NotiSpeculate: Exploring How Knowing The Notification Information In Advance Affects Users Treat Them

Tang-Jie, Chang
tjchang.cs08@nycu.edu.tw
National Yang Ming Chiao Tung University
Hsinchu, Taiwan, Taiwan

Jie, Tsai
hazel.tsai@utexas.edu
University of Texas at Austin
Hsinchu, Taiwan, Taiwan

Zi-Xun, Tang
jasonxduck.cs08@nctu.edu.tw
National Yang Ming Chiao Tung University
Hsinchu, Taiwan, Taiwan

Li-Ting, Su
liitiisu.cs10@nycu.edu.tw
National Yang Ming Chiao Tung University
Hsinchu, Taiwan, Taiwan

Yung-Ju, Chang
armuro@cs.nctu.edu.tw
National Yang Ming Chiao Tung University
Hsinchu, Taiwan, Taiwan

ABSTRACT

Prior studies on interruptibility have concentrated on identifying opportune moments or interruptibles for users to deal with notifications. However, there hasn't been much study about how notifications would be managed if users know the information of them before attending to them. In order to reach this goal, this study outlines the first research effort to change the alert (ringtone or vibration) that brings information about an arriving notification. We developed *NotiSpeculae*, which was developed in Android and replaces the original alert with a customized alert depending on user-defined categories. A 3-week ESM study with 37 users who used *NotiSpeculate* demonstrates that, first, alerts' characteristics, keyword setting of user-defined categories, perceived disturbance, speculation and selective attendance on notifications would effect each other. Second, the moments that users would probably know the source of notifications. Finally, some design recommendations for systems or apps that aim at improving notification speculation or attentiveness were made.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in ubiquitous and mobile computing**.

KEYWORDS

Mobile notifications; mobile receptivity; speculation; attentiveness; ESM

1 INTRODUCTION

Our everyday lives depend heavily on our mobile devices to be informed about events, schedules, and messages [25]. Due to widespread use of smartphones and phone communication, notifications on a mobile phone can become overwhelming. Unfortunately, as people add more apps to their smartphones, they are more likely to get alerts from notifications that they don't care about.

Many notifications, which are irrelevant to users' needs, can be disruptive and annoying [19]. There has been speculation among researchers regarding the potential benefits of disabling notifications to avoid interruptions. However, studies have shown that such a measure instead leads to increased stress among participants and does not reduce the frequency of phone checking [27, 28]. Conversely, certain studies have reported that notifications from specific applications or individuals are highly valued and appreciated by users [17, 32, 38]. Hence, the annoyance and frequent phone checking associated with notifications may also stem from users' lack of information about the content of the notifications.

Users have grown accustomed to recognizing the arrival of a new notification by its accompanying ringtone and vibration. Previous studies referred to this phenomenon as "notification speculation" [5], as users tend to make guesses about the source of the notification. Therefore, the alerts can serve as hints for users to speculate about the notification and decide whether or not to attend it. However, relying solely on sound and vibration alerts is not always reliable for accurate speculation. While some mobile apps have unique custom alerts, many still use default or similar tones. We propose that providing additional information through alerts could enhance users' speculation abilities and their selective attendance to notifications.

To examine users' reactions to notifications when they possess information from the notifications without actually attending to them, we implemented a user customization feature in which users could define their preferred categories of notifications and set keyword triggers. Subsequently, we developed an Android application named *NotiSpeculate* to deliver customized alerts corresponding to the triggered category. This research aims to address the following research questions: 1) To what extent does the usage of "NotiSpeculate" assist users in speculating the source of notifications? 2) Does "NotiSpeculate" improve the effectiveness of selective attendance of notifications? 3) To what extent does the incorporation of context awareness in "NotiSpeculate" reduce disturbances caused by notifications?

We executed a comprehensive three-week study utilizing phone logs, the experience sampling method (ESM), and interviews to investigate the topic. This study offers three key contributions. Firstly, we introduced the method of utilizing customized alerts based on the category of notifications, which significantly enhanced users' ability to speculate on notifications. Moreover, we explored how different keyword settings impacted users' speculation. Second, it reveals the relationship among alert's frequency, alert's familiarity, users' perceived disturbance, performance on speculation and selective attendance. Finally, it provides recommendations on keyword

setting and alerts for the design of future notification systems that will help support notification speculation and selective attendance.

2 RELATED WORK

Much prior notification research have examined users' notification management and their attentiveness to notifications. Research have shown that notifications are attended to within several minutes [7, 11, 26, 32], even when the phone is in silent mode [25]. Chang et al. [5] have found that users would speculate the sources of the notification when they sense it and they were more likely to attend to the notification if they speculated about the notification.However, not all notifications are considered useful by the user. Research have shown that users value the notification related to people (e.g., instant-messaging (IM)) most highly [29, 32], whereas notifications from the system app are less preferred [32]. Relationship type is found to affect users' perceptions of receiving notification [17, 20, 21, 41], also Yuan et al. [41] demonstrates that the users were more receptive to interruptions from persons they knew as opposed to complete strangers.

Additionally, researchers have sought to investigate how context affects users' receptivity to interruptions, including activity type and level [1, 3, 6–8, 13], location [20, 24, 34, 40], time of day [24, 30, 31, 34, 40], recent phone interaction [23, 32], ringer mode [9, 13], sensor information [9, 13, 15, 16, 18, 21], conversational context [35, 36], personal characteristics [41], and arousal and arousal emotional states [12, 22]. [1, 10, 19, 23] showed that predicting opportune moments and deferring notifications until these moments can help reduce distraction, researches also show that disabling notifications could help users avoid interruptions [27, 28]. However, disabling notifications may instead made users frequently checked their phone because they may afraid missing notification [27, 28]. Users' propensity to regularly check their phones can be attributed to this "more than enough" notification-reading, but it's also possible that it's caused by their low-awareness of what triggers the notification, which makes them curious. Since users' attentiveness of notification and their perception of receiving notification will be affected by the content and sender of the notification, but users have no idea of what type of content and sender before attending to the notification, such uncertainty may be the reason why users got interrupted by the notification, they afraid transfer their attention to the phone to see the notification they don't want to see but at thee same time worried about missing important notification. This study help users know more information about the notification at the time it comes, to see if it will improve their experience of receiving notification.

3 METHODOLOGY

This study contains a three-week study using a combination of approaches, including phone logs, ESM, and interviews.

3.1 Vibration And Ringtone Design

3.1.1 Designing of vibration. A vibration pattern consists of multiple small fragments of vibration and gaps, with the number of fragments being n and the number of gaps being n-1. When designing a vibration pattern, participants only need to determine two factors. Firstly, they specify the desired vibration sensation to be experienced on their phones. Secondly, they define the duration of each small fragment of vibration and the accompanying gap. In our study, we utilized the short and long patterns previously described in Saket et al. [33]. We presented users with various combinations of these two patterns to demonstrate the influence of different vibration patterns on the perception of emergency situations.

3.1.2 Designing of ringtone. Following the ringtone design guideline [4], we incorporated a collection of 140 ringtones, encompassing variations across three dimensions: mode, melodic arrangement, and timbre. First, people's emotional responses to the sound could be influenced by the mode of the sound [14]. Therefore, to ensure a broad spectrum of emotional reactions among our ringtone selections, we manipulated the mode by utilizing four chords ranging from major to minor mode.

We considered melodic patterns, specifically the melodic direction and tempo, in our ringtone design. Prior research has indicated that tempo influences emotional reactions [2, 14, 39], while different melodic directions elicit various emotions [2, 39]. To broaden the emotional range of our ringtones, we created five melodic patterns that varied in tempo and direction.

The final aspect is timbre, which, as previous studies indicate, elicits diverse emotional reactions [2]. Our study incorporates a variety of timbres, encompassing instrument sounds, real-world sounds, and synthesized sounds, totaling seven in number.

3.2 Participant Recruitment

We recruited Android smartphone users who received notifications through a ringtone or vibration. The participants, 1) whose phones are in normal or vibration mode for more than sixteen hours a day, and 2) whose phones always connect to the internet, were recruited. We posted our recruiting message in a number of forums and Facebook pages. A total of 37 participants joined the study and completed it. Among the 37 participants, 14 were female and 23 were male, ranging in age from 20 to 47, and 18 of them were students. The remainder were from a variety of job sectors, including manufacturing, information technology, and entertainment, among others. At the end of this study, 27 of them had participated in an optional interview.

3.3 Preset of the experiment

3.3.1 Self-defined notification's categories and their corresponding keyword setting. We built a website for participants to set their categories. Participants can freely set their categories (type the category name in red box in Figure 1a). For each category the participants add, they need to finish the keyword setting for it. We provided three types of information about notifications that participants can use for setting, which are: 1) sender; 2) content; and 3) app (as shown in Figure 1b's orange box). Each information the participants set would add a keyword-setting rule to its corresponding category; if the arriving notification satisfied the rule, it would be classified to the category with this rule, and *NotiSpeculate* would then play the ringtone or vibration that the participants set for this category based on the participant's phone's current ringer mode. Each category can have multiple keyword-setting rules. Rules that satisfy two or more pieces of information at the same time can also be set. So participants can set seven types of rules, which are

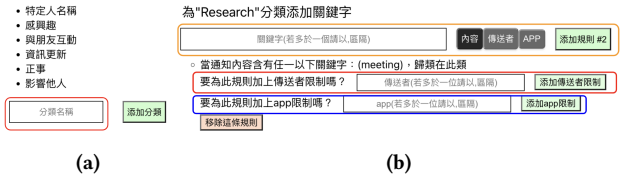


Figure 1: Interface of category setting website

1) sender, 2) content, 3) app, 4) sender-content, 5) sender-app, 6) content-app, and 7) sender-content-app.

3.3.2 ringtone setting. We display all ringtones on the website. The website has seven sub-pages; each sub-page shows one type of timbre and contains 20 ringtones (each ringtone is a combination of one of the four modes and one of the five melodic arrangements, so each sub-page has 20 ringtones). Once participants find the ringtone they want, they can go back to the category setting website to select it and complete the ringtone setting.

3.3.3 preset schedule. A pre-study meeting took place three days before the experiment, where participants were given informed consent and instructions for presetting category, ringtone, and vibration settings. A subsequent appointment was scheduled at least three days after the meeting for participants to install NotiSpeculate and complete the preset before the experiment began.

3.4 Experience Sampling Study

ESM was used to study participants' reactions and experiences with notifications. Each notification had a specific questionnaire with details about the sender, app source, arrival time, and content. Participants answered questions regarding speculation (correctness and helpfulness on a Likert scale), attendance (reasons for their decision to attend or disregard, self-evaluation of its beneficial at that time, and helpfulness on a Likert scale), and perceived disturbance (measured on a Likert scale). An "unsure" option was provided for uncertain notifications.

Our Android app, *NotiSpeculate*, facilitated ESM questionnaires by allowing participants to select 12-hour windows for receiving them each day. The app sampled notifications based on predefined criteria, ensuring an equal number from user-defined categories and other notifications. Time considerations balanced the app's waiting period for rare notifications and the possibility of an entire day without such notifications. Upon sampling, the app waited for one minute to account for immediate participant reactions.

To minimize disruption, the research app sent silent and non-vibrating notifications for ESM questionnaires. A 30-minute threshold was set to capture diverse scenarios without compromising recall. Participants received no more than eight questionnaires per day, with a minimum one-hour gap between each questionnaire phase.

3.5 Study Procedure and Data Collection

During the 21-day period, participants used the *NotiSpeculate* app on their phones, which recorded notifications, participant behavior, and phone context data such as sensor readings, location, activity, and phone status (e.g., battery life, network connectivity). Participants underwent three conditions, each lasting seven days: "AWARE" (AA), where customized ringtones or vibrations based on notification categories were played; "CONTEXT AWARE" (CA), where no customized alerts were played if the user had used specific instant messaging apps and received notifications from the same sender and app in the previous 30 minutes; and "BASELINE" (BL), where no customized alerts were played. To eliminate condition order effects, participants were randomly assigned to different condition sequences. The first two days after switching conditions were excluded from analysis to account for participants' adaptation period. The remaining five days were workdays for consistent data collection. Participants completed a familiarity test for customized alerts before entering conditions involving such alerts (AA, CA). After the ESM research concluded, participants received a reward of 1,500 NTD (about 50 USD) and were invited for interviews. During the interviews, participants were shown their notifications and ESM responses for background information and clarification. The interviews were audio-recorded and transcribed. Participants who participated in interviews received a voluntary bonus of 200 NTD (about 7 USD).

3.6 Data Cleaning and Analysis

During the study period, the participants collectively received 164,794 phone notifications and completed or partially completed 2,897 ESM questionnaires. Of which 997, 1,833, and 67 were in Normal, Vibrate and Silent modes, respectively. Among the 2,897 ESM responses we analyzed, 1,550 of them sampled notifications that were not in any category (which was called the "default" category), while the remaining 1,347 sampled notifications were in any category (which was called a "classified notification"). For the default notifications sampled by ESM, 40.77% (632) were reported as sensed, with 197 being immediately seen because they were using a phone or other device, and the remaining 435 ESM responses being reported as sensed. For classified notifications, 59.91% (or 807) were sensed when they arrived; of these sensed notifications, 241 were immediately seen because they were using a phone or other device, and the remaining 566 ESM responses were reported as sensed. On average, each participant completed 78.3 of the questionnaires that were analyzed.

For statistical analyses, we used the R software. In examining effects of specific independent variables on a dependent variable, we used a mixed-effects regression model, given that each participant contributed multiple data points, where a random effect of *participant* was included to account for individual differences.

4 RESULTS

4.1 NotiSpeculate did help speculating the source of the notification

4.1.1 Customized alert did help speculating notification's source. First, we compared the correctness of overall notifications' speculation in conditions AWARE and CONTEXT AWARE to condition BASELINE to see which condition would improve the performance on speculating. Figure 2 shows the distribution of correctness of speculation on all notifications in each condition, but it shows that excepted the correctness of speculating category in condition AA, there has no significant difference when compared to condition

BASELINE, when compared to condition AA (for category, z = 1.961, p = 0.049827; for sender, z = 1.429, p = 0.153; for content, z = -0.026, p = 0.980; for app, z = 1.255, p = 0.209), when compared to CA (for category, z = 0.401,, p = 0.688463; for sender, z = -1.104, p = 0.270; for content, z = -0.804, p = 0.422; for app, z = 0.148, p = 0.883). Since our goal is to find out whether customized alerts improve the performance of speculation, we then compared the notifications that arrived without customized alerts (which were default notifications) to the notifications that arrived with customized alerts (which were classified notifications) to see whether customized alerts improve the performance of speculating. Table 1 shows the correctness of speculation between default and classified notifications; it got significant differences on all sources (for category, z = 8.310, p < 0.001; for sender, z = 3.811, p < 0.001; for content, z = 3.504, p < 0.001; for app, z = 4.564, p < 0.001), which means that notifications that arrived with customized alerts did help participants speculating on them.

We confirmed that customized alerts would definitely improve the correctness of speculating sources. Next, we compared the mechanisms between always-play (which is condition AA) and conditionally-play (which is condition CA) to see how the difference in the mechanisms would affect the correctness of speculating. Figure 3 shows the correctness of the speculation on default and classified notifications in each condition. The reason why condition BL only had default notifications was because there were no customized alerts in condition BL. First, we compared classified notifications in condition AA to default notifications in each condition (which were default notifications in condition AA, CA, and BL), and there were significant differences when comparing to default notifications in AA (for category, z = 6.272, p < 0.001; for sender, z = 3.102, p = 0.00192; for content, z = 3.743, p < 0.001; for app, z = 4.366, p < 0.001), in CA (for category, z = 6.239, p < 0.001; for sender, z = 3.576, p < 0.001; for content, z = 2.390, p = 0.0168; for app, z = 3.637, p < 0.001), and in BL (for category, z = 5.186, p < 0.001; for sender, z = 3.420, p < 0.001; for content, z = 2.000, p = 0.0455; for app, z = 3.517, p < 0.001), this result indicated that the mechanisms of always-play did improve the performance on speculating. Following that, we compared classified notifications in condition CA to default notifications in each condition, but we discovered that customized alerts in CA did not improve the performance as well as that in AA, when comparing to default notifications in AA, it only had significant differences on speculating category and app (for category, z = 5.042, p < 0.001; for sender, z = 0.111, p = 0.911; for content, z = 1.241, p = 0.215; for app, z = 2.723, p = 0.00648), only had significant differences on speculating category when comparing to default notifications in CA (for category, z = 4.409, p < 0.001; for sender, z = 0.161, p = 0.872; for content, z = 0.092, p = 0.927; for app, z = 1.307, p = 0.19133), and had significant differences on speculating category and app in BL(for category, z = 3.934, p < 0.001; for sender, z = 0.877, p = 0.380; for content, z = 0.091, p = 0.927; for app, z = 2.399, p = 0.0164).

These results indicated that, while an always-play customized alert (condition AA) would undoubtedly help users speculate about the source of the notification, playing customized alerts on occasion (condition CA) may not improve the speculation's performance for all sources.

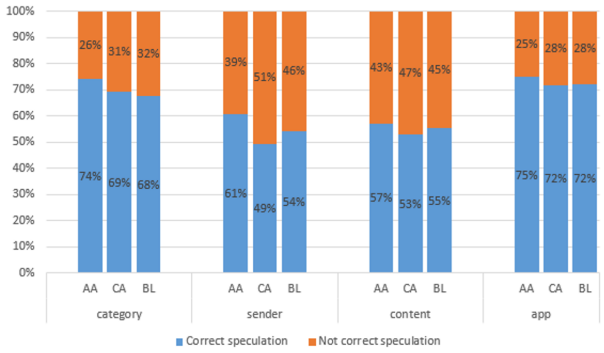


Figure 2: Correctness of the speculation on all notifications for each source (category, sender, content, and app) in AA(condition AWARE), CA(condition CONTEXT AWARE), and BL(condition BASELINE)

Table 1: Correctness of the speculation on default and classified notifications for each source (category, sender, content, and app)

	category	sender	content	app
Default	61%	52%	52%	68%
Classified	88%	61%	61%	68%

4.1.2 Correctness of speculation were relate to what trigger the notification's category. We were curious about how different types of information that the participants knew through customized alerts would affect their speculation. We then took a detailed look at the three small groups, 1) only-sender, which means the participant would only get sender information through the customized alert (from one of the possible senders). 2) only-content, which means the participant would only get content information through the customized alert (one of the possible contents). 3) only-app, which means the participant would only get app information through the customized alert (one of the possible apps)(their correctness of speculation shows in Table 2). Using only-one-clue (included only-sender, only-content, and only-app) as an IV in a logistic regression model with only-app group as a reference class, only-sender performed better on speculation (for category, z = 1.360, p = 0.17369; for sender, z = 2.177, p = 0.0295; for content, z = 1.994, p = 0.0461; for app, z = 2.937, p = 0.00331), but we did not see any significant differences on only-content (for category, z = -1.091, p = 0.27521; for sender, z = -0.461, p = 0.6448; for content, z = 0.549, p = 0.5833; for app, z = -0.823, p = 0.41032), next, we compared the performance of speculation's correctness between only-sender and only-content, found that only-sender got significant differences on speculating sender and app (for category, z = 1.822, p = 0.0685; for sender, z = 2.114, p = 0.0346; for content, z = 0.798, p = 0.425; for app, z = 3.099, p = 0.00194). To concluded, the only-sender group predicted the correctness of speculation on the sender, content, and app with a statistically significant difference; so, the sender information may be a better clue in speculating the source of the notification.

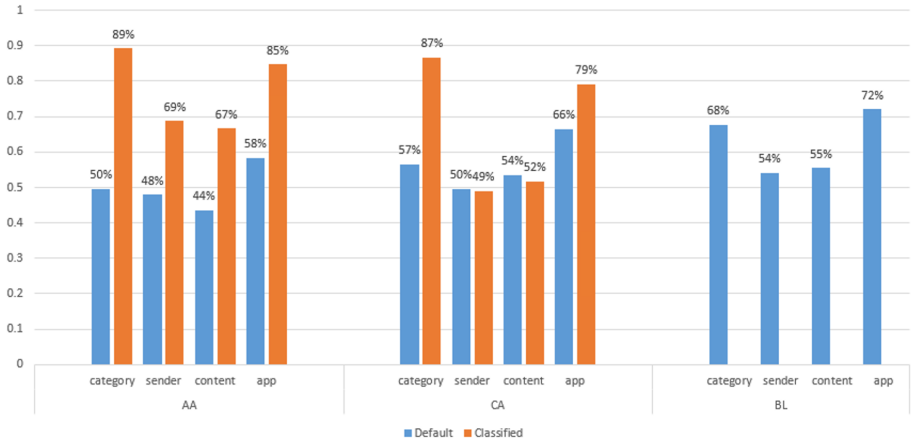


Figure 3: Correctness of the speculation on default and classified notifications for each source (category, sender, content, and app) in AA(condition AWARE), CA(condition CONTEXT AWARE), and BL(condition BASELINE)

Table 2: Comparison of the correctness on speculation between only-sender/content/app group

	category	sender	content	app
only-sender	82%	69%	37%	87%
only-content	77%	42%	58%	54%
only-app	71%	31%	27%	62%

Table 3: Rate of decide to see and not to see in each condition.

	AA	CA	BL
decide to attend	66%	68%	61%
decide not to attend	34%	32%	39%

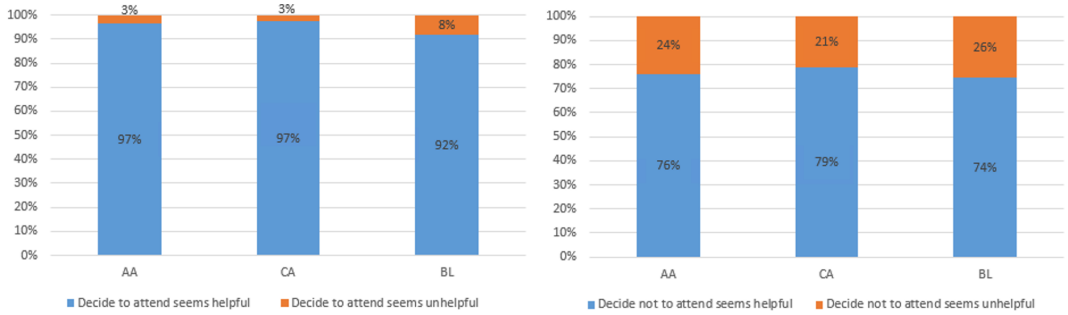
4.2 NotiSpeculate improve the effectiveness of selective attendance partially

4.2.1 Customized alerts help participants decide to attend to notifications. Table 3 shows the distribution of their rate on deciding to attend or not attend. Figure 4a shows the helpfulness of the participants' decisions after deciding to attend to notifications in each condition. Wondering if customized alerts would make participants attend to notifications more effectively, we then ran a logistic regression using condition (included AA, CA, and BL) as an IV with BL (91.87%) as a reference class. 96.62% ($z = 2.141$, $p = 0.03230$) of these decisions seemed helpful in condition AA, 97.42% ($z = 2.704$, $p = 0.00685$) of these decisions seemed helpful in condition CA. These findings indicated that customized alerts did help participants attend to notifications more effectively, as the instances of deciding to attend to a notification but later discovering it was unimportant or not wanting to see messages were reduced.

4.2.2 Decide not to attend but seems the decision was unhelpful. However, there are no significant differences between the comparisons from AA to BL ($z = 1.051$, $p = 0.293211$) and from CA to BL ($z = 1.034$, $p = 0.300976$) on the helpfulness of deciding not to attend to notifications (as shown in Figure 4b), these results indicated that customized alerts would not decrease the instances when people were missing important or wanted to see messages. Table 4 depicts the distribution of reasons for deciding not to attend but ultimately making the wrong decision. Because classified notifications are more wanted to attend to than default notifications, it is natural that a large proportion of the reasons why participants choose not to read (61.22% vs. 28.57%) is that they do not have time to do so at that time. However, another reason we believe the number of people deciding not to attend is high is due to insufficient category settings. Even though the participants have classified the notifications into different categories, there is still a different level of urgency for those notifications to be assigned to the same category. In the situation when the users were busy or did not have enough time to read, even if they knew there was a classified notification, they may decide not to attend to it because both an important and a relatively unimportant notification can trigger the same category, and they could not accurately judge whether it was necessary to immediately read it at the moment.

Table 4: Distribution of reasons for deciding not to read but at last making a wrong decision.

	default	classified
No time to read at that time	28.57%	61.22%
Think it may take time to respond	17.14%	28.57%
Think sender doesn't want to see	0%	0%
Think content doesn't want to see	11.43%	6.12%
Think sender is not important	8.57%	6.12%
Think content is not important	8.57%	6.12%
Think sender is not urgent	25.71%	12.24%
Think content is not urgent	34.29%	30.61%



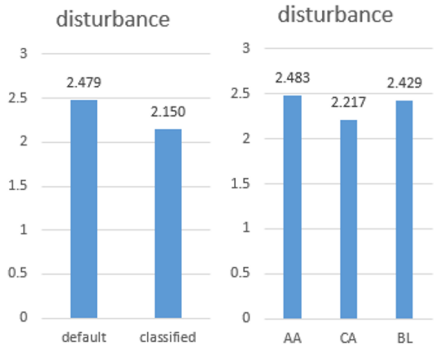
(a) (b)

Figure 4: (a) Helpfulness of the decision after deciding to attend notifications in each condition; (b) Helpfulness of the decision after deciding not to attend in each condition

4.3 NotiSpeculate did not cause additional disturbance

We were curious about whether the customized alerts played from *NotiSpeculate* would disrupt the participants or not (as shown in Figure 5a), and we found that the score of disturbance caused by classified notifications was lower than that in default notifications and had a significant difference compared to the score in default notifications ($z = -2.591$, $p = 0.00973$), which means that participants did not think customized alerts caused additional disturbance; they even thought the default alerts were more disruptive compared to the customized ones. Participants also gave positive feedback on their experience with the alert that comes from *NotiSpeculate*, of the 27 interviewees, only five said the alert that came up with the notification would cause disturbance. Some participants stated that they were accustomed to the alert in their phone without *NotiSpeculate*, so the alert played by *NotiSpeculate* caused no additional disturbance.

On the other hand, the main reason for those participants who said *NotiSpeculate* cause disturbance is the strong existence of the ringtone or vibration, like U27 said: "*Sometimes I am frightened by the vibration when I am working, because one of the vibration patterns I set for the category has a long vibrate.*", and U09 said: "*If there is a long interval between messages, then the alert is acceptable, but if there are too many messages coming suddenly, the alert will be very noisy.*". And this circumstance of many alerts all occurring at once would appear in condition AA but not in condition CA, and maybe that's the reason why condition CA got the lowest score for disturbance among all conditions (Figure 5b). In condition CA, *NotiSpeculate* would not play a customized alert if participants had used the same app that the arriving classified notification came from in the past 30 minutes and the notification had the same sender and app as the previous notifications (35.47% of classified notifications in condition CA meet this mechanism). In condition AA, if the same sender sent multiple messages in a row, there would be more than one alert. However, only the first alert gives more information about the source of the notification; the subsequent alerts may be annoying to participants.



(a) (b)

Figure 5: Perceived disturbance score (a) in default and classified notifications; (b) in three condition AWARE(AA), CONTEXT AWARE(CA), and BASELINE(BL)

4.4 Condition Context Aware help speculating in partial

Through the reduction of classified alerts, CA successfully mitigated disturbance. However, we were curious about whether reducing alerts based on context awareness from classified notifications would impact the speculate rate. Hence, we divided the classified notifications in CA into two conditions: 1) CAWO (classified notification without alert), which included notifications from the same sender and app as the last notification, and when the participants had used the app in the past 30 minutes; and 2) CAWA (classified notification with alert). We then compared these categories with the classified notifications in AA. Notifications in condition "CAWO" got the highest correctness of speculation on sender (see Figure 6), content, and app (which are 75.00%, 71.05% and 86.96%, respectively). Despite the lack of assistance, the correctness of speculation in condition CAWO is still higher than that in condition AA (for sender, $v = -0.571$, $p = 0.5678$; for content, $v = 0.021$, $p = 0.9835$; and for app, $v = -0.020$, $p = 0.98407$), implying that the user will have a clear mind on who the sender or content is in the context of an arriving notification from a recently used app compared with other

situations. But on the other hand, classified notifications in condition CAWA performed poorly (which are 86.61%, 49.04%, 51.61% and 79.09% for category, sender, content, and app, respectively), especially the speculation on sender and content. So we ran the logistic regression with condition AA as a reference class since both conditions play customized alert all the time (for category, z = -0.564, p = 0.573; for sender, z = -1.836, p = 0.06638; for content, z = -1.366, p = 0.1719; for app, z = -0.593, p = 0.55337), but found that there is no significant difference in speculating all sources of the notifications between conditions AA and CA, so the difference between the overall correctness in condition AA and CA may be influence by few of the participants who were struggled in speculating the source of classified notification in condition CAWA.

5 DISCUSSION

5.1 Keyword setting's impact on speculation and attendance

The only-sender group did better at speculating than the only-content and only-app groups. The result showed that there was a strong link between setting keywords and speculating correctly about the notification's source. Avoid general settings terms, such as the word that phone notifications frequently contain, when setting keywords; if you must set it as a keyword, add additional information to narrow down the possible notifications that would be classified to that category. We provided two ways to improve the setting of keywords: 1) system aspect; 2) setting aspect. First, the system aspect implies that users can get started quickly with the system's assistance. 1) recommend list of categories or keywords: the system can provide categories and corresponding keywords that people mostly want to set, and while setting keywords, the system can recommend some other words that are similar to those; 2) Notification-log summary: the system can trace the usage of users' notifications and record the state of attendance, when setting keywords, it will show the notifications that the users had attended before, these usage data would provide users with a basis for setting keywords. Secondly, the setting aspect means adding additional rules to fully complete the setting of keywords: 1) keywords that have a time restriction, for example, keywords that will be effective on Wednesday or from 9 a.m. to 9 p.m.; 2) keywords that are excluded; for example, notifications that contain this type of keyword would not play any alerts. We had some issues and inconveniences with the system's first attempt at setting keywords. Future research can consider the recommendations stated above when designing the keyword setting system.

5.2 Confidence on selective attendance

Previous research [5] showed that users' attendance would be affected by their speculation on notifications, and their speculation basis would be affected by the alert that a notification brings. Our results show that by setting user-defined categories and giving category information through customized alerts, we help users speculate about the sources of the notifications. With *NotiSpeculate*, users had a clearer idea of what the possible information in an arriving notification would be, which improved the effectiveness of deciding to attend to notifications. However, we did not assist users in reducing the instances of missing important messages, which were notifications that users decided not to attend but later discovered should be seen immediately; most of these notifications went unnoticed because users were busy at the time, which we attributed to the uncertainty that customized alerts bring, and this would affect their speculation on notifications and further influence their decision on whether to attend. The uncertainty mainly came from two parts: 1) Unfamiliar with the customized alerts, this would let users hesitate on speculating and affect their attendance; 2) Without a complete setting of the keywords' rules, users would not believe the categorization system because there may exist some unexpected notifications in categories. As a preliminary attempt to change the original phone's alert to a customized alert based on the different category, *NotiSpeculate* have improved the effectiveness of the selective attendance on notifications. We believed that a system like *NotiSpeculate*, with better alert selection and keyword settings, would definitely prevent users from missing really urgent and important notifications. However, there is still a lot more room to be explored and discussed in the future.

5.3 Trade off between alerts and disturbance

In prior research [37], users did not like the additional alerts because they thought the alerts were disruptive, but in our study, the customized alert caused less disturbance than the original alert. We think there are three reasons to explain this circumstance. First, we replace the original alerts with customized alerts rather than adding them after playing the originals; this setting would directly reduce half of the sounds when we think users need assistance in speculating the source of notification. Second, customized alerts truly assist users in speculating the source of notification and truly assisting them in attending to notification; the assistance may make the customized alerts more acceptable, resulting in less disturbance. Last, we provided users with 140 ringtones to choose from and let them freely set the vibration for each category; this may increase the chance for them to select the alerts they want to hear and make it more acceptable.

Some users claimed that the customized alerts caused disruption when there were a large number of messages in the same category at the same time, resulting in consecutive customized alerts, but this may be due to the user's unfamiliarity and unaccustomedness to customized alerts, which make it easier for users to notice them than default alerts. The increase in noise is due to subsequent alerts that do not provide any new information, causing disruption as the customized alerts divert users' attention. However, this disruption may be mitigated as users become accustomed to the customized alerts.

Users' perceived disturbance would be affected by many factors, we can not judge that always-play customized alerts did not cause any additional disturbance than conditionally-play customized alerts, we just can say that in our system like *NotiSpeculate*. Our system was not good enough. To improve our system, there are many aspects to consider: 1) alerts, which provide more intuitive and familiar alerts so users can get used to these alerts more quickly, leading to precisely speculating the sources of notification; 2) keyword setting, providing assistance on their setting; a complete set of rules for keyword setting would affect users' confidence in speculating on notifications' sources; and 3) alerts' playing mechanism,

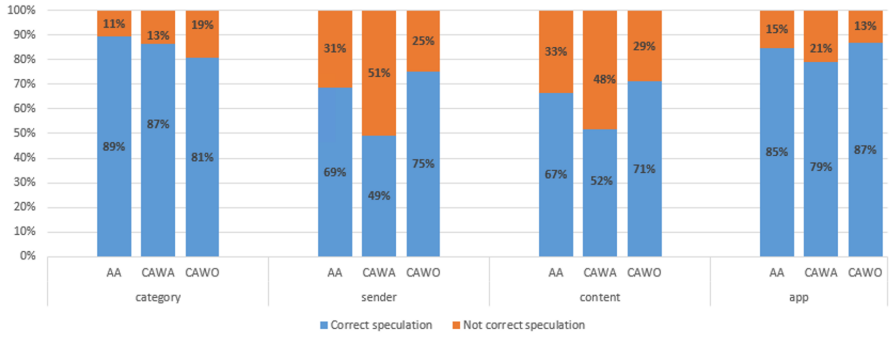


Figure 6: Comparison of the correctness of speculation in conditions AA, CAWA, and CAWO

the playing frequency of customized alerts would affected the perceived disturbance for some users and their familiarity with alerts. We believe future systems that take these three aspects into account would definitely improve users' speculation on notifications and their effectiveness in selectively attending to them.

6 RESEARCH LIMITATION

This paper is subject to several limitations. First, the way the study was set up made it impossible to track how people used desktop or web apps or wearable devices, which could affect how they really interacted with notifications. For example, a participant might decide not to answer an IM notification on their smartphone if they already replied to the relevant message on their laptop. But our ESM questionnaires did not ask participants to tell us about what they did on other devices. So, we couldn't tell how much these devices might have influenced their speculation and attendance. Second, to reduce recall bias in the ESM, we only sampled notifications that were sent within 30 minutes of the last time a participant used their phone. Due to the length of the ESM questionnaire, self-reported attentiveness could have been skewed toward times when participants were more attentive to their phones. Third, due to the specifications of our target scenario, we did not consider users who mainly kept their phones in Silent Mode. These users might display different kinds of speculation when they switch their phones to Normal or Vibrate Mode. Nor did this research consider alerts other than those based on ringtone and vibration, e.g., the flashing lights offered by certain phones. Finally, we conducted our field study in Taiwan; most of our participants were in their twenties, and half of them were students. As a result, it is unclear whether our findings can be generalized to another group of people from other age groups and/or cultures.

7 CONCLUSION

Alert systems are intended to inform users of the arrival of notifications on their phones, but users would not receive additional information about notifications through alerts. This paper had a first attempt at changing original alerts on the phone to customized alerts that bring user-defined information and found that the system using this approach was flavored by participants. Our ESM study with 37 participants had three high-level takeaways. First, our results offer a more complete picture of how speculation and attendance on notifications are related to the alerts' characteristics (e.g., frequency of appearance, intuitiveness of notifications) keyword settings, and the factors that may affect users' perceived disturbance. Second, we identified the moments that users don't need the help of alerts and would know the source of the arriving notifications; these moments are related to a recently used app and a recently contacted sender. Finally, this study provides recommendations on how to set keywords more efficiently and effectively. To summarize, an alert that brings user-defined information is worth supporting in future notification systems due to its effect on speculation and selective attendance, and it is hoped that this paper's design recommendations will assist smartphone users to speculate about notifications more accurately, and to make better decisions about whether to attend to them or not.

REFERENCES

[1] Samaneh Aminikhanghahi, Ramin Fallahzadeh, Matthew Sawyer, Diane Cook, and Lawrence Holder. 2017. Thyme: Improving Smartphone Prompt Timing Through Activity Awareness. https://doi.org/10.1109/ICMLA.2017.0-141

[2] Jan Berg and Johnny Wingstedt. 2005. Relations between Selected Musical Parameters and Expressed Emotions: Extending the Potential of Computer Entertainment. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology* (Valencia, Spain) *(ACE '05)*. Association for Computing Machinery, New York, NY, USA, 164–171. https://doi.org/10.1145/1178477.1178499

[3] Richard Boyatzis. 1998. Transforming Qualitative Information: Thematic Analysis and Code Development. (01 1998).

[4] Stephen A. Brewster, Peter C. Wright, and Alistair D. N. Edwards. 1993. An Evaluation of Earcons for Use in Auditory Human-Computer Interfaces. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (Amsterdam, The Netherlands) *(CHI '93)*. Association for Computing Machinery, New York, NY, USA, 222–227. https://doi.org/10.1145/169059.169179

[5] Yung-Ju Chang, Yi-Ju Chung, and Yi-Hao Shih. 2019. I Think It's Her: Investigating Smartphone Users' Speculation about Phone Notifications and Its Influence on Attendance. 1–13. https://doi.org/10.1145/3338286.3340125

[6] Yung-Ju Chang, Yi-Ju Chung, Yi-Hao Shih, Hsiu-Chi Chang, and Tzu-Hao Lin. 2017. What Do Smartphone Users Do When They Sense Phone Notifications?. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers* (Maui, Hawaii) *(UbiComp '17)*. Association for Computing Machinery, New York, NY, USA, 904–909. https://doi.org/10.1145/3123024.3124557

[7] Tilman Dingler and Martin Pielot. 2015. I'll Be There for You: Quantifying Attentiveness towards Mobile Messaging. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Copenhagen, Denmark) *(MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 1–5. https://doi.org/10.1145/2785830.2785840

[8] Florian Fessel, Kai Epstude, and Neal J. Roese. 2009. Hindsight bias redefined: It's about time. *Organizational Behavior and Human Decision Processes* 110, 1 (9 2009), 56–64. https://doi.org/10.1016/j.obhdp.2009.07.001

[9] Joel E. Fischer, Nick Yee, Victoria Bellotti, Nathan Good, Steve Benford, and Chris Greenhalgh. 2010. Effects of Content and Time of Delivery on Receptivity to Mobile Interruptions. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services* (Lisbon, Portugal) *(MobileHCI '10)*. Association for Computing Machinery, New York, NY, USA, 103–112. https://doi.org/10.1145/1851600.1851620

[10] Claudio Forlivesi, Utku Günay Acer, Marc van den Broeck, and Fahim Kawsar. 2018. Mindful Interruptions: A Lightweight System for Managing Interruptibility on Wearables. In *Proceedings of the 4th ACM Workshop on Wearable Systems and Applications* (Munich, Germany) *(WearSys '18)*. Association for Computing Machinery, New York, NY, USA, 27–32. https://doi.org/10.1145/3211960.3211974

[11] Jose A. Gallud and Ricardo Tesoriero. 2015. 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* (Copenhagen, Denmark) *(MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 819–824. https://doi.org/10.1145/2786567.2793706

[12] Nitesh Goyal and Susan R. Fussell. 2017. Intelligent Interruption Management Using Electro Dermal Activity Based Physiological Sensor for Collaborative Sensemaking. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 52 (sep 2017), 21 pages. https://doi.org/10.1145/3130917

[13] Scott A. Hawkins and Reid Hastie. 1990. Hindsight: Biased judgments of past events after the outcomes are known.

[14] Kari Kallinen. 2003. Emotional Responses to Single-Voice Melodies: Implications for Mobile Ringtones.

[15] Krippendorff Klaus. 1980. Content analysis: An introduction to its methodology.

[16] Andreas Komninos, Elton Frengkou, and John Garofalakis. 2018. *Predicting User Responsiveness to Smartphone Notifications for Edge Computing: 14th European Conference, AmI 2018, Larnaca, Cyprus, November 12-14, 2018, Proceedings*. 3–19. https://doi.org/10.1007/978-3-030-03062-9_1

[17] Hao-Ping Lee, Kuan-Yin Chen, Chih-Heng Lin, Chia-Yu Chen, Yu-Lin Chung, Chien-Ru Sun, and Yung-Ju Chang. 2019. Does Who Matter? Studying the Impact of Relationship Characteristics on Receptivity to Mobile IM Messages. https://doi.org/10.1145/3290605.3300756

[18] Richard Ling and Birgitte Yttri. 2002. *Hyper-Coordination via Mobile Phones in Norway*. Cambridge University Press, USA, 139–169.

[19] Abhinav Mehrotra, Robert Hendley, and Mirco Musolesi. 2016. PrefMiner: Mining User's Preferences for Intelligent Mobile Notification Management. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) *(UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 1223–1234. https://doi.org/10.1145/2971648.2971747

[20] Abhinav Mehrotra, Mirco Musolesi, Robert Hendley, and Veljko Pejovic. 2015. Designing Content-Driven Intelligent Notification Mechanisms for Mobile Applications. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 813–824. https://doi.org/10.1145/2750858.2807544

[21] Abhinav Mehrotra, Veljko Pejovic, Jo Vermeulen, Robert Hendley, and Mirco Musolesi. 2016. My Phone and Me: Understanding People's Receptivity to Mobile Notifications. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1021–1032. https://doi.org/10.1145/2858036.2858566

[22] Abhinav Mehrotra, Fani Tsapeli, Robert Hendley, and Mirco Musolesi. 2017. MyTraces: Investigating Correlation and Causation between Users' Emotional States and Mobile Phone Interaction. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 83 (sep 2017), 21 pages. https://doi.org/10.1145/3130948

[23] Chunjong Park, Junsung Lim, Juho Kim, Sung-Ju Lee, and Dongman Lee. 2017. Don't Bother Me. I'm Socializing! A Breakpoint-Based Smartphone Notification System. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 541–554. https://doi.org/10.1145/2998181.2998189

[24] Veljko Pejovic and Mirco Musolesi. 2014. InterruptMe: Designing Intelligent Prompting Mechanisms for Pervasive Applications. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Seattle, Washington) *(UbiComp '14)*. Association for Computing Machinery, New York, NY, USA, 897–908. https://doi.org/10.1145/2632048.2632062

[25] Martin Pielot, Karen Church, and Rodrigo de Oliveira. 2014. An In-Situ Study of Mobile Phone Notifications. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices Services* (Toronto, ON, Canada) *(MobileHCI '14)*. Association for Computing Machinery, New York, NY, USA, 233–242. https://doi.org/10.1145/2628363.2628364

[26] 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 SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3319–3328. https://doi.org/10.1145/2556288.2556973

[27] Martin Pielot and Luz Rello. 2015. The Do Not Disturb Challenge -A Day Without Notifications. https://doi.org/10.1145/270613.2732704

[28] Martin Pielot and Luz Rello. 2016. Productive, Anxious, Lonely - 24 Hours Without Push Notifications. (12 2016).

[29] Martin Pielot, Amalia Vradi, and Souneil Park. 2018. 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* (Barcelona, Spain) *(MobileHCI '18)*. Association for Computing Machinery, New York, NY, USA, Article 3, 11 pages. https://doi.org/10.1145/3229434.3229445

[30] Benjamin Poppinga, Wilko Heuten, and Susanne Boll. 2014. Sensor-Based Identification of Opportune Moments for Triggering Notifications. *IEEE Pervasive Computing* 13, 1 (Jan. 2014), 22–29. https://doi.org/10.1109/MPRV.2014.15

[31] Stephanie Rosenthal, Anind K Dey, and Manuela Veloso. 2011. Using decision-theoretic experience sampling to build personalized mobile phone interruption models. In *International Conference on Pervasive Computing*. Springer, 170–187.

[32] 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* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3055–3064. https://doi.org/10.1145/2556288.2557189

[33] Bahador Saket, Chrisnawan Prasojo, Yongfeng Huang, and Shengdong Zhao. 2013. Designing an Effective Vibration-Based Notification Interface for Mobile Phones. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (San Antonio, Texas, USA) *(CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 149–1504. https://doi.org/10.1145/2441776.2441946

[34] Hillol Sarker, Moushumi Sharmin, Amin Ahsan Ali, Md. Mahbubur Rahman, Rummana Bari, Syed Monowar Hossain, and Santosh Kumar. 2014. 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* (Seattle, Washington) *(UbiComp '14)*. Association for Computing Machinery, New York, NY, USA, 909–920. https://doi.org/10.1145/2632048.2636082

[35] Florian Schulze and Georg Groh. 2014. Studying How Character of Conversation Affects Personal Receptivity to Mobile Notifications. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI EA '14)*. Association for Computing Machinery, New York, NY, USA, 1729–1734. https://doi.org/10.1145/2559206.2581320

[36] Florian Schulze and Georg Groh. 2016. Conversational Context Helps Improve Mobile Notification Management. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 518–528. https://doi.org/10.1145/2935334.2935347

[37] Yi-Hao Shih, Tang-Jie Chang, Jian-Hua Jiang Chen, Hao-Ping Lee, and Yung-Ju Chang. 2020. A Preliminary Study of an Intelligent System Facilitating Selective Notification Attendance on Smartphones via Alert Assistance. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers* (Virtual Event, Mexico) *(UbiComp-ISWC '20)*. Association for Computing Machinery, New York, NY, USA, 468–473. https://doi.org/10.1145/3410530.3414432

[38] Alireza Sahami Shirazi and Niels Henze. 2015. Assessment of Notifications on Smartwatches. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (Copenhagen, Denmark) *(MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 1111–1116. https://doi.org/10.1145/2786567.2794338

[39] Sichao Song and Seiji Yamada. 2017. Expressing Emotions through Color, Sound, and Vibration with an Appearance-Constrained Social Robot. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (Vienna, Austria) *(HRI '17)*. Association for Computing Machinery, New York, NY, USA, 2–11. https://doi.org/10.1145/2909824.3020239

[40] Liam D. Turner, Stuart M. Allen, and Roger M. Whitaker. 2015. Push or Delay? Decomposing Smartphone Notification Response Behaviour. In *Proceedings of the 6th International Workshop on Human Behavior Understanding - Volume 9277*. Springer-Verlag, Berlin, Heidelberg, 69–83. https://doi.org/10.1007/978-3-319-24195-1_6

[41] Fengpeng Yuan, Xianyi Gao, and Janne Lindqvist. 2017. How Busy Are You? Predicting the Interruptibility Intensity of Mobile Users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 5346–5360. https://doi.org/10.1145/3025453.3025946