

Detecting Perception of Smartphone Notifications using Skin Conductance Responses

Pascal E. Fortin
Centre for Intelligent Machines
McGill University
Montreal, Canada
pe.fortin@mail.mcgill.ca

Elisabeth Sulmont
Centre for Intelligent Machines
McGill University
Montreal, Canada
elisabeth.sulmont@mail.mcgill.ca

Jeremy Cooperstock
Centre for Intelligent Machines
McGill University
Montreal, Canada
jer@cim.mcgill.ca

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.

CCS CONCEPTS

• **Human-centered computing** → **Interaction techniques**; *Smartphones*;

KEYWORDS

Electrodermal Activity; Smartphone Notifications; Physiological Interaction

ACM Reference Format:

Pascal E. Fortin, Elisabeth Sulmont, and Jeremy Cooperstock. 2019. Detecting Perception of Smartphone Notifications using Skin Conductance Responses. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019)*, May 4–9, 2019, Glasgow, Scotland Uk.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHI 2019, May 4–9, 2019, Glasgow, Scotland Uk

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-5970-2/19/05...\$15.00

<https://doi.org/10.1145/3290605.3300420>

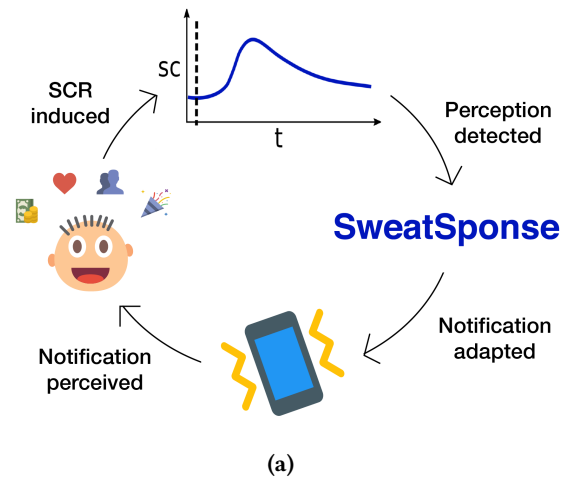


Figure 1: Overview of the proposed feedback loop. A notification is perceived by a user. The user anticipates the potential rewarding social interaction, which induces an SCR. SweatSponse captures the SCR using a wearable sensor, predicts whether it was perceived and feeds the information back to the notification system.

Scotland Uk. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3290605.3300420>

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 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 1).

2 RELATED WORK

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 [6, 13, 15]. 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 [1]. 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 [3]. 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.

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 [5]. 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 [11], followed by a slow decrease until the baseline is reached.

Physiological responses to arbitrary auditory, visual, and vibrotactile 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 [10]. This kind of activation has been correlated with heightened arousal states [16] that are known to influence electrodermal activity [4]. This makes notifications an extremely promising candidate in robustly and repeatably inducing eSCR.

In addition, smartphone notifications follow a variable-ratio reinforcement schedule [16], 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 [16] 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 [13]. As such, the difference between notifications and arbitrary stimuli should be outlined in their more complex habituation and conditioning characteristics.

3 SWEATSPONSE

The aim of Sweatspense 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 1). This relies on Sweatspense'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 [11]) 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 E4¹ 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 [2].

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?

¹Empatica E4

²Thought Technology TPS

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 [14].

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 [13]. To attenuate the impact of those events on the results, consecutive events that were logged less than one second following an initial 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 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 [9] 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:

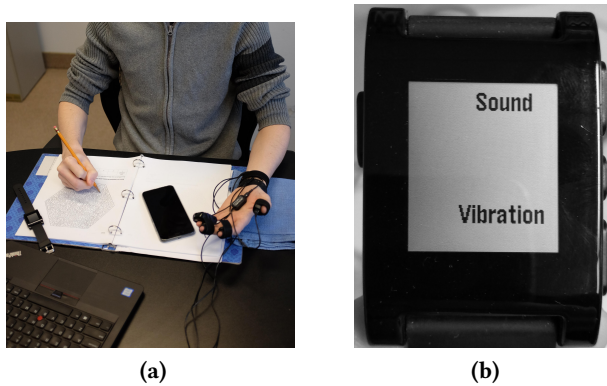


Figure 2: (a) Overall experiment setup (b) Smartwatch graphical interface

- **Inactive (IC):** Participants were asked to watch a wildlife documentary [12]. 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 mazes³ 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 [8].

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

³SRL Maze Task

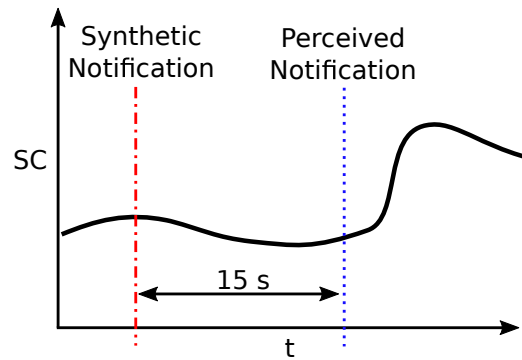


Figure 3: Sketch presenting a notification labeled as perceived by the user, preceded by a synthetic “missed” notification that was never presented to the user.

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 Figure 3). 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.

Skin conductance analysis

All skin conductance signals were processed and analyzed post-experiment using Ledalab⁴ 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 [4]. Continuous decomposition analysis (CDA) was used to extract the phasic activity from the raw skin conductance signal [2]. 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 [11].

⁴Ledalab

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 [16] will be sufficient to trigger a measurable eSCR.
- **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.

6 RESULTS

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.

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.

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.

Skin Conductance Measurements

Building on the data analysis approach used by Andersen et al. and Blum et al. [1, 3], 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 [7]. 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 [4]. Table 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 $\log_2 \text{PhasicMax}$ 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 [3]. 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

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

Null deviance 1624.2 on 1177 degrees of freedom.

Residual deviance 1588.8 on 1173 degrees of freedom

Table 1: Logistic regression analysis summary and Wald’s test output.

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. [1, 3].

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 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 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 [4]. 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 [4]. 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 [4, 7].

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.

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

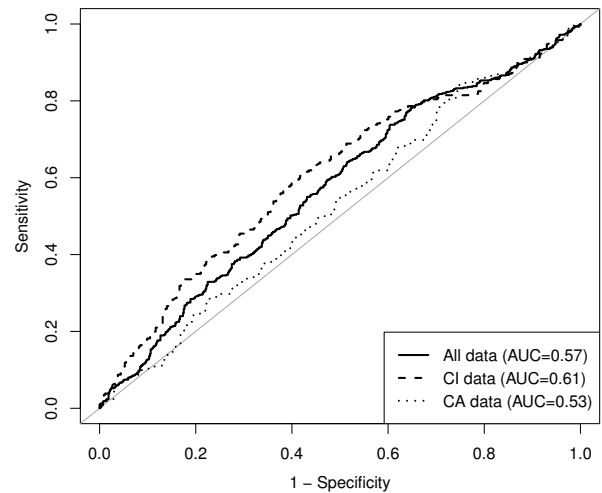


Figure 4: ROC curves generated from the proposed model for the inactive condition (IC), active condition (AC) and aggregated (all data).

the logistic regression model, and receiver operating characteristic (ROC) curves were generated using leave one subject out cross-validation (see Figure 4).

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 ($AUC_{ALL} = .573$, $AUC_{RDM} = .50$, $D = 4.37$, $n.boot = 2000$, $boot.stratified = 1$, $p < .0001$). 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).

7 LIMITATIONS

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 [4]. 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.

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 [11], we would advise against using eSCR for the detection of frequent events (<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 [9]. 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 [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. [3]. 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.

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.

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] Hans Jørgen Andersen, Ann Morrison, and Lars Knudsen. 2012. Modeling Vibrotactile Detection by Logistic Regression. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design (NordiCHI '12)*. ACM, New York, NY, USA, 500–503. <https://doi.org/10.1145/2399016.2399092>
- [2] Mathias Benedek and Christian Kaernbach. 2010. A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods* 190, 1 (2010), 80–91. <https://doi.org/10.1016/j.jneumeth.2010.04.028>
- [3] Jeffrey R Blum, Ilja Frissen, and Jeremy R Cooperstock. 2015. Improving Haptic Feedback on Wearable Devices through Accelerometer Measurements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. ACM, 31–36.
- [4] Wolfram Boucsein. 2012. Electrodermal activity. <http://site.ebrary.com/id/10517958>
- [5] John T Cacioppo, Louis G Tassinary, and Gary Berntson. 2007. *Handbook of Psychophysiology*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511546396> arXiv:arXiv:1011.1669v3
- [6] Yung-Ju Chang and John C Tang. 2015. 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 (Mobile-HCI '15)*. ACM, New York, NY, USA, 6–15. <https://doi.org/10.1145/2785830.2785852>
- [7] Carl Eisdorfer, Hans O Doerr, and William Follette. 1980. Electrodermal Reactivity: An Analysis by Age and Sex. *Journal of Human Stress* 6, 4 (dec 1980), 39–42. <https://doi.org/10.1080/0097840X.1980.9936107>
- [8] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A Hancock and Najmedin B T Advances in Psychology Meshkati (Eds.). Vol. 52. North-Holland, 139–183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- [9] Veronika Konok, Ákos Pogány, and Ádám Miklósi. 2017. Mobile attachment: Separation from the mobile phone induces physiological and behavioural stress and attentional bias to separation-related stimuli. *Computers in Human Behavior* 71 (2017), 228–239. <https://doi.org/10.1016/j.chb.2017.02.002>
- [10] Sören Krach, Frieder M Paulus, Maren Bodden, and Tilo Kircher. 2010. The Rewarding Nature of Social Interactions. *Frontiers in Behavioral Neuroscience* 4 (may 2010), 22. <https://doi.org/10.3389/fnbeh.2010.00022>
- [11] Russell A Lockhart. 1972. Interrelations Between Amplitude, Latency, Rise Time, and the Edelberg Recovery Measure of the Galvanic Skin Response. *Psychophysiology* 9, 4 (1972), 437–442. <https://doi.org/10.1111/j.1469-8986.1972.tb01791.x>
- [12] Henry M. Mix. 2009. Wild Russia - The secret forest.
- [13] Martin Pielot, Rodrigo de Oliveira, Haewoon Kwak, and Nuria Oliver. 2014. 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 (CHI '14)*. ACM, New York, NY, USA, 3319–3328. <https://doi.org/10.1145/2556288.2556973>
- [14] Andrew K Przybylski, Kou Murayama, Cody R DeHaan, and Valerie Gladwell. 2013. Motivational, emotional, and behavioral correlates of fear of missing out. *Computers in Human Behavior* 29, 4 (2013), 1841–1848. <https://doi.org/10.1016/j.chb.2013.02.014>
- [15] Alireza Sahami Shirazi, Niels Henze, Tilman Dingler, Martin Pielot, Dominik Weber, and Albrecht Schmidt. 2014. Large-scale Assessment of Mobile Notifications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 3055–3064. <https://doi.org/10.1145/2556288.2557189>
- [16] Samuel P L Veissière and Moriah Stendel. 2018. Supernatural Monitoring: A Social Rehearsal Account of Smartphone Addiction. *Frontiers in Psychology* 9 (2018), 141. <https://doi.org/10.3389/fpsyg.2018.00141>
- [17] Sunghyuk Yoon and Kun-pyo Lee. 2015. A Study on Notification System Design of Smartphone Messenger Considering the User's Stress. *Archives of design research* 28, 2 (2015), 77–91.