Journal of Ambient Intelligence and Smart Environments 11 (2019) 277–297
DOI 10.3233/AIS-190525
IOS Press

Assessing the perceptibility of smartphone notifications in smart lighting spaces

Andreas Komninos [*], Jeries Besharat, Vassilios Stefanis, Georgia Gogoulou and John Garofalakis
Computer Engineering & Informatics Department, University of Patras, Greece

Abstract. In smart spaces with connected smart lighting, there is an opportunity to deliver smartphone notifications using peripheral light, along with using standard smartphone modalities such as sound, vibration and LEDs, in order to help a user perceive them without constantly monitoring their mobile device. In this paper, we examine the effectiveness of on-device and extra-device modalities through smart lighting. We address a gap in literature by establishing a foundation that explains the role of modalities with which a notification is delivered on a mobile device. For this purpose, we conducted two ecologically valid and carefully designed experiments in a controlled environment that simulates multitasking in a smart home environment, and demonstrate that modality preferences are dependent on the environment context, by analysing subjective user data through a machine learning approach. We derive a set of guidelines for choosing notification modalities and set future research directions.

Keywords: Mobile notifications, ambient lighting notifications, notification modalities, smart homes, smart environments

1. Introduction

Notifications on mobile devices are generated by all kinds of applications and services running on our smartphones. Users have been found to receive on average more than 60 mobile notifications daily [31], a figure that remains relatively steady throughout the years, as a recent study reports [33]. Notifications are typically tended to within a few minutes, with time taken to dismiss these depending on various context factors [23,31,39]. A considerable body of literature deals with the identification of opportune moments in which to notify users of events occurring on their mobile device (e.g. [35]), and the typical behaviour with mobile notifications has been studied recently in a number of key papers (e.g. [40]). An important shortcoming in the available body of literature is lack of research into the perceptibility of mobile notifications with regard to the modality (or combination thereof) with which it is being delivered. Some previous literature attempts to address this gap (e.g. [10,20,24]), though the data it reports on comes from field studies, where a number of factors that may affect the reliability of results are not controlled for, reducing from the internal validity of these studies. Addressing the internal validity issue forms the primary goal of this paper. As a secondary goal, this paper aims to add to literature by comparing traditional modalities of delivering mobile notifications on the device, with the concept of extra-device notifications, in the form of ambient smart lighting. The synchronization of mobile notifications across multiple devices has been considered or studied in the past (e.g. [39,45]). With the proliferation of domestic connected IoT systems (e.g. Philips Hue, Apple Homekit, GE Link), questions about the integration of domestic appliances in the user's workflow of managing notifications become not just a scientific curiosity, but take practical significance. Connected lighting notifications also extend into the needs of users with disabilities (e.g. hard of hearing) or situations where other modalities are socially inappropriate.

---

[*] Corresponding author. E-mail: akomninos@ceid.upatras.gr.

2. Related work

2.1. Notifications on mobile devices

Modern smartphones generate many notifications daily [31], relating to multiple types of events (14 application categories are described in [40]). These notifications are typically dealt with in a short time frame measured in minutes, depending on which application generated it, its perceived importance to the current task, the social relationship between the user and a person relating to the notification, the hour and day of the week, the device that generated them, the user's personality and the current task a user is engaged in [23,27,35,40]. Thus, not all types of notification are important to the users under a given context. It also becomes apparent that the time taken to dismiss a notification does not only depend on the perceptibility of the notification. We can therefore frame the research relating to interaction with notification into three main topics: (a) how can the importance of a notification be assessed under context; (b) how to determine opportune moments to deliver notifications and thus reduce the impact of interruptions; and (c) what modalities are best to use under the given context, in order to efficiently deliver a notification, giving it a high chance to be perceptible while causing minimal disruption. Existing literature has provided useful insights in answering questions (a) and (b) by extrapolating the notification generation context and user availability through a range of techniques that involve device sensors and analysis of the notification content and generation parameters, to deliver the right notification at the most opportune time [23,27–29,35]. In this paper, we are concerned with determining appropriate notification modality, a topic which remains largely unaddressed in literature.

2.2. Notifications on mobile devices

Typical smartphone notification modalities are visual (including notification icons and the device status LED), auditory (including speech and sound) and haptic (vibration), though sometimes vibration can have unintended audio effects as well (e.g. when a device is vibrating against a hard surface). Users are able to control all three modalities on a modern smartphone like Android, though it is most typical that the users will switch between sound (on/off) and vibration (on/off) ringer mode combinations during the day [5]. Some users may totally silence their devices, however studies such as [24,31] show that silent mode does not prevent users from becoming aware of the notification events within a reasonable timeframe (this is explained in [5] as the users enter a proactive monitoring state). Users are up to 12 times more likely to immediately attend to a notification if it is delivered with at least one modality [20].

The main distinction between modalities relates to their persistence. Audio tones (with the exception of phone calls) and haptic modalities are momentary, meaning that if the users are unable to respond to the notification immediately, they may never become aware of it, or forget about it [12]. Visual modalities such as the screen display (status bar & lock-screen icons) or device status LEDs persist until the notification is dismissed, which helps participants in deciding to react later [12]. When trying to apply the concept of persistence in haptic feedback, researchers in [14] created a device which would constantly vibrate and whose vibrations would increase in frequency and intensity, according to the number of pending notifications active on the device. The concept was found annoying but users were still able to distinguish between the constant "idle" pulse and the more intense vibrations associated with notifications. Haptic feedback as a "peripheral display" is discussed as a prime candidate for implementing calm technology in [3], where it is acknowledged that depending on its design, it may suffer from issues such as information loss, or lack of ability by the user to arbitrarily "tune into" the signal.

A further distinction between modalities can be made by thinking of these as private or public, in terms of who can perceive them. An audio notification is public, while a haptic or visual notification is often considered private, though this is not always the case: a blinking device LED can be visible to all who can see the device even from a distance and sometimes a vibration can be heard or felt by others too, e.g. if the device is on a table. The device LED affords users some awareness of which application the notification is coming from through its colour, or importance of the notification depending on the blinking rate [20]. However, device LEDs have the disadvantage of being small and not overly bright, and not all devices incorporate colour RGB LEDs (if at all). In [11] an attempt was made to discover whether audio modalities could afford the same types of awareness to users and the researchers found that, for distinguishing between application categories, speech was the best performing modality, followed by auditory icons and lastly earcons (typically used in mobile notifications in modern smartphones). A less explored modality for notifications is the olfactory sense [8], which are less reliable, but also perceived as less disruptive and pleasant as a modality. The reseranchers recommended their use not as a replacement, but as an amplifier for existing modalities. Finally, a further consideration in mobile notifications is that they might be delivered simultaneously across multiple devices (e.g. a user having installed an instant messaging service client on their smartphone, tablet, computer and smartwatch [46]. Currently there is a profound lack of research on how to manage multi-device notifications.

The impact of modality on the perceptibility of notifications has not been widely studied. In [10], the majority of users (65%) were shown to prefer a combination of modes that includes sound. The studies that we review next, report conflicting results. All of these report findings from field trials and are based on the current device ringer mode. In [20] it was impossible to discern any statistically significant difference in notification reaction time when comparing across modalities and their combinations, even when asking participants to manually rank their preferred choice of modality (though it was found that users like to associate vibration and sound to important notifications, and that social context plays a role in determining modality choices). In [24], reaction time was found to be lowest with vibrate-only mode, followed by sound-only and sound-vibrate. In [31], reaction time was found to be faster when the phone was on vibrate-only mode, followed by silent mode and normal mode (between which there was no difference). A more recent study [42] showed that the device ringer mode is correlated to attentiveness towards mobile notifications.

However, there are two main problems with these studies: Firstly, field studies suffer from inherent internal validity problems which are very pronounced in this case. These studies did not control for a number of everyday behaviors, which might have affected the noticeability (or reaction times) of notifications significantly. For example, if a user left their device in a jacket pocket, or on a desk near other clutter, or in another room e.g. to charge, as reported in [5], then obviously the measurements would be affected. In [9], it is shown in a controlled experiment that the position of the device (backpack, table, trouser pocket) has an effect on the perceptibility of notifications, regardless of the modality used to deliver them. In this study, and for each device position, there was no noticeable difference in the perceptibility of notifications using sound or vibration. As other literature indicates [23,27,35,40], the current task and social context of the user can strongly affect the measured response times. Further from this, there are numerous issues that may impact in-the-wild notification studies, for example, some notifications are persistent (i.e. not user-dismissable) [33], and others are issued by the operating system (as implemented by individual manufacturers) and immediately dismissed as a means of intra-app communication [18]. Previous studies reporting modality impact on response time did not consider addressing these data issues, with the exception of [18]. A clearer demonstration of the internal validity issues from the aforementioned studies comes from [32], where it was discovered that ringer mode (unknown, silent, vibration or sound) is a weak predictor of the attentiveness of a user towards their mobile device with the purpose of noticing a message notification, while other indicators that are not pertinent to the notification itself (e.g. time elapsed since last “screen on” event or “hour of the day”) are stronger predictors for attentiveness. It’s not clear in this study whether sound and vibration modalities were considered separately or in conjunction with one another, but it highlights that the generalizability of findings from previous field studies is weak, owing possibly to the lack of internal validity.

The second major internal validity issue with these field studies is that they are based on capturing the user’s device ringer mode. This is problematic because ringer mode may suppress, but does not add beyond the programmed modality requests (thus will not add a LED illumination, vibration or sound to a notification which is not programmed to have one). When a phone is set on “vibrate only”, it doesn’t necessarily mean that every application generating a notification will result in a vibration. To infer thus reliable conclusions on how a notification modality influenced response time, a study should capture all types of information (what were the programmed notification modalities, user per-app preferences and what was the current ringer mode at the time of notification). In a more recent in-the-wild study [18], further proof is provided that device ringer mode is not correlated to notification response time, and that additionally, since this study attempted to determine the actual modality that was used to deliver the notification (depending on ringer mode and programmed modality), only vibration seemed to correlate negatively with response times. This finding is consistent with another recent study (although the focus was aimed at a specific type of notification) [19]. Finally, it should be noted that for the reasons described previously, in-the-wild studies offer observations but not explanations for the role of modality in notification perceptibility, because of the lack of control in these studies. While some observed correlations may make sense (e.g. a user will respond more quickly if the device screen is "ON", possibly because they are already doing something on the device and therefore likely to notice a notification more immediately, other observed correlations do not seem to make intuitive sense [42].

2.3. Notifications with smart lighting systems

The use of simple connected household devices to convey information to users (such as ambient lighting and peripheral displays) is a concept that has been discussed under the principles of Calm Computing since the early days of ubiquitous computing (e.g. [15,16]). With the affordability of connected lighting systems that can interface with smartphones, it is easy to see that a natural synergy for solving the shortcomings of mobile visual notification feedback modalities (notably, the device LED) can be achieved. In fact, a synergy of ambient lighting and the smartphone for notifications satisfies most of the criteria for ambient interface design, set out by Gross [12]. We are not aware of any literature that investigates the use of ambient lighting for the delivery of smartphone notifications as a precise mapping of the state of the mobile device LED. However, some previous work exists on ambient notification systems. In [37], it is proposed that ambient information systems may conform to four main design patterns, one being a "Symbolic Sculptural Display", i.e. a system that displays very few pieces of information, usually a single element. A system thus consisting of a single light bulb that replicates mobile notifications can be considered to fall under this category, but the authors do not propose specific ways for designing the function of such systems, other than that the system must support transitions to prevent "change blindness". A thorough survey of existing "ambient lighting systems" (ALS) that fall into this category can be found in [22]. However, there are gaps in all of the surveyed papers therein: either the systems presented are evaluated in preliminary trials with very few participants (e.g. [2]), or there are, at best, limited comparisons between alternative designs for the perceptibility of conveying simple notifications (e.g. in [30] it is argued that blinking or animated lights should be used, but only blinking vs. static light was actually examined).

In [22], the term ALS is used to describe "a system positioned in the periphery of a person's attention that conveys information using light encodings in a non-distracting way most of the time". The authors propose four general guidelines, one of which states that a light's blinking rate is the most suitable pattern for notification encoding. This guideline is partially supported in [21] via a participatory design process, but the researchers did not experimentally evaluate its effectiveness. Supporting change and state transition in ALSs is demonstrated in [25], where an RGB LED strip reflecting light on the wall behind a computer monitor, gradually changed color from green to red, depending on how much time remained for a user to complete a task with a deadline. This can be seen as a persistent notification system, but has little practical relationship with the majority of spontaneously issued notifications that are typically issued through events in mobile devices and do not contain a temporal dimension. In [38] the smartphone is augmented with additional LEDs able to project light surrounding the device, however this work was presented as a prototype and not evaluated. Other research such as [26,34], extend the modalities of a smartwatch or a tablet using additional LEDs, but, just as in [38], these extensions are on-device and demand the user's attention is already on the device itself. Hence, they do not conform to the definition of an ALS.

ALS notification systems have privacy implications, as highlighted for example in [13], where participants raised significant concerns. So far, only [43] have investigated the issuing of ambient notifications with the user engaged in a social activity and found that the presence of another person in the room did not affect the acceptability of the notification, regardless of intrusiveness of the modality. However, the participants in the study were real-life couples and the close relationships (trust and familiarity between individuals) might have affected this finding. Finally, as far as the positioning of ALSs is concerned, in [1], it was found that the most suitable location for a general use ALS is a living room or office, while specific room types (e.g. the kitchen) should be reserved for special purpose objects. Some guidelines on the design of smart lighting can be found in [6]. The researchers explored a variety of information encodings using colour, brightness and rates of change, for the purpose or remotely monitoring an elderly person's activity state. From these guidelines the general findings applicable to other scenarios were that brightness might be less appropriate than colour coding, but if used, sudden intensity changes should be avoided. Changes should correspond to a user's mental model of what is being represented and changes should be intuitive, to create a dynamic understanding of the context. Colour should be chosen carefully as it evokes emotion, and lighting should avoid over-illumination (eye irritation) and pay attention to positioning (not directly in the user's field of view, but in the periphery). We can conclude thus that the guidelines for designing ALS notification services are still not definitive but there seems to be some evidence that perceptibility and interpretation can benefit from gradual state transitions and appropriate use of color or blinking patterns.

2.4. Summary and research questions

Summarizing the previous literature, we can derive the following open issues. Firstly, although a range of studies report findings on the impact of modality on the perceptibility of smartphone notifications, the generalizability is limited because they did not control for extrinsic contextual factors. Additionally, these studies focus on ringer mode, which does not necessarily reflect the true modality of a notification. Secondly, research on notification-based ALSs focuses on the use of these systems as stand-alone replacements and not extensions of a smartphone. Where lighting has been investigated as an extension to the modalities available on a device, this has been done by augmenting the device itself, hence negating the definition of an ALS by placing the lighting at the center of the user's attention and not their periphery. There is hence no present understanding of how an ALS can extend the modalities for issuing notifications on a smartphone. Based on the above, our main research questions are

(R1) "*How do modality combinations affect the perceptibility of smartphone notifications in a typical use environment?*";

(R2) "*How does the extension of mobile notifications to ambient lighting affect the perceptibility of smartphone notifications, alone or in conjunction with existing smartphone modalities, in a typical use environment?*", and

(R3) "*Do variations in ALS modality settings affect its perceptibility?*".

In a previous artlcle [17], we explored the answers to questions R1 and R2. This investigation is presented again here, since this is paper is an extended version of our previous work. In addition to this previous work, we present also further experimentation to answer research question R3. Moreover, since R1-3 concern the behaviour of users in a specific situational context, we present additional work to explore a further research question that relates to the attitude of users towards modality preferences under different contextual situations, hence formulate the question:

(R4) "*How does change in situational context affect user attitude towards notification modality preference?*"

3. Experiment 1: Multimodal notifications

3.1. Experiment setup and participants

To answer our research questions R1 and R2, we decided to proceed with a laboratory experiment, whose controlled conditions would complement existing studies by focusing on adequate ecological validity and maintaining strong internal validity. As such, our aim was to examine user behaviour in perceiving notifications in a controlled but realistic setting.

Choice of environment Mobiles are used in a variety of environments and settings. For the purposes of this experiment we chose to simulate an environment of a home office, where the user might be engaged in multiple simultaneous tasks, hence not constantly paying attention to their mobile device. A study of 693 participants [50] demonstrates that the most probable place for a user's mobile to be found at any time, regardless of environment, is "out on the table or desk" ("right now": 68%, "over 24 hrs": 83%), followed by the front trouser pocket ("right now": 14%, "over 24 hrs": 64%). Hence, though smartphones are highly mobile, in reality they are mostly stationarily placed on flat surfaces near the user. Typical environments that contain such flat surfaces are home or office environments, which are also the natural "habitats" for an ALS [1]. Thus, the home-office set-up reflects a highly ecologically valid scenario. We considered also a set-up where we might position the device in second most common location, i.e. the front trouser pocket of users. However, this would reduce the audibility of sound notifications, depending on the pocket lining material of our users' clothing and the possibility of some female users also wearing thick tights (the experiment was during the winter), thus creating uneven conditions. Additionally, this set-up would remove the visibility of the LED notification, preventing us from examining it. Furthermore, in [9], it is shown that the LED is not percepti-

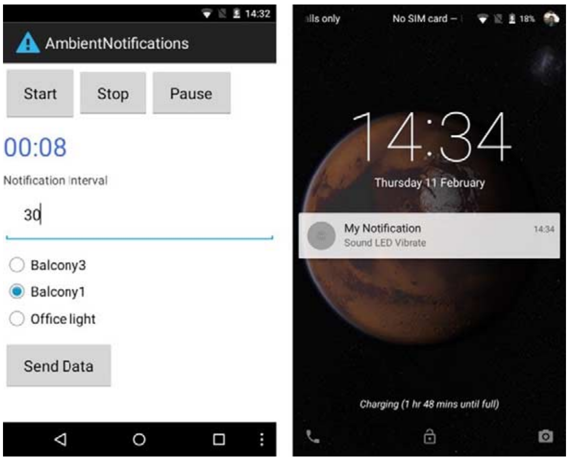
Fig. 1. Participant workstation. A notification is being displayed via device LED and lighting behind the monitor.

ble in any device location other than a table (i.e. backpack, trouser pocket), therefore positioning the device on a table is the only way to obtain a comparison with this modality. Though there exists no previous literature on where exactly users place their devices on desk, in [7,50] it is shown that users mostly keep their device within arm's length for easy reach and maximum perceptibility of notifications. We thus selected a smartphone position as per Fig. 1 so as plausibly emulate a user's behaviour (i.e. the device being within easy reach, facing up and within our participants' field of vision, so that the device LED at the bottom of the device can be easily seen). Some smartphone screens wake up from sleep when a notification is received, but we excluded this option. Finally, our set-up included an ALS using a single bulb, placed directly behind the participant's monitor, positioned close to the wall (Fig. 1). Hence the participant could not directly see the bulb, but was aware of its state as light reflected on the wall and desk surface behind the monitor, as per the recommendations in [6].

Activities during experiment Using multiple activities to overload the user's cognitive processing ability in multiple channels was intentional and essential to our experimental design, as it would help prevent the participants' intentional focus on the mobile device and remove their ability to direct their attentiveness to it, something that would bias their behaviour [5]. According to Wickens' Multiple Resource Theory [48,49], a human operator has several pools of resources that can be simultaneously tapped by the same, or multiple concurring tasks. These resources include the user's sensory organs and the cognitive processing of stimuli (perception, processing, action, and reasoning). When multiple tasks occur, a competition for these resources emerges, whereas when a single task is taking place, these resources can be tapped in parallel, enhancing user performance (e.g. if a user misses a notification sound, they can still perceive it through visual feedback, hence successfuly perceiving it). Therefore, we aimed to experiment in an environment of multiple tasks that "overload" the users' sensory channels while performing a cognitive process that employs each channel. For example, not just having an audio stream, but having the user actually do something with the incoming sound.

It is possible to devise and select from many examples of such tasks, as long as the choice of task is a representative instance of the attention overload abstraction. Hence, we selected to engage the participants in two representative parallel tasks, which would overload their vision and hearing channels. The first task was to play a game on a desktop computer screen – for this we selected the well-known Bubble Shooter game, which is very simple to learn and known for its addictiveness. We asked participants to engage in the game, without worrying about high-scores or losing (if they lost, they could start over). At the same time, we played a recording of a basketball game through speakers in the room, at an average volume level of 50 db, roughly equivalent to the volume of a conversation. Participants were asked to pay attention to the game and take a note every time a particular well-known player's name was mentioned. We developed an application running on a smartphone, which was able to generate notifications using all device modalities including an ALS. Participants were asked to make dismissing the notifications on the smartphone as they noticed them their top priority, and to position the device back to its marked position on the table, as shown in Fig. 1, after dismissing a notification. Notifications were dismissed from the device's lock screen (Fig. 2 right) by swiping on the notification. If a participant did not dismiss a notification before another was issued, both notifications persisted on the screen and the participant was asked to swipe both.

Software, devices and settings For this experiment, we built a simple Android application that generates notifications to the user's mobile (Fig. 2 left). These notifications are issued with all the possible combinations including the modalities of sound, device LED, vibration and ambient lighting, with the exception of



00:08

Notification interval

30

○ Balcony3
● Balcony1
○ Office light

14:34

Thursday 11 February

My Notification
Sound LED Vibrate
14:34

Charging (1 hr 48 mins until full)

Fig. 2. The notification generator application control interface (left) and locked device screen with an issued notification (right).

Table 1

Notification modality combinations (S: sound, L: LED, V: vibration, A: ambient lighting)

Modality combination	Sound	Device LED	Vibration	ALS
A				*
V			*	
VA			*	*
L		*		
LA		*		*
LV		*	*	
LVA		*	*	*
S	*			
SA	*			*
SV	*		*	
SVA	*		*	*
SL	*	*		
SLA	*	*		*
SLV	*	*	*	
SLVA	*	*	*	*

issuing a notification with none of the above modalities (as the user would have no way of perceiving it). This resulted in 15 different notification modality combinations (Table 1, modality presence in a combination denoted by a *). The application issues each modality combination twice in the experiment session, resulting in 30 total notifications to the user. Notifications are issued at a regular interval of 30 seconds and the sequence of notifications (with regard to the modality combination) is wholly random. We did not tell our participants that notifications would arrive at regular intervals. This was a restriction due to the task of taking note of the basketball player name mentions, and, as will be explained below, performance in this task could only be equally measured if the task was executed in an equal time period by all participants. We anticipated that because of the random selection of notification modality and the high cognitive load placed by the parallel tasks, participants would not be able to learn and anticipate the notification timing. We will demonstrate in Section 3.3.1 that this approach did not lead to learning effects, hence our results are not invalidated by this aspect of the experiment design. For our experiment, we used a Google Nexus 5 device. For notification sound we selected one of the in-built Android tones ("Tethys"), a simple two-tone short earcon, with the volume level set to 90% of the maximum supported by the device. This was the highest volume setting without the speaker producing "tinny" noises, which could alter participants' perceptions of annoyance of the modality. The device LED color was set to white, blinking at a repeating pattern of 3 seconds ON, 3 seconds OFF. The vibration was set to a pattern of 200 ms ON, 200 ms OFF, 200 ms ON (i.e. two vibrations for each notification). These patterns are similar to popular applications where notifications are an integral part of the user experience (SMS, Gmail, Facebook messenger). In this experiment, we are not concerned with the encoding the importance of a notification but only with the perceptibility of the notification event, hence we chose a white LED color and patterns that are arguably of average intensity so as not to convey priority semantics.

Finally, issuing notifications vial an ALS, we extended our application to work with Philips HUE, using a single A19 9W bulb, setting the colour of the bulb to white and brightness of 70% (empirically set with 6 colleagues, 3 female, for a comfortable and perceptible level) and with the bulb remaining ON until the user dismisses the notification. Because we are not concerned with the encoding of information in a notification, we used a neutral light colour (white) and did not employ blinking, so as to prevent any assumptions regarding the importance of notifications [6]. Based on [6,30], we implemented a gradual transition between the "OFF" and "ON" states, to animate the light and attract attention. Finally, the notification text was generic, to exclude any priority semantics (Fig. 2).

Participants We recruited 25 participants, 7 females. All participants were Computer Science students, and in the age bracket of 18–29 (the questionnaire presented a list of ages brackets and did not require the precise age). Most participants were Android device users (20) but we had also 2 iOS and 3 Windows Mobile users. They were not incentivized for their participation.

3.2. Experiment results

Validation of the task and method Our first concern was to validate the appropriateness of the parallel tasks (i.e. playing the game and listening for the basketball player's name) for occupying the participants' attention. At the start of our experiment, we asked participants to report using 5-point Likert scales both familiarity with the sport (all mentioned often or very often watching or attending basketball games) and the player's name (all reported him to be known, or well known to them). Because the player's name is mentioned a specific number of times during the recording, we needed participants to complete the task in the same timeframe, so as to effectively measure their performance in listening for the player's name. We expected that users would not learn to anticipate the timing of notifications, because of the random selection of notification modalities and the high level of cognitive engagement in the tasks at hand. Indeed, it is common psychology knowledge that human cognitive ability suffers from a processing bottleneck which reduces our ability to perform multiple tasks at the same efficiency as single tasks, so much so that cognitive overload in one or more channels with essential or incidental information, removes our ability to learn new information. Though humans are able to monitor several streams of information for a specified target, if those streams contain a target at the same time, or close together in time, some targets will be unavoidably missed.

Still, we examined the average response times for the starting 10, middle 10 and final 10 notifications issued to each participant (Fig. 3). A Friedman test (due to the distribution of the data), failed to reveal any statistically significant differences in these three subsets ($\chi^2 = 2.880$, $p = 0.237$). Post-hoc Bonferroni corrected pairwise Wilcoxon Signed Rank tests (2-tailed) between the individual sets also did not reveal any statistically significant differences. Since Friedman tests do not allow the computation of effect sizes (Cohen's d), these are presented for the pairwise Wilcoxon signed rank tests (A–B $d = 0.18$, A-C $d = 0.27$, B–C $d = 0.13$). This outcome can mean that either there is no learning effect to be detected, or our study is insufficiently powered to detect one. Common practice where no prior study data is available (as is our case), is to calculate sample sizes according to a general estimate, e.g. Cohen's $d = 0.5$ (a medium size effect), $\alpha = 0.05$ and study power $(1 - \beta) = 0.80$. Based on the above, a power analysis of our study shows that it is sufficiently powered $(1 - \beta = 0.81)$ to detect a one-tailed medium size effect (because we are only interested in whether the response time average decreases due to learning effects), at the lower boundaries of Cohen's "medium effects" category ($d = 0.53$). Thus, our sample size can be considered as appropriate. Given the size of the reported effects, our study is not sufficiently powered to determine statistical significance in the discovered small effects, but we can confirm that if any learning effects were indeed present, these were either very small or small, according to Cohen's d categorisation (the lower thresholds for very small, small and medium effects are 0.1, 0.2 and 0.5 respectively). The long average response time shows that our participants' attention was occupied very heavily by the two parallel tasks. Further corroborating evidence for the heavy cognitive load comes from the fact that out of the total of 19 times that the basketball player's name was mentioned in the 15 minutes of the experiment duration, the average number of instances in which the participants were able to capture the player's name was just 4.36 times (sd = 3.2).

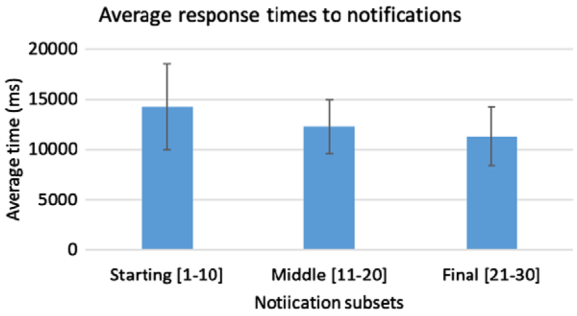


Fig. 3. Perceptibility of notification modalities (error bars at 95% c.i.).

Experiment results The results of our experiment are summarized in Fig. 4. In the figure, the different modality combinations are depicted with the initials A (ambient lighting), S (sound), V (vibration) and L (LED), as per Table 1. The left half depicts all conditions where sound was absent in the notifications, while the right half where sound was present. To examine the results for statistical significance, we used non-parametric tests, due to the non-normal distribution of the data.

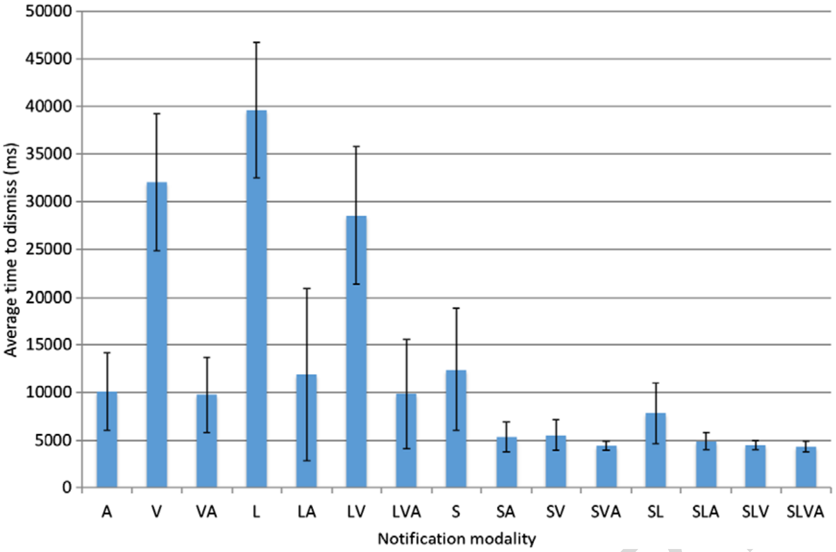


Fig. 4. Notification dismissal time averages (error bars at 95% c.i.).

Table 2

Statistical significance results on pairwise condition comparisons for the effect of modalities in notifications

Ambient lighting	Mod. combo	VA-V	LA-L	LVA-LV	SA-S	SVA-SV	SLA-SL	SLVA-SLV
	Z (p-value)	**−3.740 (0.000)**	**−3.727 (0.000)**	**−3.484 (0.000)**	−2.489 (0.013)	−0.740 (0.459)	−1.332 (0.183)	−1.251 (0.211)
Vibration	Mod. combo	VA-A	LV-L	LVA-LA	SV-S	SVA-SA	SLV-SL	SLVA-SLA
	Z (p-value)	−0.202 (0.840)	−2.085 (0.037)	−0.309 (0.757)	−2.489 (0.013)	−0.309 (0.757)	−0.794 (0.427)	−0.821 (0.412)
Device LED	Mod. Combo	LA-A	LV-V	LVA-VA	SL-S	SLA-SA	SLV-SV	SLVA-SVA
	Z (p-value)	−1.251 (0.211)	−0.040 (0.968)	−0.821 (0.412)	−1.440 (0.150)	−0.094 (0.925)	−0.121 (0.904)	−0.390 (0.696)
Sound	Mod. Combo	SA-A	SV-V	SVA-VA	SL-L	SLA-LA	SLV-LV	SLVA-LVA
	Z (p-value)	−2.139 (0.032)	**−4.130 (0.000)**	−2.516 (0.012)	**−4.319 (0.000)**	−0.444 (0.657)	**−4.103 (0.000)**	−0.498 (0.619)

Firstly, with regard to single modalities (A, L, V or S), a Friedman tests shows that a statistical significance between the mean response times exists ($\chi^2_{(3)} = 33.336$, $p = 0.00$). Following post-hoc Wilcoxon signed rank tests, with a Bonferroni correction setting the significance level at $p < 0.0083$, we find statistically significant differences between V-A ($Z = -3.807$, $p = 0.00$), L-A ($Z = -4.184$, $p = 0.00$), S-V ($Z = -3.538$, $p = 0.00$) and S-L ($Z = -3.700$, $p < 0.00$). The remaining two comparisons do not show any statistical significance (L-V $Z = -1.251$, $p = 0.211$; S-A $Z = 0.256$, $p = 0.798$). From these results we find that as singular modalities, ALS and Sound are equally associated with the fastest response times, therefore being the most perceptible modalities, while the LED and vibration are equally associated with the slowest response times. While it is not surprising that the device LED did not seem to affect the perceptibility of the notifications, we were surprised to find that vibration did not affect the perceptibility of notifications as it did in previous field studies in [20,24,31].

Following on, we examine the effect of adding a single modality to others, therefore creating modality combinations of one or more modalities. A Friedman test across all distributions shows that a statistical difference in the means for each modality combination exists in our data ($\chi^2_{(14)} = 125.892$, $p = 0.00$). To proceed with the analysis, we performed post-hoc Wilcoxon signed rank tests (Table 2) to assess the effect of each modality on combinations with others, applying Bonferroni correction resulting in a significance level set at $p < 0.0071$.

From these analyses, we obtain the following interesting insights. Firstly, adding the LED and Vibration modality to other modalities seems to have no impact on their perceptibility. This is an expected outcome, since our previous analysis showed that these two are the least perceptible modalities. Following on, adding ambient lighting seems to have an effect only on modality combinations that do not include sound, and also adding the sound modality only has an effect on those combinations that do not already include ambient lighting. This confirms our previous analysis which found no statistically significant difference between the perceptibility of these two modalities. Our results indicate that perceptibility of notifications can be strongly affected by the presence of audio or ambient lighting cues only.

Subjective feedback We asked participants to rate, post-experiment, how perceptible and annoying each notification modality was. We also asked them if they would prefer a different intensity level for each of the modalities (e.g. sound level, lamp and LED brightness, vibration intensity) and to note their preferences for these questions on 5-point Likert scales. Participants reported high levels of perceptibility for the sound and ambient lighting modalities, while mostly indicating that vibration and the device LED were hard to perceive (Fig. 5 top). These findings support our quantitative analysis. Understandably, the LED provides a rather small visual cue which is also situated at the periphery of the participants' field of view. Despite being persistent and blinking, the LED brightness and size are not enough to draw participants' attention away from their current tasks. With regard to vibration, it can be theorized here that the low perceptibility of this modality is due to the placement of the device on the table and not in direct contact with the user's body. However, although this explanation sounds plausible, it cannot be unreservedly accepted. There is a significant body of literature documenting the experience of "phantom vibrations", with the evidence pointing towards a mental mechanism of manifestation of false perceptions, linked to the context of use and the level of use of the mobile device (e.g. [41]). It is therefore questionable how much of the perceptibility of a vibration is truly dependent on the sensory capability of humans and not affected by mental processes. In [4], for example, it is shown that the perceptibility of haptic feedback is severely impeded under high cognitive workload. Given the cognitive load imposed on our participants and the short duration of the vibrations (as is standard on mobile devices), it is plausible that vibration perceptibility was low, not because the participants weren't able to sense it, but rather, their brain could not process it due to the multitasking conditions.

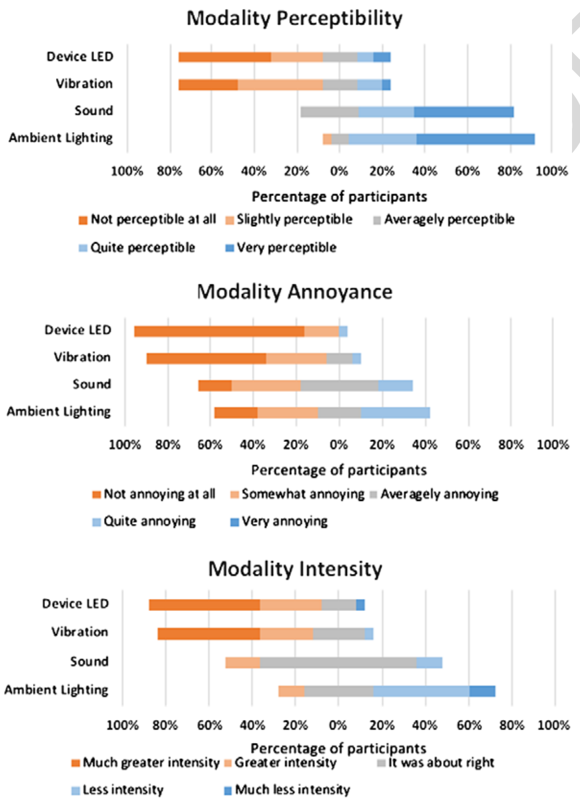


Fig. 5. Response distribution to subjective questions.

Modality annoyance was reported higher for the ambient lighting and sound modalities, an expected result, as these were the two modalities that our participants noticed the most (Fig. 5 middle). These two modalities, particularly ambient lighting, provided a sense of urgency to remove the notification for our participants, as reported in post-experiment discussions. Some participants expressed concern on the impact of the longevity of the bulb or the consumption of energy if the bulb was left on for too long, given the fact that the ambient lighting notification continues to remain switched on until the user dismisses the notification. Inversely, the sound caused a sense of urgency, since it was a non-repeating modality and thus some users felt they should dismiss the notification right away and postpone any ongoing tasks, for fear that they might forget about it if left for later.

Finally, we asked our participants to indicate for each modality, how they might change the intensity with which it was delivered. As can be seen (Fig. 5 bottom), and in line with their previous reports, 20 participants would like to intensify the device LED and 18 the vibration levels. While the latter is not possible on the Android platform, vibration intensity can be changed by prolonging the ON period in the pattern. Extend-

Table 3

Notification modality combinations

Notification setting	Description	Persistence	Lighting pattern
AL1	On and stay on	Yes	Single Off–On transition duration 300 ms
AL2	Blink twice	No	On (500 ms), Off (500 ms), On (500 ms), Off
AL3	Slow pulsing	Yes	Repeated Off–On–Off transitions every 2000 ms (transition duration 1000 ms)
AL4	Rapid pulsing	Yes	Repeated Off–On–Off transitions every 500 ms (transition duration 300 ms)

ing the ON period or repeating the ON-OFF pattern for longer, might also help with the perceptibility of the haptic feedback when under heavy cognitive load, as suggested by [4]. 18 participants felt that the sound level was about right. Participants were somewhat negative regarding the brightness of the ambient lighting, with 14 participants preferring a lower setting.

4. Experiment 2: Ambient lighting notifications

As mentioned in Section 2, notification modalities can be characterized in terms of their persistence. In the previous experiment, we used a permanent setting for the ALS notifications, i.e. switching the light on and keeping it on until the user dismissed the notification. Intrigued by the efficacy of the ALS notifications in terms of making a notification immediately perceptible, we aimed to further explore whether persistence plays a role in this context. Therefore, we repeated the experiment, aiming to investigate different persistence settings, in order to answer research question R3.

4.1. Experiment setup and participants

The setup for our second experiment was identical to the previous one, with the exception that this time, users were only asked to play the "bubble shooter" game, since the only sensory channel which we aimed to overload was the visual channel. One other change that also took place was that, since some of our participants were left-handed, we asked them to place the device close to their dominant hand, in order to afford quick and easy reachability of the device and thus ensure that the response times to notifications were not affected by reachability issues due to handedness.

To ensure that the participants were focused on the assigned task, we motivated participants by telling them that whoever achieved the highest score would be rewarded with a €30 gift card. Participants were explicitly instructed to immediately dismiss notifications as soon as they perceived them, while striving to achieve the best possible score. We investigated the four settings for the ALS notification system, as show in Table 3.

For the experiment, participants were issued 6 instances for each of the ALS notification types, using the same app as in our previous experiment, thus totaling 24 notifications per participant. The notifications were issued by randomly selecting between the notification types. We also slightly modified our experimental application so that the notifications arrived at pseudo-random time intervals, since precise timing was not required (as was in our previous setup). To achieve this, we issued notifications every 30 seconds, ± a "jitter" time of a random value between [0, 8] seconds.

We invited a further set of participants, who had not participated in our previous experiment. In total, 36 participants attended the experiment (19 female). Participants were mostly engineering students at our university. Their age was between 18–29 (34 participants) and 2 were researchers (aged 30–39). Most participants owned an Android device (31), iOS (4) and we had one non-smartphone user.

4.2. Experiment results

In analysing our participants' results, we noticed that five participants were intensely focused on achieving a high-score in the bubble-shooter game, thereby missing, or ignoring, a large percentage of the notifications issued to them. We excluded these participants from the ensuing analysis, therefore reporting on a sample of 31 participants. From this sample, a further 5 participants did not dismiss some of the final notifications until after the experiment finished (i.e. 30 seconds after the last notification was issued), therefore we set the dismissal time for these notifications to be the time of the end of the experiment.

Users' data across the 6 instances of each notification setting were averaged and analysed via boxplots, to identify and exclude any extreme outliers. This process led to the elimination of a further 3 users from the dataset, who displayed outlier values in three or more of the notification settings, therefore the analysis proceeded with 28 users. A Shapiro-Wilk test for distribution and visual examination of Q–Q plots showed that the values in none of the ALS settings were normally distributed ($p < 0.01$), therefore the tests used for statistical significance are non-parametric. In Fig. 6, we observe some differences in the mean response times across the different ALS settings. Notably, slow pulsing light (AL3) shows the longest response time (m = 7960.42 ms, sd = 7397.51 ms), followed by the permanent light (AL1, m = 7194.27 ms). The rapid changes (blinking, AL2 and rapid pulsing, AL4) showed the lowest response times (m = 5264.04 ms, sd = 3090.48 ms and m = 5566.95 ms, sd = 3644.68 ms respectively). A Friedman two-way analysis of variance however, showed that the observed differences are not statistically significant ($\chi^2_{(3)} = 4.629$, $p = 0.201$). Further pairwise examination with post-hoc Bonferroni-corrected Wilcoxon Signed Rank tests, resulting in a p-value threshold of $p = 0.008$, also confirmed that the differences were not statistically significant across any two pairs (Table 4).

In examining the data more closely, we were able to identify some interesting characteristics of participant behaviour during the experiment. We noted that the participants' response times across a timeline were "peppered" with incidents where the response time was unusually large. For example, in Fig. 7, we demonstrate the response time of four randomly picked participants across their experiment participation timeline. These events indicate either a total miss of the notification event (i.e. the participant failed to notice it altogether), or a purposeful postponement of the engagement with the notification (i.e. the participant chose to ignore it for some time, focusing instead on their game performance).

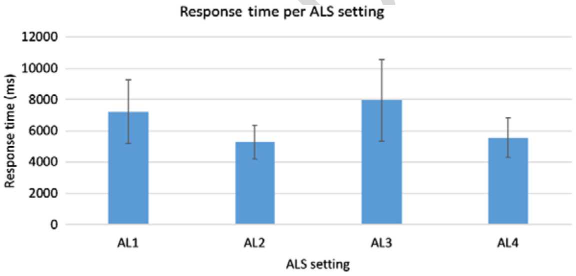


Fig. 6. Average response time per ALS notification setting (error bars at 95% c.i.).

Given these plausible explanations, and since we could not discern which instances were indeed cases of having noticed a notification but choosing to ignore it, we proceeded to prune our dataset to exclude such cases. Analysing the data and identifying the threshold at 80% of the dataset for each ALS setting, we trimmed those instances (records) of response time that fell outside the respective thresholds. This gave us a clearer representation of the true behaviour of participants, namely better insight into how quickly a notification was noticed, and subsequently dismissed according to the instructions given to the participants. Regarding the outlier events, we found 70 such cases (9.4% of all events) across 28 of the participants (min = 1, max = 4, m = 2.5, sd = 1.14), which were similarly distributed across the different ALS settings (AL1:20, AL2:19, AL3:13, AL4:18). Out of the 70 events, only 27 notification events (3.6%) from 12 users exceeded the maximum 38 seconds before another notification was issued, thus certainly being unnoticed (missed) events. This provided some reassurance that no particular ALS setting caused excessive "misses" of the notifications, and that participants behaviour was attributable mostly to choosing to postpone engagement in these cases. As can be seen in Fig. 8, a different picture emerges in the pruned dataset, as the response times between all ALS settings appear similar, except AL3 (which had the fewest extreme response time events). Repeating the same analysis for the pruned dataset, we find that a Friedman two-way analysis of variance, showed that the observed differences are statistically significant ($\chi^2_{(3)} = 15.360$, $p < 0.01$). Further pairwise examination with post-hoc Bonferroni-corrected Wilcoxon Signed Rank tests, resulting in a p-value threshold of $p = 0.008$, confirmed that the differences were statistically significant between AL1-AL3 ($Z = 3.383$, $p = 0.001$), AL2-AL3 ($Z = 3.733$, $p = 0.000$) and AL3-AL4 ($Z = 2.993$, $p = 0.003$), thereby concluding that the slow pulsing light is actually the least noticeable of the four combinations (all others having no statistically significant differences).

Table 4

Wilcoxon signed rank test results (full dataset)

	AL2-AL1	AL3-AL1	AL4-AL1	AL3-AL2	AL4-AL2	AL4-AL3
Z	−1.685a	−0.159b	−0.979a	−2.300b	−0.296a	−1.571a
Asymp. Sig. (2-tailed)	0.092	0.873	0.327	0.021	0.767	0.116

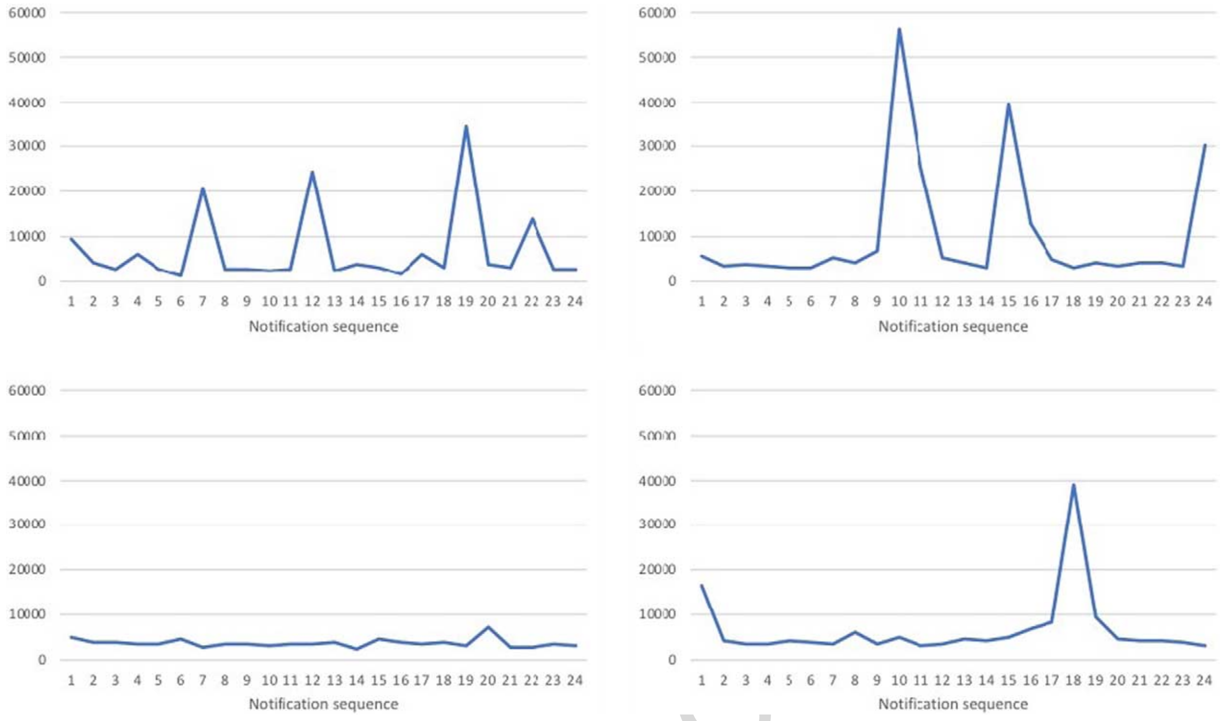


Fig. 7. Random participant response times across a timeline. Episodes of sporadic “forgetufulness” can be clearly identified in three out of four cases (clockwise from top left).

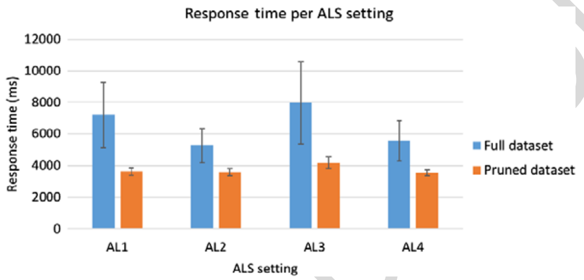


Fig. 8. Average response time per ALS notification setting in the full and pruned dataset (error bars at 95% c.i.).

Subjective feedback In terms of subjective feedback, we asked participants to rate post-experiment the different ALS settings for perceptibility and annoyance, and also asked them to indicate whether they would have preferred a different light intensity setting to achieve a better balance between annoyance and perceptibility during the experiment. For perceptibility, participants believed that the most perceptible setting was the rapid pulsing light (66%), followed by the permanent-on (50%) and blinking (39%), while only 31% found that the slowly pulsing light was perceptible. These findings align perfectly with the observed differences in the response times shown in the results from our pruned dataset. In terms of modality annoyance, most participants agreed that the ALS in general was not annoying, however more participants thought that the rapid pulsing light was annoying or very annoying (31%), than any of the other settings. Finally, most participants (67%) felt that the light intensity level during the experiment was “about right”, while 33% felt that the light intensity should have been greater, or much greater.

5. Subjective notification modality preference under context

As a last step in our work, to answer R4, we present an analysis of an exploratory survey issued to the participants of the two previous experiments. Research in the past has focused on determining the user’s context in order to determine an appropriate time to deliver a notification to them. While the temporal aspect is important, another consideration is whether the user’s current context also affects the modality with which a notification should be delivered. To determine whether such adaptive behaviour may be useful in context-aware notification delivery, we asked users after the

Table 5
Context types used in the survey scenarios

Context type	Context value categories	Context examples
User task	Working, Relaxing	Watching TV, reading books, performing household chores, having a meeting
User social context	Single user, Close relationships, Work relationships	Alone, with family, with friends, with colleagues
User location	Domestic, Workplace	Home, office

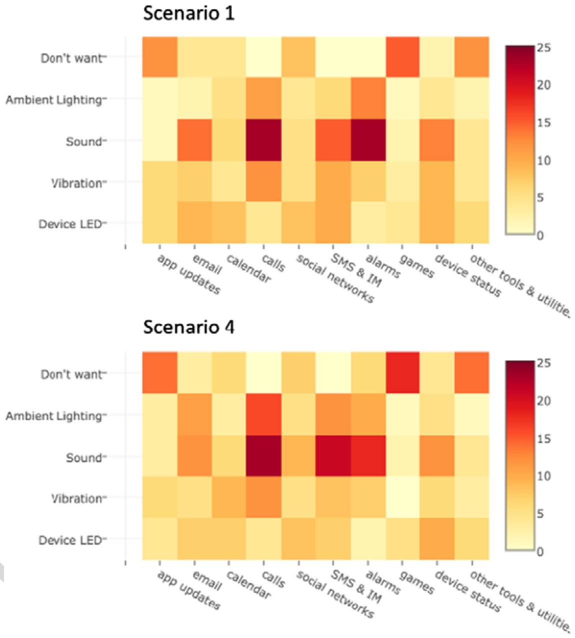


Fig. 9. Sample notification modality preference changes, based on context differentiation (here, the task type: relaxing or working).

end of each experiment, to indicate which modality they would prefer for common smartphone applications, under a range of 5 context scenarios. Each user was asked to report an opinion on each of the modalities they experienced during the preceding experiment. The scenarios were constructed to examine a combination of the user's current task type and their urgency (e.g. work, domestic or leisure) and the user's social context (alone, with friends or family, with coworkers). Table 5 describes the context types used to construct the scenarios.

We constructed thus five context scenarios using combinations of the above, as follows:

1. I am relaxing at home on my own (e.g. watching TV, reading books)
2. I am sat at the office, working, without any colleagues nearby
3. I am at home, in the same space (e.g. living room, dining room) with family or friends
4. I am at home, alone, carrying out domestic chores (e.g. cooking, cleaning, laundry)
5. I am at work, having a meeting with other colleagues in my office

In Fig. 9, we show one example of the emerging differences between the responses, for participants of experiment 1. As can be seen, changing just one context factor (i.e. in this example, the location semantics), yields noticeable differences in the perception of desired, and by extension, appropriate modalities. These results demonstrate that users require that a context-aware notification delivery mechanism considers the nature of the notification (i.e. which application it's coming from), but also the current task and social context, in order to choose an appropriate modality for delivering the notification.

To explore further, we explored these outcomes with two approaches, both from a statistical and a machine learning perspective. For this, we used a multinomial logistic regression, and decision trees respectively, analysing separately the responses of experiment 1 and 2 participants. Both techniques attempt to build models to explain the outcome values of a polytomous variable (in our case, the preferred modality). For this purpose, we transformed the participant responses into a five-dimensional vector, consisting of the scenario attributes:

1. "Task": Working, Alone;
2. "Location": Home, Work;
3. "Social": Alone, Work relations, Close relations;
4. "Application Type": Alarms, App updates, Calendar, Calls, Device status, Email, Games, Other tools & utilities, SMS & IM, Social networks;
5. "Modality": ALS, LED, Sound, Vibration, Do not want

5.1. *Multinominal logistic regression*

This technique attempts to statistically model a polytomous variable based on multiple ordinal, nominal or continuous variables. In our case, the dependent variable was "Modality", since we try to predict the

Table 6

Multinomial logistic regression model classification results

	Experiment 1					
	True LED	True Sound	True Vibration	True ALS	True DW	Class precision
Pred. LED	55	41	44	23	48	12.60%
Pred. Sound	139	315	179	141	46	48.80%
Pred. Vibration	40	5	61	6	16	7.60%
Pred. ALS	0	0	0	0	0	0.00%
Pred. DW	103	28	89	24	277	31.00%
Class recall	16.30%	81.00%	16.40%	0.00%	71.60%	

	Experiment 2					
	True ALS1	True ALS2	True ALS3	True ALS4	True DW	Class precision
Pred. ALS1	14	2	4	12	0	1.90%
Pred. ALS2	60	251	174	108	113	41.80%
Pred. ALS3	7	8	9	5	4	2.00%
Pred. ALS4	63	25	22	67	17	11.50%
Pred. DW	20	102	61	30	512	42.90%
Class recall	8.50%	64.70%	3.30%	30.20%	79.30%	

desired modality depending on the other context attributes. For experiment 1 participants, the final model rejects the null model hypothesis ($\chi^2_{(52)} = 808.042$, $p < 0.01$) and demonstrates a good model fit (Pearson $\chi^2_{(144)} = 77.655$, $p = 1.00$). However, the pseudo R^2 value is 0.399 (Nagelkerke), showing that only part of the variance in declared preferred modality is explainable by the contextual factors Task, Location, Social and Application Type. Looking further, likelihood ratio tests show that three out of these factors have a significant impact on the modality, namely Application Type ($\chi^2_{(36)} = 610.616$, $p < 0.01$), Social ($\chi^2_{(8)} = 100.359$, $p < 0.01$) and Location ($\chi^2_{(4)} = 13.215$, $p = 0.01$), while the user Task seems to have no impact at all ($\chi^2_{(4)} = 5.023$, $p = 0.285$).

The results are slightly different for experiment 2 participants. The final model rejects the null model hypothesis ($\chi^2_{(52)} = 1101.915$, $p < 0.01$) and demonstrates a good model fit (Pearson $\chi^2_{(144)} = 135.300$, $p = 1.00$). The pseudo R^2 value is somewhat better than for experiment 1 participants, at a value of 0.505 (Nagelkerke), explaining more of the variance in declared preferred ALS modality. Likelihood ratio tests show that the factors that have a significant impact on the modality are different this time, namely Application Type ($\chi^2_{(36)} = 950.422$, $p < 0.01$), Social ($\chi^2_{(8)} = 144.654$, $p < 0.01$) and Task ($\chi^2_{(4)} = 11.612$, $p < 0.05$), while the Location context seems to have no impact at all ($\chi^2_{(4)} = 5.200$, $p = 0.267$).

Overall, for Experiment 1 the model recall performance is good for situations where notifications are not wanted at all (DW) or where sound is the preferred modality, but precision in both these metrics is quite low (Table 6). In Experiment 2, again recall performance of the model is reasonably good in situations where notifications are not wanted, or where the ALS2 variation is desirable, but similar to experiment 1, precision is again low.

5.2. *Decision tree classification*

We also employed a basic machine-learning process to determine whether it would be possible to infer context-based rules from the responses received in our survey. For the purposes of inferring rules, we use a decision tree, since the output of the tree is meaningful and easy to interpret as a set of rules. Note here that our target is primarily to observe any emergent component factor hierarchy for choosing a modality under context, rather than to accurately predict a classification, something which was not possible with the regression model. Nevertheless, to observe the modelling ability of the ruleset over our data, we also perform a 10-fold cross-validation on the datasets and report its performance, since the interesting question that can be answered with this technique is whether it is possible to predict a given person's preferences in modalities, based on the preferences of a wider set of users. For replicability purposes, the decision tree parameters

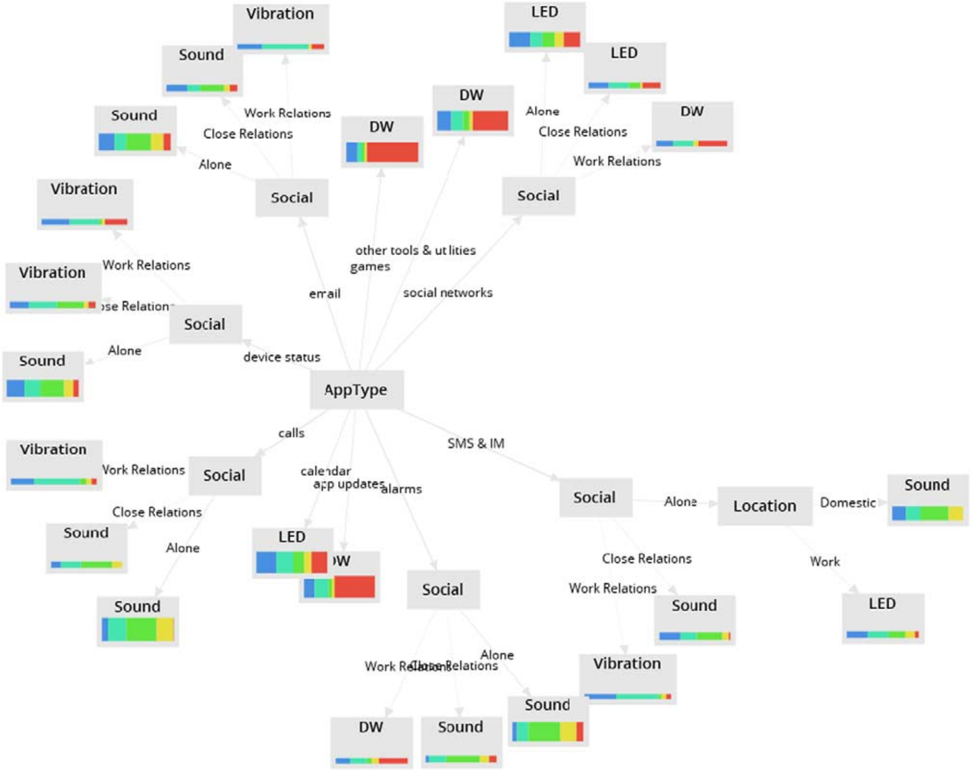


Fig. 10. Radial visualization of the decision trees in experiment 1. The “Application Type” feature is at the center, since it is the root node (DW = “do not want”).

were [criterion: gini-index, maximal depth: none, confidence: 0.25, minimal gain: 0.01, minimal leaf size:2, minimal split size: 4].

As it turns out, for both experiments the decision tree models put the “Application Type” at the root of the tree (Fig. 10 and 11), therefore affording this context feature greater importance over others. Sequentially, this is followed by “Social” context, and “Location” and “Task” does not appear to feature at all in the model for the Experiment 1 dataset, while they feature minimally in the Experiment 2 dataset. Overall accuracy of the models is quite low (Exp.1: m = 39.70%, sd = 2.60%; Exp2. m = 46.86%, sd = 2.75%), however, the recall metrics are actually reasonably good for the “do not want” (shown as “DW”) and “sound” modalities (see confusion matrices in Table 7), as with the regression model previously discussed. Certainly thus there appears to be some consensus across participants about when they do not want to be disturbed, however, the choice of modality in other context situations might possibly be a subject of personal preference of individual users.

6. Discussion of results

Our work has clear implications for the design of notification modalities during multitasking. Firstly, we would like to emphasize the importance of having tested the perceptibility of notification modalities in a controlled environment, compared to previous in-the-wild studies. As literature shows, users’ response times to notifications are dependent on a wide range of factors. Hence previous studies reporting on the correlation of notification modalities with perceptibility are largely unreliable. We were not able to replicate the users’ better response times with vibrations found in other literature (e.g. [20,24,31]), placing doubt on the external validity of these studies. While intuitive explanations of some of our results could be given (e.g. that vibrations were not highly perceptible because devices were placed on the desk), literature such as [4,41] shows that other explanations involving the human cognitive system, not considered in previous studies, may offer better plausibility. The findings and analysis of our quantitative and qualitative data in light of the current literature, lead to some design recom-

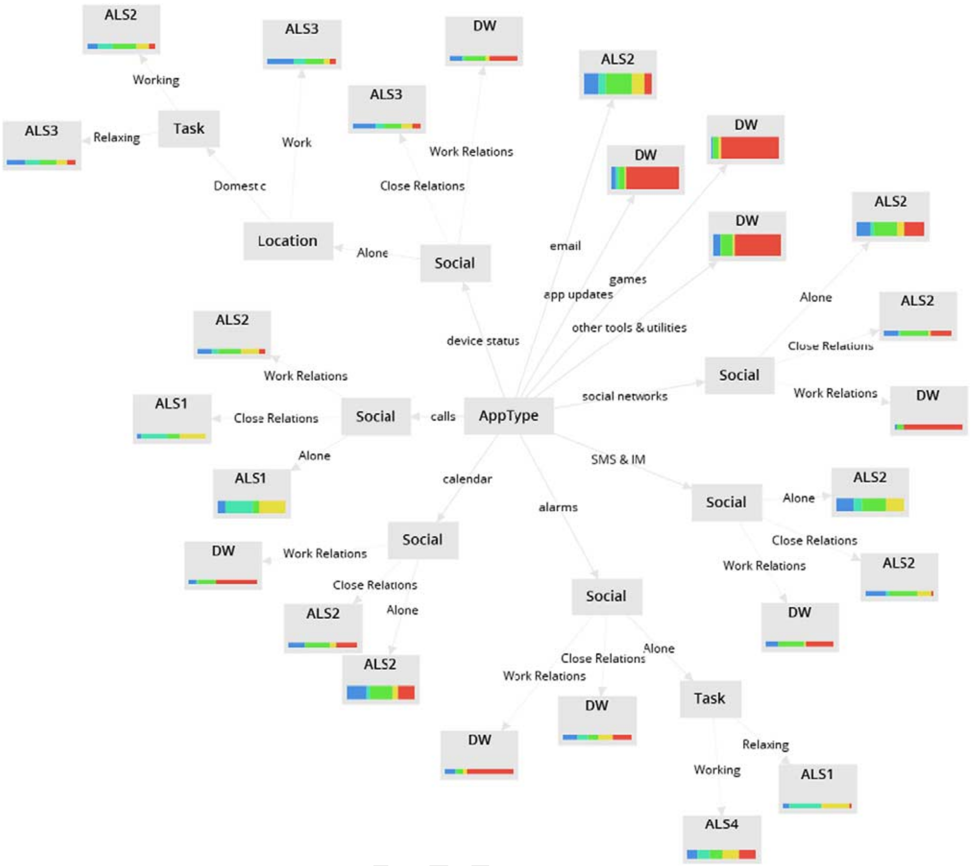


Fig. 11. Radial visualization of the decision trees in experiment 2. The “Application Type” feature is at the center, since it is the root node (DW = “do not want”).

Table 7

Decision tree model classification results

	Experiment 1					
	True LED	True Vibration	True Sound	True ALS	True DW	Class precision
Pred. LED	44	60	39	19	54	20.37%
Pred. Vibration	73	69	27	18	27	32.24%
Pred. Sound	129	165	291	136	43	38.09%
Pred. ALS	0	0	0	0	0	0.00%
Pred. DW	91	79	32	21	263	54.12%
Class recall	13.06%	18.50%	74.81%	0.00%	67.96%	

	Experiment 2					
	True ALS1	True ALS2	True ALS3	True ALS4	True DW	Class precision
Pred. ALS3	18	35	25	15	17	22.73%
Pred. ALS1	36	9	10	51	1	33.64%
Pred. ALS2	54	206	153	107	108	32.80%
Pred. ALS4	37	20	12	17	12	17.35%
Pred. DW	19	118	70	32	508	68.01%
Class recall	21.95%	53.09%	9.26%	7.66%	78.64%	

mendations for notification modalities when a user is multitasking in a smart environment:

1. Audio and ALS notifications are the best for attracting the users' immediate attention. The combination of these modalities with vibration or LED modalities increases the perceptibility of a notification, but it is not necessary to use both audio and ALS at the same time. However, these two notification modalities have significant privacy implications and must be considered as distractive and also potentially inappropriate, depending on the users' context.
2. ALS notifications are not affected by persistence – instead, we noted that response time to pure ALS notifications was longer only when the transitions between the on-off states were slow and gradual. An ALS system is thus more successful in attracting user attention when designed with a rapid off->on transition time and calmer, non-urgent notifications should be delivered using gradual transitions.
3. Vibration and device LED notifications are easy to miss. These modalities are more protective of users' privacy, though under certain circumstances they can be publically perceptible.
4. The device LED is appropriate for private, non-urgent notifications (the user will eventually see it but with a likely long delay). It also serves as a reminder for notifications issued via other modalities (audio, haptic) which the user might have temporarily ignored or missed altogether. However, when an ALS is present, it is best to use the ALS for this role as the LED is the least perceptible modality.
5. Vibrations are appropriate for private, urgent notifications only if the duration of the pattern is long enough to allow it a chance to be perceived. They also encourage "phantom vibration" experiences, especially under certain contexts of use.
6. The intelligent handling of notifications should strongly consider contextual cues about the importance of the application to the user (this can be mined, for example by looking at app usage statistics), and also contextual cues about the presence of other people around the user (e.g. discoverable via location semantics, or bluetooth scanning for unknown devices, such as other phones).

Expanding on the latter point, it appears from our results that a common consensus exists across users for the context cases where notification disturbance is not wanted, and for those cases where sound is an appropriate notification modality to use (perhaps alongside others). From that point on, specific notification modality combinations seem to be a subjective preference of each individual, or that perhaps more contextual attributes (or even attribute values) need to be taken into account in order to adequately model notification modality preference. However, this result means that in the future, it might be possible for a notification management service (or even individual apps) to proactively adjust device ringer modes or per-app notification modalities, based on the application type and surrounding context, assuming of course that the latter can be accurately inferred.

7. Conclusions

Our paper offers a comprehensive review of literature on the perceptibility of smartphone notifications and discusses implications from literature about how smartphone notifications can be extended to work with ALS. As can be seen from the literature that we have explored, and from papers in very recent workshops on notifications [36,44,47], the topic of modality perceptibility and appropriateness under context is still very much open for research, particularly when considering notifications involving an ALS. Our work does not aim or claim to provide a comprehensive finding to all the open issues in this domain – this is a target far beyond the scope of a single paper. However, it is the first study in this area which attempts to establish a baseline regarding the perceptibility of smartphone and ALS notifications, using a thorough controlled environment trial. Although we cannot claim generalizability of these results (the only true test would be studies that replicate ours), through this investigation, we have aimed to ground the discourse on notification modalities on a more solid basis and to provide some reliable design guidelines, representing one highly ecologically valid scenario of mobile device use in smart environments.

Our results for R1 & R2 show that sound, whether on its own or in combination with other modalities is strongly related to the perceptibility of a notification, when the user is engaged in an attention-demanding task. Additionally, in our investigation of R3, we found that ambient lighting is successful in attracting users' attention during such circumstances (depending on how gradual transitions between on and off states are).

We did not find any evidence to support the use of the device LED, or, in contrast to other researchers, vibration. For the latter, literature suggests that the reasons might be due to the underlying psycho-cognitive mechanism of dealing with haptic stimuli under multitasking. We would have expected the results of these field studies to at least partially confirm our own, since field studies are largely regarded as more externally valid. This contrast in results highlights that previous field studies might have been premature (without considering the role of contextual factors in perceptibility and response time, or the full mechanism behind Android notifications) or too loosely structured to ensure a reasonable degree of internal validity.

Our study has some limitations as can be expected from a laboratory environment. We examined only one contextual setting, where multitasking overloaded participants' visual and audio perception. The experiment should be repeated with the user engaging in other types of task, such as leisurely ones. A further limitation was the placement of the mobile device on the desk. Although literature indicates that this is the most representative and common case in real use of mobile devices, it would be interesting to repeat the experiment with the device placed in contact with the user's body, something that might increase the perceptibility of vibration. This positioning of the device (e.g. in a pocket), might diminish the user's ability to perceive visual and audio cues. Hence it might be worth expanding the notification space to wearable devices (e.g. vibrating smart watch) that form part of the user's device ecology. A final consideration is the use of ALS modalities in a shared space, particularly regarding privacy issues or conflicts, e.g. if used simultaneously as a room's main lighting source.

In general, we can state here that appropriate context awareness and consideration for the users' physiology and cognitive ability has the potential to improve the design of notifications in terms of their modality, increasing their effectiveness and reducing annoyance, as indicated by our findings on R4. An added advantage of careful notification design is the conservation of device battery resources, by avoiding the use of modalities that are ineffective under the given context (especially when being able to off-load notifications to an ALS). There is scope for operating system developers to implement APIs that can be utilized by all third-party developers. As demonstrated by [50], a simple periodic assessment of the device's placement context using integrated sensors can provide cues for selecting notification modalities can either by the application, or managed by the OS itself.

References

[1] F. Alt, A.S. Shirazi, A. Kaiser, K. Pfeuffer, E. Gurkan, A. Schmidt, P. Holleis and M. Wagner, Exploring ambient visualizations of context information, in: *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2010 8th IEEE International Conference on, IEEE*, 2010, pp. 788–791, 978-1-4244-6606–978-1-4244-1. ISBN 978-1-4244-6605-4. doi:10.1109/PERCOMW.2010.5470542.

[2] M. Altosaar, R. Vertegaal, C. Sohn and D. Cheng, AuraOrb: Social notification appliance, in: *CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06)*, ACM Press, 2006, p. 381. ISBN 1-59593-298-4. doi:10.1145/1125451.1125533.

[3] J.R. Blum, J.R. Cooperstock and J. Cauchard, Pseudo-ambience: Filling the gap between notifications and continuous information displays, in: *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, UbiComp '18, ACM*, 2018, pp. 1222–1227. ISBN 978-1-4503-5966-5. doi:10.1145/3267305.3274111.

[4] A. Chan, K. MacLean and J. McGrenere, Learning and identifying haptic icons under workload, in: *Eurohaptics Conference, 2005 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, World Haptics 2005. First Joint, IEEE, 2005, pp. 432–439. ISBN 0-7695-2310-2. doi:10.1109/WHC.2005.86.

[5] Y.-J. Chang and J.C. Tang, Investigating mobile users' ringer mode usage and attentiveness and responsiveness to communication, in: *Proceedings of the 17th International Conference on Human–Computer Interaction with Mobile Devices and Services, MobileHCI '15, ACM*, 2015, pp. 6–15. ISBN 978-1-4503-3652-9. doi:10.1145/2785830.2785852.

[6] K. Davis, E.B. Owusu, L. Marcenaro, L. Feijs, C. Regazzoni and J. Hu, Effects of ambient lighting displays on peripheral activity awareness, *IEEE Access* **5** (2017), 9318–9335, ISSN 2169-3536. doi:10.1109/ACCESS.2017.2703866.

[7] A.K. Dey, K. Wac, D. Ferreira, K. Tassini, J.-H. Hong and J. Ramos, Getting closer: An empirical investigation of the proximity of user to their smart phones, in: *Proceedings of the 13th International Conference on Ubiquitous Computing, UbiComp '11, ACM*, 2011, pp. 163–172. ISBN 978-1-4503-0630-0. doi:10.1145/2030112.2030135.

[8] D. Dobbelstein, S. Herrdum and E. Rukzio, InScent: A wearable olfactory display as an amplification for mobile notifications, in: *Proceedings of the 2017 ACM International Symposium on Wearable Computers, ISWC '17, ACM*, 2017, pp. 130–137. ISBN 978-1-4503-5188-1. doi:10.1145/3123021.3123035.

[9] A. Exler, C. Dinse, Z. Günes, N. Hammoud, S. Mattes and M. Beigl, Investigating the perceptibility different notification types on smartphones depending on the smartphone position, in: *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Com-*

puters, UbiComp '17, ACM, 2017, pp. 970–976. ISBN 978-1-4503-5190-4. doi:10.1145/3123024.3124560.

[10] J.A. Gallud and R. Tesoriero, Smartphone notifications: A study on the sound to soundless tendency, in: *Proceedings of the 17th International Conference on Human–Computer Interaction with Mobile Devices and Services Adjunct, MobileHCI '15, ACM*, 2015, pp. 819–824. ISBN 978-1-4503-3653-6. doi:10.1145/2786567.2793706.

[11] S. Garzonis, C. Bevan and E. O'Neill, Mobile service audio notifications: Intuitive semantics and noises, in: *Proceedings of the 20th Australasian Conference on Computer–Human Interaction: Designing for Habitus and Habitat, OZCHI '08, ACM*, 2008, pp. 156–163. ISBN 978-0-9803063-4-7. doi:10.1145/1517744.1517793.

[12] T. Gross, Ambient interfaces: Design challenges and recommendations, in: *Human–Computer Interaction: Theory and Practice, CRC Press*, 2003, pp. 68–72. ISBN 978-0-8058-4931-8.

[13] R. Hansson, P. Ljungstrand and J. Redström, Subtle and public notification cues for mobile devices, in: *Ubicomp 2001: Ubiquitous Computing*, G.D. Abowd, B. Brumitt and S. Shafer, eds, Vol. 2201, Springer, Berlin Heidelberg, 2001, pp. 240–246, ISBN 978-3-540-45427-4. 978-3-540-42614-1. doi:10.1007/3-540-45427-6_20.

[14] F. Hemmert and G. Joost, Ambient Life: Calm and Excited Pulsation as a Means of Life-like Permanent Tactile Status Display in Mobile Phones, 2008.

[15] A. Kameas, S. Bellis, I. Mavrommati, K. Delaney, M. Colley and A. Pounds-Cornish, An architecture that treats everyday objects as communicating tangible components, in: *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications*, 2003. (PerCom 2003), pp. 115–122. doi:10.1109/PERCOM.2003.1192733.

[16] F. Kawsar, K. Fujinami and T. Nakajima, Augmenting everyday life with sentient artefacts, in: *Proceedings of the 2005 Joint Conference on Smart Objects and Ambient Intelligence: Innovative Context-Aware Services: Usages and Technologies, SOc-EUSAI '05, ACM*, 2005, pp. 141–146. ISBN 978-1-59593-304-1. doi:10.1145/1107548.1107587.

[17] A. Komninos, J. Besharat, V. Stefanis and J. Garofalakis, Perceptibility of mobile notification modalities during multitasking in smart environments, in: *2018 14th International Conference on Intelligent Environments (IE)*, 2018, pp. 17–24. doi:10.1109/IE.2018.00011.

[18] A. Komninos, E. Frengkou and J. Garofalakis, Predicting user responsiveness to smartphone notifications for edge computing, in: *Ambient Intelligence*, A. Kameas and K. Stathis, eds, Lecture Notes in Computer Science, Springer International Publishing, 2018, pp. 3–19. ISBN 978-3-030-03062-9. doi:10.1007/978-3-030-03062-9_1.

[19] Z. Ma, S. Mirzamohammadi and A.A. Sani, Understanding sensor notifications on mobile devices, in: *Proceedings of the 18th International Workshop on Mobile Computing Systems and Applications, HotMobile '17, ACM*, 2017, pp. 19–24. ISBN 978-1-4503-4907-9. doi:10.1145/3032970.3032978.

[20] A. Mashhadi, A. Mathur and F. Kawsar, The myth of subtle notifications, in: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, UbiComp '14 Adjunct, ACM*, 2014, pp. 111–114. ISBN 978-1-4503-3047-3. doi:10.1145/2638728.2638759.

[21] A. Matviienko, W. Heuten, V. Cobus, H. Müller, J. Fortmann, A. Löcken, S. Boll, M. Rauschenberger, J. Timmermann and C. Trappe, Deriving design guidelines for ambient light systems, in: *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*, ACM Press, 2015, pp. 267–277. ISBN 978-1-4503-3605-5. doi:10.1145/2836041.2836069.

[22] A. Matviienko, M. Rauschenberger, V. Cobus, J. Timmermann, J. Fortmann, A. Löcken, H. Müller, C. Trappe, W. Heuten and S. Boll, Towards new ambient light systems: A close look at existing encodings of ambient light systems, *Interaction Design and Architecture(s) Journal* **26** (2015), 10–24.

[23] A. Mehrotra, M. Musolesi, R. Hendley and V. Pejovic, Designing content-driven intelligent notification mechanisms for mobile applications, in: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '15, ACM*, 2015, pp. 813–824. ISBN 978-1-4503-3574-4. doi:10.1145/2750858.2807544.

[24] A. Mehrotra, V. Pejovic, J. Vermeulen, R. Hendley and M. Musolesi, My phone and me: Understanding people's receptivity to mobile notifications, in: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, ACM*, 2016, pp. 1021–1032. ISBN 978-1-4503-3362-7. doi:10.1145/2858036.2858566.

[25] H. Müller, A. Kazakova, M. Pielot, W. Heuten and S. Boll, Ambient timer – unobtrusively reminding users of upcoming tasks with ambient light, in: *Human–Computer Interaction – INTERACT 2013*, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2013, pp. 211–228, 978-3-642-40483–978-3-642-2. ISBN 978-3-642-40482-5. doi:10.1007/978-3-642-40483-2_15.

[26] H. Müller, A. Löcken, W. Heuten and S. Boll, Sparkle: An ambient light display for dynamic off-screen points of interest, in: *Proceedings of the 8th Nordic Conference on Human–Computer Interaction: Fun, Fast, Foundational, NordiCHI '14, ACM*, 2014, pp. 51–60. ISBN 978-1-4503-2542-4. doi:10.1145/2639189.2639205.

[27] T. Okoshi, J. Ramos, H. Nozaki, J. Nakazawa, A.K. Dey and H. Tokuda, Attelia: Reducing user's cognitive load due to interruptive notifications on smart phones, in: *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2015, pp. 96–104. doi:10.1109/PERCOM.2015.7146515.

[28] V. Pejovic and M. Musolesi, InterruptMe: Designing intelligent prompting mechanisms for pervasive applications, in: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '14, ACM*, 2014, pp. 897–908. ISBN 978-1-4503-2968-2. doi:10.1145/2632048.2632062.

[29] V. Pejovic, M. Musolesi and A. Mehrotra, Investigating the role of task engagement in mobile interruptibility, in: *Proceedings of the 17th International Conference on Human–Computer Interaction with Mobile Devices and Services Adjunct, MobileHCI '15, ACM*, 2015, pp. 1100–1105. ISBN 978-1-4503-3653-6. doi:10.1145/2786567.2794336.

[30] F. Perteneder, E.-M.B. Grossauer, J. Leong, W. Stuerzlinger and M. Haller, Glowworms and fireflies: Ambient light on large interactive surfaces, in: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, ACM*, 2016, pp. 5849–5861. ISBN 978-1-4503-3362-7. doi:10.1145/2858036.2858524.

[31] M. Pielot, K. Church and R. de Oliveira, An in-situ study of mobile phone notifications, in: *Proceedings of the 16th International Conference on Human–Computer Interaction with Mobile Devices & Services, MobileHCI '14, ACM*, 2014, pp. 233–242. ISBN 978-1-4503-3004-6. doi:10.1145/2628363.2628364.

[32] M. Pielot, R. de Oliveira, H. Kwak and N. Oliver, Didn'T you see my message?: Predicting attentiveness to mobile instant messages, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14, ACM*, 2014, pp. 3319–3328. ISBN 978-1-4503-2473-1. doi:10.1145/2556288.2556973.

[33] M. Pielot, A. Vradi and S. Park, Dismissed!: A detailed exploration of how mobile phone users handle push notifications, in: *Proceedings of the 20th International Conference on Human–Computer Interaction with Mobile Devices and Services, MobileHCI '18, ACM*, 2018, pp. 3–1311. ISBN 978-1-4503-5898-9. doi:10.1145/3229434.3229445.

[34] H. Pohl, J. Medrek and M. Rohs, ScatterWatch: Subtle notifications via indirect illumination scattered in the skin, in: *Proceedings of the 18th International Conference on Human–Computer Interaction with Mobile Devices and Services, MobileHCI '16, ACM*, 2016, pp. 7–16. ISBN 978-1-4503-4408-1. doi:10.1145/2935334.2935351.

[35] B. Poppinga, W. Heuten and S. Boll, Sensor-based identification of opportune moments for triggering notifications, *IEEE Pervasive Computing* **13**(1) (2014), 22–29, ISSN 1536-1268. doi:10.1109/MPRV.2014.15.

[36] B. Poppinga, M. Pielot, N. Henze, N. Oliver, K. Church and A.S. Shirazi, Smarttention, please! Intelligent attention management on mobile devices, in: *Proceedings of the 17th International Conference on Human–Computer Interaction with Mobile Devices and Services Adjunct, MobileHCI '15, ACM*, 2015, pp. 1066–1069. ISBN 978-1-4503-3653-6. doi:10.1145/2786567.2795399.

[37] Z. Pousman and J. Stasko, A taxonomy of ambient information systems: Four patterns of design, in: *Proceedings of the Working Conference on Advanced Visual Interfaces, AVI '06, ACM*, 2006, pp. 67–74. ISBN 1-59593-353-0. doi:10.1145/1133265.1133277.

[38] Q. Qin, M. Rohs and S. Kratz, Dynamic ambient lighting for mobile devices, in: *Proceedings of the 24th Annual ACM Symposium Adjunct on User Interface Software and Technology, UIST '11 Adjunct, ACM*, 2011, pp. 51–52. ISBN 978-1-4503-1014-7. doi:10.1145/2046396.2046418.

[39] A. Sahami Shirazi and N. Henze, Assessment of notifications on smartwatches, in: *Proceedings of the 17th International Conference on Human–Computer Interaction with Mobile Devices and Services Adjunct, MobileHCI '15, ACM*, 2015, pp. 1111–1116. ISBN 978-1-4503-3653-6. doi:10.1145/2786567.2794338.

[40] A. Sahami Shirazi, N. Henze, T. Dingler, M. Pielot, D. Weber and A. Schmidt, Large-scale assessment of mobile notifications, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14, ACM*, 2014, pp. 3055–3064. ISBN 978-1-4503-2473-1. doi:10.1145/2556288.2557189.

[41] V.J. Sauer, S.C. Eimler, S. Maafi, M. Pietrek and N.C. Krämer, The phantom in my pocket: Determinants of phantom phone sensations, *Mobile Media & Communication.* **3**(3) (2015), 293–316, ISSN 2050-1579, 2050-1587. doi:10.1177/2050157914562656.

[42] L.D. Turner, S.M. Allen and R.M. Whitaker, Reachable but not receptive: Enhancing smartphone interruptibility prediction by modelling the extent of user engagement with, *Notifications* **40** (2017), 480–494, ISSN 1574-1192. doi:10.1016/j.pmcj.2017.01.011.

[43] M.H. Vastenburg, D.V. Keyson and H. de Ridder, Considerate home notification systems: A user study of acceptability of notifications in a living-room laboratory, *International Journal of Human–Computer Studies* **67**(9) (2009), 814–826, ISSN 10715819. doi:10.1016/j.ijhcs.2009.06.002. doi:10.1016/j.ijhcs.2009.06.002.

[44] A. Voit, B. Poppinga, D. Weber, M. Böhmer, N. Henze, S. Gehring, T. Okoshi and V. Pejovic, UbiTtention: Smart and ambient notification and attention management, in: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, UbiComp '16, ACM*, 2016, pp. 1520–1523. ISBN 978-1-4503-4462-3. doi:10.1145/2968219.2968542.

[45] D. Weber, Notifications in a Multi-Device Environment, University of Stuttgart, 2015. http://elib.uni-stuttgart.de/handle/11682/3515.

[46] D. Weber, Towards smart notification management in multi-device environments, in: *Proceedings of the 19th International Conference on Human–Computer Interaction with Mobile Devices and Services, MobileHCI '17, ACM*, 2017, pp. 68–1682. ISBN 978-1-4503-5075-4. doi:10.1145/3098279.3119921.

[47] D. Weber, A.S. Shirazi, S. Gehring, N. Henze, B. Poppinga, M. Pielot and T. Okoshi, Smarttention, please!: 2Nd workshop on intelligent attention management on mobile devices, in: *Proceedings of the 18th International Conference on Human–Computer Interaction with Mobile Devices and Services Adjunct, MobileHCI '16, ACM*, 2016, pp. 914–917. ISBN 978-1-4503-4413-5. doi:10.1145/2957265.2965025.

[48] C.D. Wickens, Multiple resources and performance prediction. Theoretical issues in ergonomics, *Science* **3**(2) (2002), 159–177, ISSN 1463-922X. doi:10.1080/14639220210123806.

[49] C.D. Wickens, Multiple resources and mental workload. Human factors, *The Journal of the Human Factors and Ergonomics Society* **50**(3) (2008), 449–455, ISSN 0018-7208. doi:10.1518/001872008X288394.

[50] J. Wiese, T.S. Saponas and A.J.B. Brush, Phoneprioception: Enabling mobile phones to infer where they are kept, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, ACM*, 2013, pp. 2157–2166. ISBN 978-1-4503-1899-0. doi:10.1145/2470654.2481296.