, would have to authenticate by placing their finger on the sensor, and then move their hand in order to provide input through the touch screen to the self-reporting interface appearing on the phone.

To adapt the field of unlock journaling to modern authentication techniques, this work introduces a new reporting mechanism that takes advantage of Android's fingerprint sensor gesture API to allow co-located fingerprint authentication and self-reporting based on on-screen prompts. We report here on the results of a twelve-day user study, comparing performance of this reporting mechanism with Slide-to-X methods, quantified by the time required to enter data, perceived intrusiveness, and response compliance.

6.2 Fingerprint Sensor Gesture Interface

We designed an unlock journaling interface using the Android fingerprint gesture API. The interface captures fingerprint gestures applied to the smartphone sensor in order to support data entry based on an on-screen prompt. Participants respond by performing one of up to four different single-action gestures to indicate four different levels of a variable (see Figure 6.1)

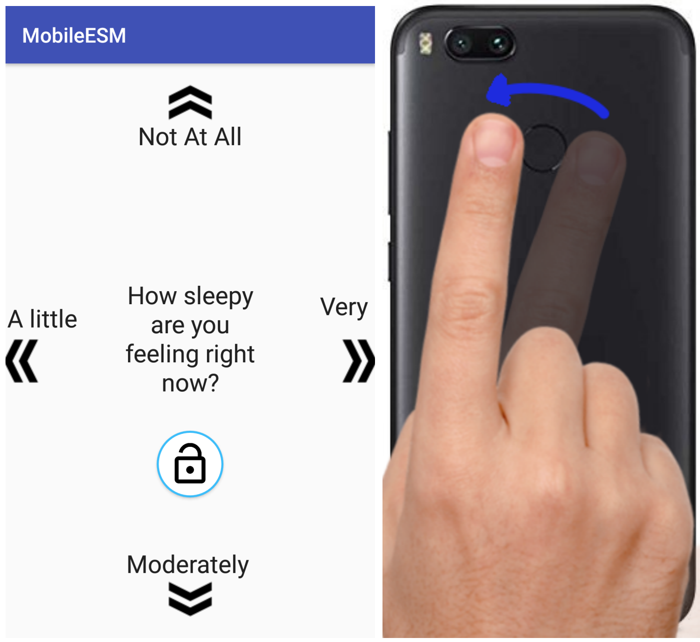


Fig. 6.1 Screenshot of the fingerprint sensor gesture interface and example gesture required to enter ”Very”. The inversion of the swiping gesture is due to it being applied to the back of the device

The interface appears as soon as users authenticate themselves and disappears immediately after the user performs a fingerprint sensor gesture. The user can also manually dismiss the interface by pressing the Android built-in ”Back” or ”Home” button, or the padlock button built into the reporting interface.

Interestingly from a privacy perspective, the fingerprint sensor gesture API does not grant access to the fingerprint data itself. Rather, it only returns the direction when a swiping gesture is detected, independently from any information regarding the owner’s identity.

6.3 User Study

A within-subjects study was designed to compare the performance of the proposed fingerprint sensor gesture reporting in comparison to Slide-to-X reporting, as introduced in prior work [16, 17]. Following one of the scenarios used by Zhang et al. [17], the study was framed as a sleepiness tracking activity.

6.3.1 Recruitment

Recruitment was achieved using a combination of internal mailing lists, posts on official university community Facebook pages and the university's student society classified ads. To be eligible, participants had to be active users of an Android O (or more recent) smartphone equipped with a fingerprint sensor on the back of their device. Use of the sensor for device unlocking was not an inclusion criterion for the study.

It was discovered during the pilot study that only a small subset of Android O phones, equipped with a fingerprint sensor would respond as per the Android documentation to fingerprint sensor gesture function calls. Therefore, participants were asked to submit their smartphone brand and models for a screening before the initial meeting.

6.3.2 Technical Framework

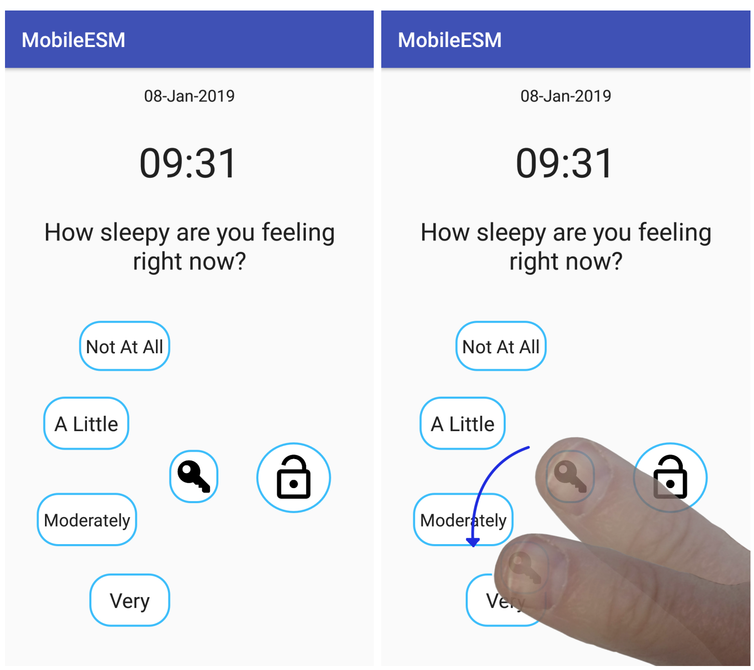


Fig. 6.2 Screenshot of the Slide-to-X interface and example gesture required to enter "Very".

In order to provide a common point of comparison, a Slide-to-X interface replicating a unidimensional ordinal data reporting example from Zhang et al. [17] was adapted and implemented (see Figure 6.2).

With the Slide-to-X before unlock (SXBEF) approach, the sequence of actions required for users to unlock their device is the following:

1. Press home/unlock/power button
2. Enter self-report by sliding a marker or finger to the rating item
3. Authenticate using usual mechanism

However, this sequence of actions differs from that experienced when using fingerprint sensor gesture reporting (FG) and identification:

1. Apply finger on fingerprint sensor to authenticate
2. Enter self-report by swiping finger on sensor

Consequently, a third self-reporting interface (SXAFT) was included in the study, in order to employ the same sequence of actions as fingerprint reporting, but with the Slide-to-X interface, i.e., authenticate and then enter self-report, as used for on-screen methods. This third condition should offer insights into the importance of the order of actions when implementing unlock journaling interfaces. More precisely, it is hypothesized that since users have already unlocked their device, they would want to be able to engage in their regular activities immediately, as opposed to being stopped by the reporting interface. This could have the effect of allowing the collection of less data points than Slide-to-X before unlock due to its position in the smartphone unlocking sequence.

6.3.3 Protocol

Participants were welcomed with a brief information session about the project. After reading and signing the institutionally approved consent form, they were administered a pre-experiment questionnaire aimed at understanding their prior experience with self-tracking instruments and smartphone unlocking mechanisms (e.g., pin, pattern, fingerprint sensor, face ID).

The experiment application was then installed on the participants' smartphones. Time was taken to familiarize them with the three reporting interfaces and ensure that they knew basic troubleshooting procedures for the application.

Participants left the laboratory and used the application for twelve (12) consecutive days, starting on the day following the initial meeting. Each interface was presented for two consecutive days in the first and second half of the study. The presentation order of each condition was randomized within each half (e.g., 1) FG, FG, SXAFT, SXAFT, SXBEF, SXBEF, 2) SXAFT, SXAFT, FG, FG, SXBEF, SXBEF). The application was set to automatically change to the next reporting interface in the sequence every two days. Every morning, a notification was delivered to remind the participants of the interface used on that day to minimize confusion due to the different reporting interface. Participants were asked to use an authentication method that was aligned with the self-reporting interface used on that given day, i.e., if the day's mode was SXBEF, the user was told to make sure to use an on-screen authentication mechanism (no security, PIN or pattern). The authentication mode had to be changed manually by participants, since the Android operating system does not allow applications to make modifications to security related mechanisms.

6.3.4 Measurements

The experiment application records to a local log the participants' entries and response times through the reporting interface, as well as the self-reporting mechanism used for each event. In order to ascertain the subjects' perception of each journaling interface, a short questionnaire was presented at the end of each day at a time selected during the initial meeting. Participants were asked to rate the perceived intrusiveness of that day's interface using a Likert scale (1-not intrusive at all to 5-extremely intrusive). Following the completion of each daily questionnaire, the log file was transferred to the cloud for subsequent analysis.

We considered the following measurements employed by Zhang et al. for the purpose of evaluating and comparing the three self-reporting interfaces:

1. Response time: The time elapsed between the presentation of the interface and the completion of a self-report or interface dismissal.

2. Intrusiveness: Perceived intrusiveness of the reporting interfaces as measured using daily questionnaires.

3. Frequency: Number of self-reports entered per hour

Since all reporting interfaces are relying on device unlocking, it is expected that frequency will not vary significantly between interfaces for an individual subject. As such, we introduce a new metric: response compliance, mathematically expressed as $N_{entry}/(N_{entry} + N_{dismiss})$, where N_{entry} is the number of time a data point was entered and $N_{dismiss}$ is the number of times the interface was dismissed within the same time period. This metric is thought to be representative of how often users decide to enter a data point instead of dismissing the interface. A response compliance of 1 suggests that the interface never dismissed, and conversely, a value that approaches 0 indicates a high interface dismissal rate.

A post-experiment questionnaire was also employed to collect more information on their perception of the different interfaces at the end of the twelve-day period. More specifically, it asked participants to identify which interface they found to be fastest, most error-prone, most intrusive, least intrusive, and preferred overall.

6.3.5 Hypotheses

Based on the prior literature on experience sampling methods and unlock journaling, the following hypotheses were formulated:

- H1: Because of the hypothesized smaller movement amplitude to enter a rating, and the co-location of the unlocking and rating instruments, fingerprint gestures are anticipated to require less time on average to enter a self-report than either of the Slide-to-X approaches.
- H2: Assuming H1 is confirmed, fingerprint gestures should be perceived as less-intrusive than Slide-to-X methods, as reflected in participants' daily questionnaire entries.
- H3: The frequency of ratings should not vary significantly between self-reporting methods because of the anticipated participants' regular usage patterns. However, response compliance will be higher with the fingerprint gesture than Slide-to-X methods because of its anticipated shorter completion time and lower perceived intrusiveness.

In addition, since participants were instructed to use an authentication mechanism that was appropriate for each self-reporting interface, we anticipate significant interaction between the self-reporting interfaces and participants' usual unlocking mechanisms on all measurements. For example, PIN-users might rate fingerprint gestures as being more intrusive and require more time to complete self-reports since they are not initially used to placing their finger on the sensor to unlock their device.

6.4 Participants

In total, ten subjects (8M/2F) 20 to 38 years old, $\overline{x} = 25.2$ participated in this tiredness tracking study. Half of the subjects reported being currently engaged in some form of automated self-tracking activities (e.g., step counting or heart-rate tracking).

Even though recruitment was not specifically targeting fingerprint sensor *users*, 100% of subjects reported using this as their main authentication mode. This prevalence, even on a small sample size, suggests that users of smartphone equipped with a fingerprint sensor might be more likely to use this authentication mechanism than touch-screen based methods. Given the homogeneity of default authentication mode by our recruited subjects, we were precluded from carrying out the planned analysis of the interaction between subject's usual authentication mechanisms and the metrics employed in this study.

6.5 Results

6.5.1 Technical issues

During the study, certain smartphone models caused the interface to fail to change automatically every two days. This problem could not be replicated on the experimenters' devices. Nevertheless, the number of time a particular reporting interface was presented is approximately balanced ($\overline{x}$ = 1643, σ = 106.7). Out of caution, the predictive ability of the number of days spent in each condition on the perceived interface speed, error-proneness, intrusiveness and overall interface preference was explored using multinomial logistic regression analyses. With the exception of error-proneness, the time spent in each condition was not found to be a statistically significant predictor of the data collected as part of the post-experiment questionnaire. As such, and even though correlation does not equal causation, this measurement will not be used as part of the discussion of the results.

6.5.2 H1: Response time

Some participants' devices were configured such that their screen would not automatically turn off when plugged in. This resulted in the self-reporting interfaces being displayed for extended periods of time (e.g., hours) without data being entered. To prevent these outliers from skewing the data set, response times exceeding two standard deviations over the mean were not considered for analysis.

Table 6.1 presents summary statistics of the response times observed for each interface. A within-subjects ANOVA was used to investigate the difference between the average response time in each reporting interface. No main effects of the self-reporting interface was noted on response

Table 6.1 Response times for each interface

	Response time (ms)	
Interface	**Mean**	**Standard Deviation**
FG	2419	1563
SXBEF	2570	1803
SXAFT	2361	1714
Aggregated	2451	1702

time, suggesting that FG does not outperforms Slide-to-X methods when it comes to response time alone. These quantitative findings reject our first research hypothesis (H1) that predicted a shorter response time for fingerprint gestures than for Slide-to-X methods.

On the other hand, data collected in the post-experiment questionnaire reveals that 90% of participants perceived the FG condition as being the fastest of all three tested interfaces. The fact that it was perceived as being the fastest is arguably more important, as it is the user's perception of the system that would influence its adoption and compliance with the data collection procedure.

6.5.3 H2: Perceived Intrusiveness

A within subjects ANOVA was used to compare the influence of the reporting interface on the perceived intrusiveness as reflected by the subjective data collected using daily questionnaires. Since participant 10 did not complete any of the four daily questionnaires for one of the experimental conditions, this individual's data is not considered for this part of the analysis. A statistically significant main effect of the reporting interface on the perceived intrusiveness was observed ($F(2,16)=4.36$, $p< .05$). To understand where the observed difference resides, multiple comparisons were computed using Tukey's HSD test, correcting p-values using the Bonferonni-Holm method. A statistically significant difference in perceived intrusiveness was observed between FG and SXBEF ($p< .05$). However, no significant differences were observed between fingerprint sensor gestures and the Slide-to-X after unlock mode, and between the two Slide-to-X methods (see Figure 6.3). These results suggest that fingerprint sensor gesture reporting compares favorably to the Slide-to-X method before unlock proposed by Truong et al. [16] and Zhang et al. [17]. Data collected using the post-experiment questionnaire supports this claim. Indeed, 80% of subjects selected FG as being the least intrusive of the three interfaces used.

These results partially confirm our second research hypothesis (H2) that predicted less overall perceived intrusiveness for fingerprint gesture reporting interface than for both Slide-to-X

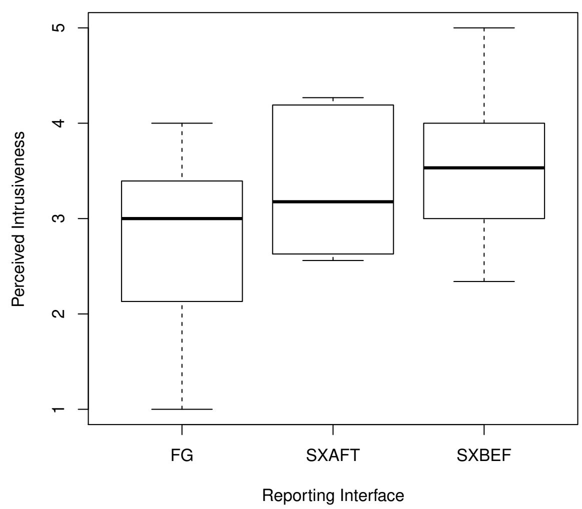


Fig. 6.3 Perceived intrusiveness as reported by the daily questionnaires for each mode

methods.

6.5.4 H3: Frequency & Response Compliance

As predicted in Section 6.3.4, a within subjects ANOVA showed that the frequency of hourly reports did not change significantly based on the reporting mode used for each subject.

However, a significant main effect of the reporting interface on response compliance was observed ($F(2,708)=409.3$, $p< .0001$). Indeed, a multiple comparisons test revealed that FG statistically significantly outperformed both SXAFT ($p< .0001$) and SXBEF ($p< .0001$). A statistically significant difference was also observed between both Slide-to-X methods ($p< .05$).

These finding allow us to confirm our third hypothesis (H3), stating that the response frequency would not vary significantly across reporting interfaces based on their common reliance on smartphone unlocking, and that FG's response compliance would be superior to that of other interfaces.

6.5.5 Overall preference

While no formal research hypothesis was associated with the users' interface preference, fingerprint gesture reporting was selected as the overall preferred interface by eight of ten participants,

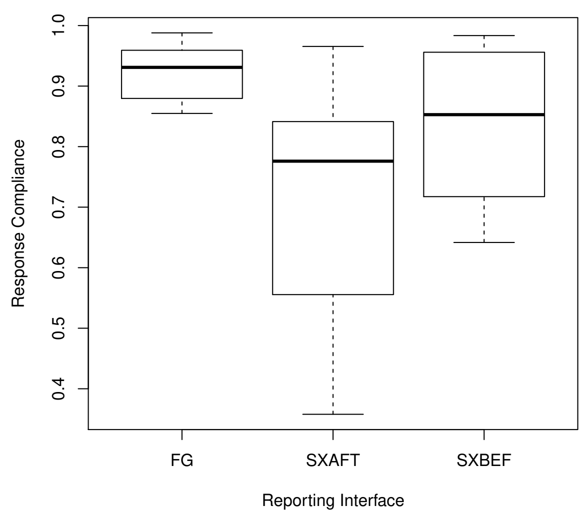


Fig. 6.4 Response compliance for each of the self-reporting interfaces.

with both Slide-to-X modes receiving only one vote each. However, this result should be viewed with suspicion since all the participants used fingerprint authentication as their default unlocking mechanism.

6.5.6 Journaling Before or After Unlocking the Device?

With the exception of response compliance, no statistically significant differences were observed between the two Slide-to-X interfaces. These results go against our working hypothesis that SX-AFT would perform worst by all measures. Rather, they suggest that the performance advantages reported by Zhang et al. [17] in their comparison of SXBEF against traditional notification-based ESM may have been due to other factors than the use of the same swiping gesture for unlocking and reporting. Otherwise, we would expect SXBEF to be faster, perceived as less intrusive, and offer greater response compliance, because it requires the least amount of gestures. The advantages of unlock journaling methods could instead mainly reside in their presentation at an ecological time during task transitions, in addition to the co-location of the reporting instruments and authentication interfaces.

6.6 Limitations

While the results presented in this paper are promising for the proposed fingerprint gesture method, two important limitations should be considered.

First, the comparison presented in this paper was limited to the reporting of four discrete elements, a constraint imposed by the Android fingerprint gesture API, which can only recognize four discrete gestures. This limitation would preclude the application of fingerprint gesture reporting in applications requiring the collection of more than four ordinal or nominal elements. It also makes this approach ill-suited to the reporting of multi-dimensional data such as the pleasure-arousal space used by Zhang et al. in their study [17]. Such data could be entered using consecutive swipes on the sensor to enter its different components, but doing so would increase response time, and likely have a steeper learning curve for participants.

Second, as noted previously, all participants were users of fingerprint authentication. This might have negatively influenced the qualitative and indirectly the quantitative measurements for both Slide-to-X methods. Nevertheless, we argue that the evaluation and comparison presented in this paper is representative of the fingerprint sensor gesture reporting interface intended audience.

6.7 Conclusion

This work extends the toolbox of unlock journaling techniques by introducing the use of fingerprint sensor gestures to enter self-reports. Based on results from a twelve-day user study with fingerprint authentication users, this new reporting mechanism is not only the first to be coherent with fingerprint authentication, our results suggest that it offers better response compliance, perceived intrusiveness and perceived reporting speed than Slide-to-X methods, and performed competitively when comparing response time. Unsurprisingly since they were all fingerprint sensor users, the proposed reporting interface was preferred by the majority of subjects over the current state of the art Slide-to-X approaches which rely on on-screen authentication methods. While we believe that fingerprint sensor gesture reporting could be employed alongside facial and iris recognition authentication because of its reliance on phone holding positions that allow finger placement on the sensor [20], further work is warranted to explore how unlock journaling techniques can be applied to these contactless authentication methods.

References

[1] S. Consolvo and M. Walker, "Using the experience sampling method to evaluate ubicomp applications," *IEEE Pervasive Computing*, vol. 2, no. 2, pp. 24–31, 2003.

[2] S. Shiffman, A. A. Stone, and M. R. Hufford, "Ecological Momentary Assessment," *Annual Review of Clinical Psychology*, vol. 4, no. 1, pp. 1–32, Mar. 2008. [Online]. Available: https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

[3] N. V. Berkel, D. Ferreira, and V. Kostakos, "The Experience Sampling Method on Mobile Devices," *ACM Comput. Surv.*, vol. 50, no. 6, pp. 93:1—-93:40, Dec. 2017. [Online]. Available: http://doi.acm.org/10.1145/3123988

[4] A. Ponnada, C. Haynes, D. Maniar, J. Manjourides, and S. Intille, "Microinteraction Ecological Momentary Assessment Response Rates: Effect of Microinteractions or the Smartwatch?" *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 92:1—-92:16, Sep. 2017. [Online]. Available: http://doi.acm.org/10.1145/3130957

[5] P. D. Adamczyk and B. 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*, ser. CHI '04. New York, NY, USA: ACM, 2004, pp. 271–278. [Online]. Available: http://doi.acm.org/10.1145/985692.985727

[6] B. P. Bailey and J. 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*, vol. 22, no. 4, pp. 685–708, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S074756320500107X

[7] V. Kostov, T. Tajima, E. Naito, and J. Ozawa, "Analysis of appropriate timing for information notification based on indoor user's location transition," in *Fourth Annual IEEE International Conference on Pervasive Computing and Communications (PERCOM'06)*, 2006, pp. 6 pp.–250.

[8] S. T. Iqbal and B. P. Bailey, "Leveraging Characteristics of Task Structure to Predict the Cost of Interruption," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '06. New York, NY, USA: ACM, 2006, pp. 741–750. [Online]. Available: http://doi.acm.org/10.1145/1124772.1124882

[9] T. Tanaka and K. Fujita, "Study of User Interruptibility Estimation Based on Focused Application Switching," in *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, ser. CSCW '11. New York, NY, USA: ACM, 2011, pp. 721–724. [Online]. Available: http://doi.acm.org/10.1145/1958824.1958954

[10] V. Pejovic, M. Musolesi, and A. 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*, ser. MobileHCI ’15. New York, NY, USA: ACM, 2015, pp. 1100–1105. [Online]. Available: http://doi.acm.org/10.1145/2786567.2794336

[11] S. T. Iqbal and B. P. Bailey, “Effects of Intelligent Notification Management on Users and Their Tasks,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’08. New York, NY, USA: ACM, 2008, pp. 93–102. [Online]. Available: http://doi.acm.org/10.1145/1357054.1357070

[12] V. Pejovic and M. Musolesi, “InterruptMe: Designing Intelligent Prompting Mechanisms for Pervasive Applications,” in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp ’14. New York, NY, USA: ACM, 2014, pp. 897–908. [Online]. Available: http://doi.acm.org/10.1145/2632048.2632062

[13] H. Sarker, M. Sharmin, A. A. Ali, M. M. Rahman, R. Bari, S. M. Hossain, and S. Kumar, “Assessing the Availability of Users to Engage in Just-in-time Intervention in the Natural Environment,” in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp ’14. New York, NY, USA: ACM, 2014, pp. 909–920. [Online]. Available: http://doi.acm.org/10.1145/2632048.2636082

[14] M. Pielot, B. Cardoso, K. Katevas, J. Serrà, A. Matic, and N. Oliver, “Beyond Interruptibility: Predicting Opportune Moments to Engage Mobile Phone Users,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 91:1—-91:25, Sep. 2017. [Online]. Available: http://doi.acm.org/10.1145/3130956

[15] I. Chaudhri, B. Ording, F. A. Anzures, M. Van Os, S. O. Lemay, S. Forstall, and G. Christie, “Unlocking a device by performing gestures on an unlock image,” Feb. 2010, uS Patent 7,657,849.

[16] K. N. Truong, T. Shihipar, and D. J. Wigdor, “Slide to X: Unlocking the Potential of Smartphone Unlocking,” *Proceedings of the 32nd annual ACM conference on Human factors in computing systems.*, pp. 3635–3644, 2014.

[17] X. Zhang, L. R. Pina, and J. Fogarty, “Examining Unlock Journaling with Diaries and Reminders for In Situ Self-Report in Health and Wellness.” *Proceedings of the SIGCHI conference on human factors in computing systems. CHI Conference*, vol. 2016, pp. 5658–5664, May 2016. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/28804795http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5553980

[18] J. Jung, M. Nour, M. Allman-Farinelli, and J. Kay, “Harnessing the ”Ambience” of the Mobile-phone Lockscreen for Ultra-lite Logging,” in *Proceedings of the 29th Australian*

Conference on Computer-Human Interaction, ser. OZCHI '17. New York, NY, USA: ACM, 2017, pp. 21–30. [Online]. Available: http://doi.acm.org/10.1145/3152771.3152774

[19] R. Vaish, K. Wyngarden, J. Chen, B. Cheung, and M. S. Bernstein, "Twitch Crowdsourcing: Crowd Contributions in Short Bursts of Time," in *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: ACM, 2014, pp. 3645–3654. [Online]. Available: http://doi.acm.org/10.1145/2556288.2556996

[20] S. Hoober, "Design for Fingers, Touch, and People, Part 1," 2017. [Online]. Available: https://www.uxmatters.com/mt/archives/2017/03/design-for-fingers-touch-and-people-part-1.php

Chapter 7

Discussion

This thesis presented the first exploration of the use of physiological signals as a means to better understand smartphone users' notification experience. By studying physiological responses to smartphone users' perception of their own notifications in and outside of the laboratory, this work lays the foundations of this new area of research and demonstrates the feasibility of conducting *in situ* notification research based on the analysis of physiological signals. Recognizing the limitations of current physiological sensors and psychophysiological inferences, it also proposes two systems that aim to support the collection of physiological signals and subjective data in the wild.

In addition to its core contributions, the work presented in this thesis made use of creative methodological approaches that expanded possible notification research directions and circumvented logistical limitations imposed by the Covid-19 pandemic. To avoid redundancy with each chapter's discussion, this section aims to highlight the process through which these research contributions were obtained, and how it advances notification research methodologies. In addition, practical recommendations, promising areas of future research as well as non-research applications of the systems presented in this thesis are discussed.

7.1 Physiologically Informed Notification Research

As introduced in Chapter 2, two main methodological currents exist in notification research. The passive approach relies on the background observation of participants' notification interactions to explore quantitative behavioral research questions, without requiring participants to engage with data collection instruments (e.g., self-reporting interfaces). The active approach instead uses various questionnaires delivered after the presentation of a notification or at fixed time intervals to explore attitudinal research questions pertaining to participants' notification perception and experience. These two methodologies can be employed individually, but are frequently combined to attempt to draw correlations between behavioral and attitudinal independent variables. Despite their proven track record, these approaches suffer from limitations that motivated the methodological explorations presented in this thesis:

1. By relying on direct notification interactions to determine perception, they do not have the means to assess whether a notification was actually perceived *when* it was presented [1]. In addition, it is impossible to determine whether the smartphone interaction used to infer perception coincidentally followed the presentation of a notification or resulted from it. Therefore, making a notification perception assumption from smartphone-based actions may at best lead to the delayed confirmation that a notification was detected, and in the worst case to the mistaken assumption than an alert was perceived.

2. Subjective self-reports are known to be susceptible to different sources of biases that can negatively impact data quality (e.g., retrospective bias, demand characteristics). For example, knowing that a researcher is studying the impact of notifications on stress, participants could adapt their responses to match their expectations of what the researcher is seeking, as opposed to remaining true to their personal experience.

3. The frequent interruption of the self-reporting interfaces can increase the notification volume, directly influencing the phenomenon that is being measured. Indeed, it is hypothesized that this may further amplify negative affect towards notifications and their perceived interruption burden.

This thesis makes the argument that recent advances in wearable sensing, affective and physiological computing make possible the use of biosignals as a means to assess users' experience of notification, minimizing reliance on subjective self-reports. Towards this objective, Chapters 3 and 4 presented evidence that it is indeed possible to capture significant changes in physiological signals resulting from notification perception, both in and outside of the laboratory. By demonstrating that these changes can be employed to classify whether a notification was perceived by a smartphone user, they also showed that they could be used to support the research and design of future intelligent, perception-aware notification systems.

Beyond its proof of concept, the interaction-less notification confirmation method introduced in Chapter 3 and extended in Chapter 4 provides a source of information that was previously inaccessible to researchers, and enables the investigation of unexplored research directions. In their work, Pielot et al. introduced a five-stage model of notification consumption [1]. Based on their model, a notification can be shown, seen, checked, consumed and finally removed. By their admission, “shown” events are not considered for analysis in the current notification literature since researchers do not have the means to know whether the stimulus was actually presented by a system. For example, an application’s request to post a notification may be blocked by the operating system or its presentation may fail due to a hardware failure (e.g., broken speaker or vibration actuator). Similarly, within that framework, unlocking a smartphone or receiving a notification while the device is unlocked carries the assumption that a notification was “seen”, limiting the applicability of their model to the visual component of the notifications and neglecting their typical auditory and/or vibrotactile components. Employed within that framework, the proposed physiologically informed perception confirmation system would not only allow the inclusion of “shown” and “seen” events in the analysis, but also extend the model from a visual-only paradigm to the vibrotactile and auditory modalities, significantly broadening its generalizability. This new capacity means that the majority of existing research on notification responsivity and handling [1, 2] could be re-visited to increase its granularity. Breaking down analyses between perceived and missed notifications will allow researchers to better understand the individual contributions of contextual factors that influence notification responsivity, handling behavior and ultimately, their impact on participants. For example, while notification attendance time is traditionally computed as the difference between the time at which an alert is presented and removed, using the proposed system to confirm that a notification was perceived or missed significantly changes possible interpretations of longer response times, which in the literature are attributed to other factors such as the users’ activity, the type of application that generated the notification and/or its originating contact. In contrast, the approach described in this thesis allows to confirm whether a notification was perceived, providing a potential alternative explanation for long response times.

7.1.1 Practical Recommendations for *In Situ* Physiological Signals-based Notification Research

While knowing how to *use* a physiological sensing platform might be sufficient to conduct laboratory studies, taking the research into the real world requires a thorough understanding of a device’s sensing principles, how it communicates and logs data, and the different ways it can (and will) fail. Building on the experience I acquired working with physiological signals both in and outside of the laboratory, this subsection presents practical recommendations and information I would have wanted to know when I first started in this research direction. Many of these tips are not specific to notification research and physiological signals, but are particularly important within that research context.

In the lab, environmental conditions are ideal for physiological data collection, e.g., constant comfortable temperature and humidity level. However, in the wild, participants and equipment can be exposed to liquids, sudden changes in temperatures and other unpredictable events that can introduce significant changes in physiological signals [3]. Knowing a sensor's sensing principles allows researchers to adequately constrain the conditions in which data can be collected in order to maximize the quality of the signals and the depth of their analysis. For example, electrodermal activity is typically acquired by measuring the electrical current resulting from the application of a fixed voltage between two electrodes in contact with the skin. As such, collecting this signal in conditions where the electrodes could be exposed to liquids (e.g., shower, rain, snow) is unlikely to yield consistent skin conductance measurements. It is therefore important for the experimenter to specify in their participants' instructions that the electrodes and surrounding skin should not be exposed to liquids. Another alternative is to collect sensor data in all conditions that are safe for the participant and sensor, and reject segments that do not meet signal quality criteria. On that last point, it is important to adjust one's expectations with regards to the quality of physiological signals collected in the wild, i.e., what is typically considered poor to passable signal quality in the lab is essentially the norm when considering data collected in the wild. In Chapter 4, since the Shimmer3 GSR+ sensor is not waterproof, we requested that participants leave the sensor indoors if they were leaving their residence during data collection. They were also asked to remove the device when engaging in activities that would risk exposing the unit to liquids.

The signal quality of practically all physiological sensors depends on robust mechanical coupling with the users' skin and adequate positioning on the body. In lab studies, experimenters are typically responsible for attaching and positioning the sensors on participants. With their experience and understanding of the sensing principles, they know where to position the sensing elements and how hard they should be pressed against the skin to maximize signal quality. However, when engaging in in-the-wild research, participants may be responsible for this task. This is particularly true of studies conducted during the current global pandemic, where the equipment may be dropped at the participant's residence without physical contact to minimize risks of virus propagation. In the study presented in Chapter 4, multiple strategies were successfully employed to assist participants in wearing the wearable sensor at the right locations and pressed sufficiently hard against the skin. First, the equipment package delivered to participants contained detailed device positioning instructions that included illustrations from different points of view. Figures were used to communicate where the sensing elements should be placed on the fingers and the route wires should take between the fingers to reach the main sensing unit. Second, a video call was scheduled during which the experimenter walked the participant through the process of putting on the sensor using the instructions and pictures from the equipment package. Third, feedback was provided on how the sensors were positioned and how tight the sensing elements should feel. A participant expressed his apprehensions with regards to putting on the sensor correctly. He later explicitly stated that the detailed instructions and feedback provided made him feel confident in his abilities to collect data by himself. It is hypothesized that TightHR, introduced in Chapter 5, could have assisted in this last step by autonomatically providing tightness adjustment recommendations based on its contact force estimation. However, considering the separate leads for skin conductance and photoplethysmograph sensing elements, doing so would have only helped for the optical heart rate and heart rate variability estimations.

Using physiological signals for notification research requires tight time synchronization between participants' smartphones, on which alerts are presented, and the device to which physiological measurements are logged. The amount of acceptable desynchronization depends on the physiological signals employed. For example, electroencephalography responses (< 1 second) to stimuli take place on a much shorter time frame than that of electrodermal activity (1 to 4 seconds). While such synchronization can be attained between multiple devices in laboratory conditions using tools such as the Lab Streaming Layer [1], the unpredictable network connectivity encountered in the wild and variable support for mobile platforms makes using such tools prohibitively difficult. The most robust solution we found to this problem is to minimize the total number of devices that need to be synchronized by directly streaming and logging data to the participants' smartphone. Using this approach, the Bluetooth connection between the sensor and smartphone is the only remaining source of consistent de-synchronization and is sufficiently small to not be problematic.

In laboratory contexts, experiments rarely last longer than two hours and data can be logged directly to lab-owned high capacity hard drives. Due to the extended data collection period, the volume of physiological data generated *in situ* can introduce serious logistical challenges. For example, in the study presented in Chapter 4, all sensors were sampled at a rate of 64 Hz for

[1] https://github.com/sccn/labstreaminglayer

a duration of approximately 8 hours. This resulted in approximately 1.8 GB of sensor data per measurement session. This volume of data quickly becomes difficult to handle when it needs to be stored on participants' personal devices that are often already at capacity, or have to be transmitted using Canada's prohibitively expensive data plans. The solutions we adopted consisted of requiring at least 2 GB of free storage on participants' smartphones as part of the study's eligibility criteria, which was verified by an experimenter before starting data collection. In addition, the application allowed for data compression, which greatly facilitated their submission at the end of the session. While this was effective and required minimal implementation efforts, compressing the data as it is received, and transmitting it in bulk when the device is connected to a wifi network, would make better usage of limited resources and reduce the risks of data losses.

Finally, do not forget that the "best" is the enemy of the "good". Conducting in the wild research can be extremely challenging, especially if one is used to the rigour and control of laboratory experiments. Without tight control over participants' activities, environment, social interactions and how they wear the physiological sensors throughout the day, ensuring that the signals being collected are representative of what is intended to be measured is much more complex than in laboratory conditions. That being said, what is lost in control is compensated for by the increased ecological validity of the findings. Conducting in-the-wild research is the only way to understand how users truly engage with technology on a daily basis, or in the case of notification research, how they respond to and are impacted by notifications. Even though the human-computer interactions community generally understands the value of in-the-wild research, its limitations and how they impact results, it is important to remember that in the wild studies involving physiological signals are relatively recent. As such, paper submissions will most likely be reviewed by peers who are used to conducting research with physiological signals in laboratory conditions. It is therefore crucial to be particularly explicit about the limitations of the results and to carefully frame the contributions.

7.2 Conducting Research during a Lockdown

The Covid-19 pandemic has disrupted and continues to impede our ability to conduct research with human participants. Indeed, accessing the laboratory, physically meeting with study subjects, and even exchanging equipment within the research team was either impossible or prohibitively complex for most of 2020. This subsection discusses some of the creative approaches we adopted to continue our research activities despite these constraints.

Chapter 5 presented a concrete example of how a laboratory study was adapted to the constraints of the pandemic. Traditionally, TightHR would have been developed using data collected in the lab from a large set of participants to account for differences in factors known to influence optical heart rate measurements' accuracy, e.g., baseline heart rate levels, skin tone, body fat and hairiness. Unfortunately, due to the enforcement of country-wide self-isolation measures,[2] the research team only received permission from the university's research ethics board for one of the co-authors to self-administer the data collection protocol. Given these constraints, the methodology was modified to attempt to collect data from that single participant under different conditions so as to increase the variability of the measurements. The intent was to manipulate variables that are known to have an impact on heart rate in an attempt to increase the measurements' distribution spread and increase its similarity to that of a larger study population.

Knowing that mean heart rate varies significantly throughout the day [4], seven measurement sessions were held each day in order to introduce variability in the collected physiological signals. In addition, three of those measurement sessions were immediately preceded by two minutes of jumping jacks to further modify the participant's heart rate. Data were collected over two days, alternating between the non-dominant and dominant wrist to vary the sensor placement, and thus, the blood vessel and tissue topology directly under the sensor.

As per Figure 7.1, our efforts were successful at introducing variability in heart rate between the different measurement sessions. While whether the distribution obtained is comparable to that of a larger population remains to be validated, these manipulations were sufficient to demonstrate the feasibility and promise of the proposed contact force estimation technique.

Even though self-experimentation is generally frowned upon, results from Chapter 5 demonstrate that it can be used meaningfully in *certain* scientific inquiries. However, it is important to note that such methodology cannot be employed if the variables of interest have any subjective component or could otherwise be biased by the experimenter being aware of the research hypothesis. For example, for the development of SweatSponse, knowledge that physiological responses were hypothesized to occur after the presentation of a notification could have been sufficient to amplify or modify the observed signals and make them unusable. Similarly, the subjective or qualitative evaluation of a system by its creator cannot be considered valid, as it is extremely difficult to remain objective with regards to one's own project. A proof of concept for TightHR, however, was a perfect candidate for such approach. Indeed, both measurements (i.e., contact force and PPG signal) were purely objective, with little potential for experimenter-introduced

[2] `https://en.wikipedia.org/wiki/2019%E2%80%9320_coronavirus_pandemic`

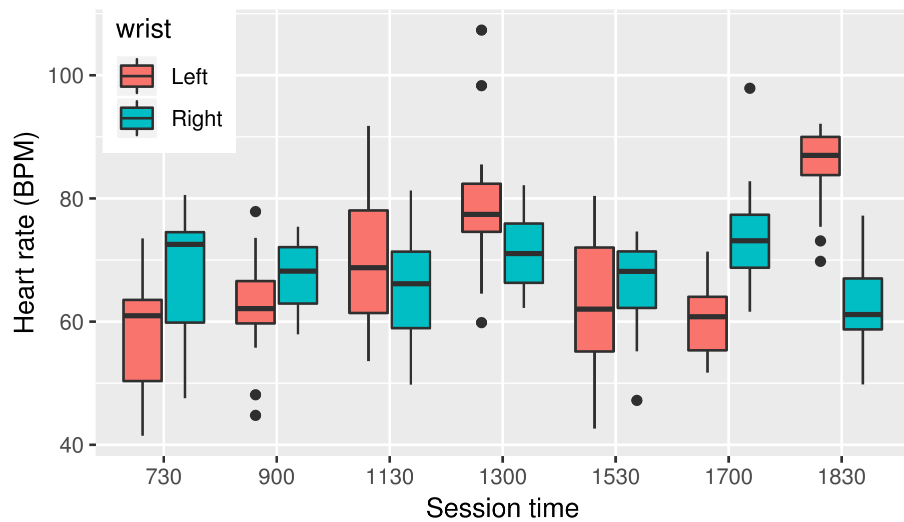


Fig. 7.1 Heart rate distribution at each session time. Differences in means and overall distributions can be seen across session times, suggesting that efforts to create variability were successful.

biases. It is also crucial to be honest about the limitations introduced by using such methodology when reporting the results and fully acknowledge the factors that prevent their generalizability.

The study presented in Chapter 4 was initially designed to be a laboratory extension of the experiment introduced in Chapter 3. That extended laboratory study was supposed to include gaze tracking and pupilometry in addition to electrodermal activity, heart rate, heart rate variability and wrist motion measurements. The intent was to explore the impact of notifications on these signals, but also how participants' physiology is affected by the perception of *someone else's* notifications. Measuring responses to someone else's alerts would have allowed the partial disambiguation of the social relevance and sensory components' contribution to the observed physiological responses. Despite spending months preparing for this new study, pandemic-related public health measures prevented us from engaging in data collection. The experiment design and the physiological signals to be collected had to be reconsidered for an *in situ* study. Despite initially being disappointed by McGill's slow response with regards to the resumption of research involving human participants, the shift from a laboratory to an *in situ* study proved to be a positive one with regards to the findings. Indeed, even though a more thorough understanding of the social and sensory components' contribution to the physiological response would have been interesting, reporting for the first time on physiological changes induced by the perception of notifications *in situ* demonstrated the feasibility of the proposed research methodology.

7.3 Applications

Three mutually reinforcing systems were introduced in this thesis with the objective of advancing the state of the art in notification research and engineering. The following subsections briefly discusses these three systems, how they reinforce one another, and most importantly, envisions their meaningful application in domains beyond notification research.

7.3.1 SweatSponse

In Chapter 3, SweatSponse, a system aimed at confirming the perception of smartphone notifications from skin conductance measurements, was introduced. It was later extended in Chapter 4 by including new physiological sensing channels in its perception prediction system, and most crucially, by demonstrating its performance with data collected outside of laboratory conditions. To the best of our knowledge, SweatSponse is the first technique that allows the confirmation that a notification was perceived, after its presentation, without the need for users to interact with any of their devices (e.g., smartphone, smartwatch or personal computer). While SweatSponse has not been validated in other domains than notification research, we envision that the physiological responses on which it is based extend to other application domains.

For instance, physiological signals are already used to assess users' experience of marketing materials in laboratory conditions [5]. However, this investigation is often constrained to the development stage, as opposed to post-deployment. Assuming perfect time synchronization between users' devices and digital billboards, in the wild perception confirmation could allow this perceptual investigation to be extended into consumers' lives, allowing for the collection of richer, better contextualized responses. In addition, perception-aware advertisement could open the door to a new ad pricing scheme. Indeed, whereas web-based advertisements produce revenue based on the leads or clicks they can generate, their real-world counterparts (e.g., billboards, screens) are priced based on their location and the estimated number of views. Using SweatSponse, real world advertisement could operate with a pay-per-perceive pricing scheme, as opposed to broad assumptions associated with their location.

Finally, SweatSponse could be applied to the problem of emergency signal perception. This would be particularly beneficial in contexts where users' ability to sense their environment is impeded either due to elevated ambient sound or vibration levels, or due to sensory impairments. For example, older adults living alone or in retirement homes may suffer from sensory impairment that could prevent them from perceiving a smoke or carbon monoxide detector's alarm. In such case, the system could be used to confirm whether the alarm was perceived, allowing for a change in the modality used to present the signal (e.g., auditory to vibrotactile, or visual) if needed. In addition, knowledge that a resident has not perceived an alert could be used to allocate first responder personnel more effectively. Similarly, the system could be used in factory or military settings, where elevated ambient sound levels might mask alerts, to ensure the timely communication of safety critical information.

7.3.2 TightHR

One of SweatSponse's most serious limitations is the difficulty associated with collecting high-quality physiological measurements outside of laboratory conditions. Indeed, participants' activity can introduce significant amounts of motion artefacts in these sensitive signals. While automated signal processing tools can be used to detect and remove such artefacts, their models can introduce signal segments that differ from participants' actual physiology. Ensuring that participants wear the sensors at the right location and at an adequate tightness for data collection goes a long way in maximizing data quality and minimizing reliance on techniques that risk denaturing the signal.

Towards the objective of better controlling device tightness in uncontrolled settings, Chapter 5 introduced TightHR, a technique that uses waveform properties of the raw optical heart rate sensor, or photoplethysmograph (PPG), to estimate contact force between the sensor and its user's skin. While researchers would traditionally rely on vague subjective guidelines (e.g., "tight yet comfortable" or use a dedicated force sensor (e.g., force sensitive resistor or load cell), this method allows to estimate contact force with approximately the same precision as a force sensitive resistor, without the need for supplementary dedicated hardware and accompanying electronics. In doing so, TightHR can support researchers conducting laboratory and in the wild studies based on physiological measurements and haptic perception by better capturing the conditions in which data was collected.

To be actionable for end-users, a system using TightHR should ideally use the force estimation to present clear feedback and guide them through the tightness adjustment process. Indeed, without proper guidance, relying on users to choose a subjective "reasonable" strap tightness for coupling is unlikely to yield optimal and repeatable mechanical properties [**?**]. As exemplified in Figure 7.2, TightHR could be integrated in a tightness adjustment recommendation system to guide users in achieving optimal coupling based on a specific applications' requirements.

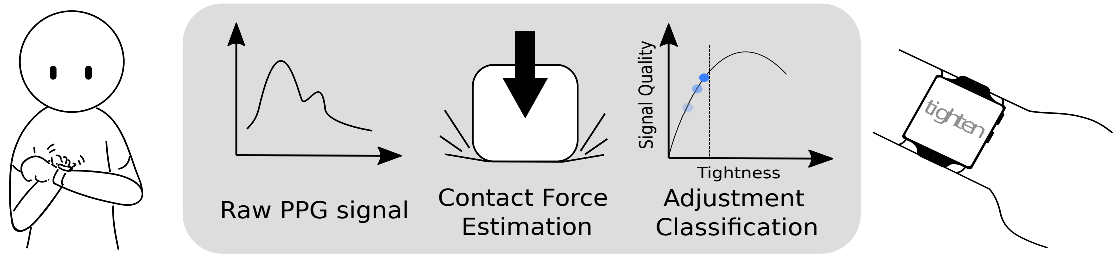


Fig. 7.2 Visual representation of TightHR used in a tightness recommendation system. 1) A user is putting a wearable on. 2) Signals from the optical heart rate sensors are acquired. 3) Contact force is estimated from the properties of the signal's waveform. 4) Tightness of the band is classified. 5) Feedback is presented to the user who adjusts the tightness as needed.

If one's intent is to collect the most accurate heart rate measurements as possible, such a recommendation system could provide feedback to users, ensuring that the device is worn sufficiently tight to collect optimal photoplethysmograph (PPG) signal, while avoiding the over-compression of tissues and blood vessels. The exact contact force to which it should be set would have to be experimentally determined ahead of the deployment and would be dependent on the application. If an optical heart rate sensor is located sufficiently close to the measurement site, this tightness adjustment recommendation could be beneficial to the collection of any contact-based physiological signal, e.g., skin conductance, surface electromyography and electrocardiography, particularly for cases where dry electrodes are used. This application of TightHR could enable better quality physiological parameter estimation in end-user devices (e.g., smartwatches, activity trackers, smartglasses), but also in ambulatory medical devices.

Another application domain that could benefit from more robust and repeatable coupling is that of wearable haptic communication. Blum et al. have identified coupling as one of the most significant challenges to overcome when attempting to conduct haptic research outside of the laboratory. In the worst case, inadequate coupling can result in users' completely missing a haptic signal. However, even if the signal is detected, coupling characteristics are also known to influence the perceived intensity and properties (e.g., frequency) of vibrations, risking signal misinterpretation [**?**]. For commercial devices relying heavily on haptic communication, ensuring that tactile effects are perceived consistently between users and across time can make the difference between a product's success and failure.

7.3.3 Fingerprint Sensor Gesture Reporting Interface

Finally, as outlined by the room for improvement in SweatSponse's classification performance and the need to develop a system such as TightHR to increase the quality of physiological measurements, it is important to acknowledge that sensors, physiological or otherwise, cannot always offer the level of insights and context one would be interested in capturing when conducting research. Beyond the cases where self-reporting would be employed to provide user experience insights that are inaccessible using sensor data, this research methodology could be applied to the collection of information that is redundant with that provided by a physiologically informed system. While measuring a psychological construct using two different methodologies may seem like a significant experimental overhead, the data obtained explicitly via self-reporting could be used to correct erroneous inferences, increasing the reliability of the results. On the other hand, a physiologically informed notification research approach could provide a best estimate for a given psychological construct even when no self-reports are entered. This mixed method would offer the advantages of both methodologies (i.e., the validity of explicit self-reports and the objective implicit insights afforded by psychophysiological analyses), while reducing the potential impact of their limitations on the results of a study. Furthermore, such an approach would facilitate the collection of subject-specific labelled data for continuous system refinements with the objective of reducing its reliance on self-reporting over time.

As an advancement in this methodological space, Chapter 6 introduced a novel gestural interface that embeds itself in the smartphone unlocking process to allow the seamless collection of subjective self-reports. By employing fingerprint sensor gestures, this work adapted the unlock journaling concept to an authentication mechanism that is used by an increasingly large number of smartphone users. In doing so, it enabled its use with a broader audience than was previously possible, while offering reporting performance and perceived intrusiveness that are directly comparable to current state of the art unlock journaling interfaces. Beyond its research application, this novel unlock journaling interface could be used by "quantified selfers" who are collecting various types of data about themselves (e.g., mood, tiredness) to better understand how their environment and behavior influence their physical and mental health [6].

References

[1] M. Pielot, A. Vradi, and S. Park, “Dismissed!: A Detailed Exploration of How Mobile Phone Users Handle Push Notifications,” in *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, ser. MobileHCI ’18. New York, NY, USA: ACM, 2018, pp. 3:1—-3:11. [Online]. Available: http://doi.acm.org/10.1145/3229434.3229445

[2] D. Weber, A. Voit, and N. Henze, “Clear All: A Large-Scale Observational Study on Mobile Notification Drawers,” in *Proceedings of Mensch Und Computer 2019*, ser. MuC’19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 361–372. [Online]. Available: https://doi.org/10.1145/3340764.3340765

[3] W. Boucsein, “Electrodermal activity,” New York, 2012. [Online]. Available: http://site.ebrary.com/id/10517958

[4] D. J. Ewing, J. M. Neilson, C. M. Shapiro, J. A. Stewart, and W. Reid, “Twenty four hour heart rate variability: effects of posture, sleep, and time of day in healthy controls and comparison with bedside tests of autonomic function in diabetic patients.” *Heart*, vol. 65, no. 5, pp. 239–244, 1991. [Online]. Available: https://heart.bmj.com/content/65/5/239

[5] N. Lee, A. J. Broderick, and L. Chamberlain, “What is ‘neuromarketing’? A discussion and agenda for future research,” *International Journal of Psychophysiology*, vol. 63, no. 2, pp. 199–204, 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167876006001073

[6] E. K. Choe, N. B. Lee, B. Lee, W. Pratt, and J. A. Kientz, “Understanding Quantified-Selfers’ Practices in Collecting and Exploring Personal Data,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’14. New York, NY, USA: Association for Computing Machinery, 2014, pp. 1143–1152. [Online]. Available: https://doi.org/10.1145/2556288.2557372

Chapter 8

Conclusion

Just like smartphones, notifications are here to stay, whether we like it or not. It is our responsibility as HCI practitioners and engineers to study and develop technologies to support the research of their impact and to foster a healthier digital future. To do so, however, the notification research community needs to go beyond its current research methodologies to better understand the impact of smartphone notifications on users' psychological and physiological states. Towards this objective, this thesis explored the physiological impact of notification perception and its application to the development of systems aimed at providing unique insights into participants' internal state, a perspective that had not yet been explored in the literature.

Through the presentation of novel findings resulting from a combination of laboratory and *in situ* studies, this thesis successfully fulfilled the objectives it set for itself in Chapter 1. The observation of repeatable skin conductance, heart rate, heart rate variability and wrist motion response patterns following the perception of a notification allows us to conclude that they have a significant impact on physiological signals even outside of laboratory conditions. Beyond reporting on these physiological responses, the promising perception classification performance observed for SweatSponse and its multi-physiological extension demonstrated that these responses are sufficiently robust to allow their use in the study of notifications' impact on user's psychophysiological state. Combined, these two significant contributions prove that current wearable physiological sensing platforms are sufficiently reliable to be employed *in situ*, and that these measurements can, and do, provide insights into participants' psychophysiological state that were previously unavailable to notification researchers. These findings significantly expand the depth and breadth of possible research questions by offering new methodological opportunities, and by demonstrating that perception-aware notification systems could be implemented.

In addition to these fundamental contributions, this thesis introduced two novel systems that partially address serious methodological limitations of physiologically informed notification research and the current active notification research approaches. The first, TightHR, provides contact force estimation between an optical heart rate sensor and its user's skin, offering rich coupling information that can assist in reliably and robustly putting on wearable physiological sensors. Furthermore, its exclusive reliance on raw optical heart rate measurements makes this technique particularly interesting for large scale deployment on existing wearable and medical devices. The second system takes advantage of fingerprint sensor gesture recognition to allow users to complete self-reporting tasks during the smartphone unlocking process. In doing so, it adapted the unlock journaling approach to the reality of fingerprint sensor users, significantly reducing the perceived intrusiveness of the self-reporting interface while remaining comparable in terms of reporting performance.

While novel systems were introduced on the basis of the recent observation of physiological responses to smartphone notifications, this thesis only scratches the surface of new possible research directions and end-user applications these inventions enable. Nevertheless, it already unveils a realm of research questions that will allow the community to understand more thoroughly how notifications impact smartphone users:

- Beyond binary perception prediction, is it possible to use physiological responses to notifications to determine *how* a notification was perceived without the need for self-reports, e.g., stressful, pleasant, interrupting, exciting?

- Knowing users consume their notifications across a number of devices (e.g., smartwatch, smartphone, computer), how does the system on which a notification is presented influence the physiological responses observed?

- Following up on the work that investigated which properties of notifications contributed to the users' stress, what are the relative contributions of the social and sensory components of a notification to its physiological response?

- How are these responses influenced by the experience of problematic smartphone usage patterns, other potentially relevant psychological constructs (e.g., fear of missing out, self-consciousness, social anxiety) or even personality traits?

- Would it be possible to use these physiological changes to anticipate or detect the onset of problematic smartphone usage patterns?