Hush now! Context factors behind smartphone ringer mode changes

Andreas Komninos*, Antonis-Elton Frengkou, John Garofalakis

Computer Engineering and Informatics Department, University of Patras, Rio 26504, Greece

Abstract

In this paper we examine the contextual factors driving users' decisions to change smartphone ringer modes, and in particular switching between a full (normal) mode where all notification modalities are enabled, and a discreet ringer mode (i.e. where audio is disabled). We add to the limited current literature through a qualitative study based on the theory of Reasoned Action Approach, and the longitudinal collection of empirical data from users' devices. Our findings demonstrate that temporal, spatial and social context are strong determinants for the decision-making process of switching ringer modes. In particular, social context emerges in a much more complex form than described in previous studies, questioning if it can be accurately captured in all its richness. On the other hand, temporal and spatial context are found to explain much of the variation in ringer mode use, enabling prediction of appropriate ringer mode states using machine learning. Our findings open new directions for cross-cultural, longitudinal research.

Keywords: Smartphones, ringer mode, context awareness, machine learning

1. Introduction

Smartphones are undeniably a major element of modern day life. Their ability to connect us to people and events in our locality and at a global scale is a key feature that makes them indispensable. Much of this connectivity is initiated through multimodal notifications, which are issued through visual, audio and tactile (vibration) cues. The ability of modern smartphones to notify us about events is both a blessing and also a curse, and we have all undoubtedly experienced the inconvenience of disturbance during times where we would prefer the device to be silent. To

*Corresponding author

Email addresses: akomninos@ceid.upatras.gr (Andreas Komninos), frengkou@ceid.upatras.gr (Antonis-Elton Frengkou), garofala@ceid.upatras.gr (John Garofalakis)

date, the management of notification modality has been left to the individual user. Through physical or touchscreen controls, or using simple time and calendar-based rules, the user is able to manually or programatically set their device's ringer mode to a *discreet* mode (either vibrate-only, or muting all notification modalities except the visual one). For the rest of the paper we treat these discreet ringer modes jointly as the *"silent mode"*, as per the work of [1, 2].

Users often set their mobiles to silent according to their personal preference and goals (e.g. when working or studying), or because contextually relevant societal norms dictate a level of discreetness [3]. This manual handling of ringer mode can cause problems for the users. As daily experience shows, users might frequently forget to perform the switch, resulting in much embarrassment or annoyance when the device rings at an inappropriate time. At other times, the user might not even be aware of societal norms that dictate switching the phone to silent (e.g. when travelling to a new country, or being present at a location where the local micro-culture differs from the user's habits, for example, attending a meeting at another company's offices). Of course, the reverse also applies: users might forget to switch back to normal mode after a period of keeping their device silent, thereby missing incoming events such as phone-calls and messages.

To mitigate these problems, modern smartphone operating systems and third party applications offer a range of rule-based automation functions, but extensive user effort is needed to imagine, set-up and fine-tune good rules [2]. Most users simply adopt manufacturer default settings and don't delve into exploring these functions [4]. Even so, rule-based systems have difficulty in accounting for abrupt changes in user contexts (e.g. when travelling to another country, or when taking time off for holidays, health, or other reasons).

Hence, the question arises, whether it might be possible for a mobile device to learn the user's preferences and the social rules that drive the decision to switch to silent, and thereby be able to offer assistance to users, either by performing autonomic actions (i.e. automatically switching between full and silent modes) or prompting the user to do it themselves. Although the idea was proposed back in 2003 by Sieworek et al. [24], surprisingly, there exist very few publications addressing the challenge of automated ringer mode management. Related studies either lack explicative interpretations of empirical findings [2], or are based on self-reported information, without empirical validation [1, 25].

The goal of our paper is to obtain insights about the contextual factors that drive users to switch ringer modes, and to take a first step towards determining whether the autonomous learning of these factors can be achieved. Adding to previous research, we report findings from a user survey based on a structured theory of human behaviour (Reasoned Action Approach), which offer explicative interpretations of the contextual factors leading to ringer mode changing behaviour. Further, we contribute by presenting empirical results on ringer mode changing behaviour, collected both through experience sampling and via automatic context capture on user smartphones. We demonstrate, using machine-learning approaches, the practicalities and limitations for predicting appropriate circumstances to manage ringer modes.

2. Related work

The disruptive nature of mobile notifications is repeatedly documented in research. To cope, users adopt various interruption management strategies, which include manually setting their device ringer mode to silent. In Voit et al. [4], the researchers found this strategy to be employed by very few of their 16 participants, but larger studies report this practice to be much more extensive (e.g. [5]). Several works including [6–9] demonstrate that setting one's phone to silent and avoiding notifications improves concentration and productivity and reduces hyperactivity, but on the other hand, can leave the user anxious about appearing less responsive, worried about missing important events and feeling less connected to their social groups.

To mitigate negative aspects of manual management of ringer mode, modern smartphones offer some assistance, by including a "do not disturb" (DND) feature. This allows users to schedule periods of time where the device can automatically switch ringer modes (e.g. "between 23:00 and 07:00"), to temporarily mute the device for a fixed period of time (e.g. "for 1 hour"), or when a scheduled calendar event is programmed [10]. Settings for notification modality options on a per-app basis are also available. However, this programming is fixed and thus offers little flexibility for users whose contexts vary significantly during the day [2]. Users also may face difficulty in correctly estimating appropriate lengths for temporary fixed muting periods. Location-based support for switching ringer modes depending on geographical coordinates, or the detection of a known Wi-Fi network, has been offered since Android 10 through the "Rules" settings, although this is currently only available on Google Pixel devices. Consumer apps like IFTTT (http://www.ifttt.com) can offer better levels of automation, since the user can define rules that associate various context features with ringer modes (e.g. location, time, scheduled meetings) but all these options require an extensive initial effort from the user to set up. Maues et al [11] demonstrate that some of these rules can be automatically inferred from past user behaviour, however learning based on historical records cannot adapt rules quickly enough to respond to the very dynamic nature of daily life. Voit et al. [4] found that most users do not alter their device settings and adopt the default manufacturer settings.

Previous research has identified a range of challenges and opportunities in intelligent notification management. An extensive survey of related work is offered by Anderson et al. [12], where a pipeline for the operation of an "attention management system" is presented, including steps for sensing, processing, inferring (context), modelling interruptibility, and finally managing interruptions, e.g. deferring notifications, or altering the delivery modality. An intelligent system could apply the latter approach to *individual notifications* (e.g. allow important or urgent notifications to use more than just the visual channel), or to *all notifications* (equivalent to setting the device ringer mode to a discreet setting). In fact, the latter approach equates to what users currently do without assistance, i.e. to set their device too silent mode. Of course, an intelligent system would still be helpful even with such blanket approaches, because the burden of remembering to switch (to either setting) would be removed from the user [3].

There is a rapidly growing body of publications related to individual notification management that take into account the contexts of user activity, location, time and social circumstances amongst others. So far, the literature has focused on determining user receptivity and response time (e.g. depending on content, social relationship with sender, generating application etc. [5, 13–16]), predicting opportune moments to disrupt the user and deliver a notification (e.g. [17–21], or to assess the perceptibility of notifications (e.g.[3, 13, 22, 23]). For the interested reader, [12] provides excellent coverage of further important work in this domain. However, we were able to identify just four publications which directly focus on the topic of managing the ringer mode of the device [1, 2, 24, 25], which we discuss next.

Early work by Sieworek et al. [24] outlined how the appropriate ringer mode could be inferred by analysing data from the device microphone, calendar information, light sensor and accelerometer. No formal evaluation of the effectiveness of the approach was offered. More recently, Qin et al. [2] combined sensor data from the device (microphone, ambient light, location, compass, Bluetooth and wi-fi) to form rules describing four user activity contexts. Using machine learning classifiers, they achieve up to 90% accuracy in correctly guessing the appropriate ringer mode for these contexts, using however only fixed training and validation sets. Another issue with this study is the limited extent of the user context scenarios that it is able to recognise, hence it is uncertain how this system may adapt to the variety of contexts of everyday life. Finally, this work is more descriptive rather than explicative, since it does not offer the necessary insights behind the motivations for ringer mode switching behaviour, which would be essential for developing a more flexible system.

To explicate the motivating factors behind ringer mode switching, Chang & Tang [1] and Exler et al. [25] provide some preliminary findings. In [1] a survey and a diary study demonstrated frequent use of silent mode in contexts where the user did not want to be disturbed themselves (e.g. sleeping, at work) but also in contexts where the user is conscious about not disturbing others. On the other hand, users reported reverting to normal mode when expecting some incoming communication and to maintain awareness of notifications. Overall, users reported setting their device to normal mode for most of the time. The most recent study relating to ringer mode use is [25]. In this study, participants were asked to imagine various situational contexts (locations and location-based activities) and indicate their preference towards notification modalities. The main findings related to strong correlations between location type, and receptivity to notifications, perceived disruptiveness and level of task engagement. Hence, certain locations appear to be strongly associated with specific activities and carry common semantic connotations across users.

Unfortunately, neither of these two studies focus much on the negative aspects of switching too silent. Another issue is that both studies relied exclusively on self-reported data from a limited population sample. While a range of insights are uncovered, their findings were not empirically validated. Further, neither study is based on any structured theoretical approach to explaining human behaviour. As such, they focus mostly on explaining the semantics of context that can be automatically inferred from sensors (e.g. location, activity), and less on the types of context such as the social environment of the user, which can be harder to capture.

Thus, for this paper, we aim to expand the literature in this specific problem, by adopting a dual approach. First, we perform a qualitative survey with 30 participants, using the Reasoned Action Approach theory (RAA) [26] as a lens to focus our understanding of the factors related to the behaviour of switching ringer modes. This more structured approach yields new insights about the decision-making process of users and significantly adds to the current work of Chang & Tang [1] and Exler et al. [25]. Next, we perform a study of ringer mode use with 44 participants, through the automated collection of context factors (spatial and temporal), to empirically validate the survey findings about these factors. We demonstrate that these two alone are able to explain a large proportion of the variation of ringer mode switching behaviour.

3. Study 1: Qualitative survey

To improve the theoretical grounding of our work, we started with a survey based on the RAA theory [26]. RAA provides a framework for predicting human behaviour, by determining their intentions. Intentions are viewed as the result of interplay between three factors: a) an individual's attitudes towards the behaviour, (e.g. "I think that the *[behaviour]* is beneficial to me and I want to adopt it"); b) the norms affecting the behaviour as perceived by the individual (e.g. "Other people would want me to adopt this *[behaviour]*") , and; c) the control of the individual over the behaviour, as perceived again by the individual, and which may depend on cognitive, contextual, skill or other factors (e.g. "I find it difficult to perform the *[behaviour]* because...").

As a necessary prerequisite, first the behaviour to be examined must be specified, and for our study this was defined as *"To change my smartphone's ringer mode to silent, when appropriate for my current environment and goals"*. Secondly, the research population target also needs to be specified, and for our study it is defined as *"young adults including students and professionals who own a mobile phone with switchable ringer modes"*. While this specification excludes a significant part of the population, we consider it appropriate, especially since literature in directly related work (e.g. [1, 2, 25]) also refers to a similar user group, hence allowing us to directly compare our findings and better position them within the literature.

Based on this, we designed a survey to elicit readily accessible behavioural outcomes, normative referents and control factors, which form the core constructs of RAA. The survey was administered online, publicising to mailing lists of our university, including students, researchers and alumni. No incentive was provided for participation. In total, 30 people responded to our survey. Introductory questions revealed a balanced gender split (14 female, 15 male, 1 did not wish to indicate). The respondents' age was indicated by selecting from specified age groups and was mostly people aged 18-30 (18-24: 17; 25-30: 12; 31-35: 1). Most respondents indicated their occupation as full-time students (18), while other choices included alumni in full-time employment (8), part-time employed (2) or unemployed (2). Most respondents use an Android device as their daily smartphone (22), 7 respondents use an iPhone and 1 respondent indicated use of a feature phone.

The core survey questions were organised in sections, to elicit the salient beliefs that pertain to the core RAA constructs (behavioural outcomes, normative referents and control beliefs). The list of questions were taken verbatim from [27] under the "pilot questionnaire" section, replacing the behaviour with the one defined for our study. We asked respondents to provide a free-form textual response to the questions, providing between 3-6 statements for each question. The free-form response content was manually coded independently by three researchers from our team, each using inductive coding and a flat coding frame. Related codes were merged as appropriate through consensus, to identify pertinent themes in the responses.

3.1. Behavioural outcomes

3.1.1. Advantages of "switching to silent"

Here, we identified 8 main advantages. The most commonly mentioned was "Respecting others" with 16 responses (e.g. *"There are places where the mobile can annoy other people, like when you are at the cinema"*). This was closely followed by "Avoiding interruptions while concentrating on other tasks" (13), "Avoiding interruptions and annoyance in general" (11) and "Avoiding noise" (9). Other responses were more specific towards the context of interruption: to "Avoid interruption while relaxing or sleeping" (6), to "Avoid other people" (4), to "Save battery" (3) and to "Maintain privacy" (1). From this, we notice that overall the avoidance of unwanted interruptions towards others and towards the individual seems to be the predominant perceived advantage, and respondents mention a range of contexts under which interruptions are unwanted.

3.1.2. Disadvantages of "switching to silent"

In total, a further 8 disadvantages were identified, with the predominant category here being "Missing important or urgent communication" (26). This was followed by "Missing important app notifications" (7). Other disadvantages of switching to silent included "Upsetting other people" (3), "Missing alarms" (2), "Becoming isolated" (1), "Forgetting to return to normal mode" (2), "Missing reminders" (1) and being "Unable to locate the device" (1).

3.1.3. Other thoughts

The last question asked respondents to write any other thoughts that came to mind about switching their phones to silent. In this section, some respondents again mentioned aspects which could be re-classified as advantages or disadvantages. The former include some aspects already mentioned in the advantages section, such as "avoiding interruptions" (6), "avoid disturbing others" (3), "saving battery" (3) and "maintain privacy" (3). A novel aspect was "reducing my addiction to the mobile" (2). One respondent saw this from a reverse perspective, mentioning that if other users switched to silent, then he wouldn't be annoyed as often. In terms of further disadvantages, some respondents mentioned "forgetting to switch back to normal mode" (2), that "vibration is loud" and hence also "annoying" (1), and a having "feelings of insecurity" (1). Further from these comments, respondents mentioned being unable to select which apps to silence (1), that new ways to make notifications less annoying should be developed (2), and that silent mode should be used with care (1). One more participant mentioned that for all perceived disadvantages of "normal mode", she keeps her device generally on "silent".

3.2. Normative referents

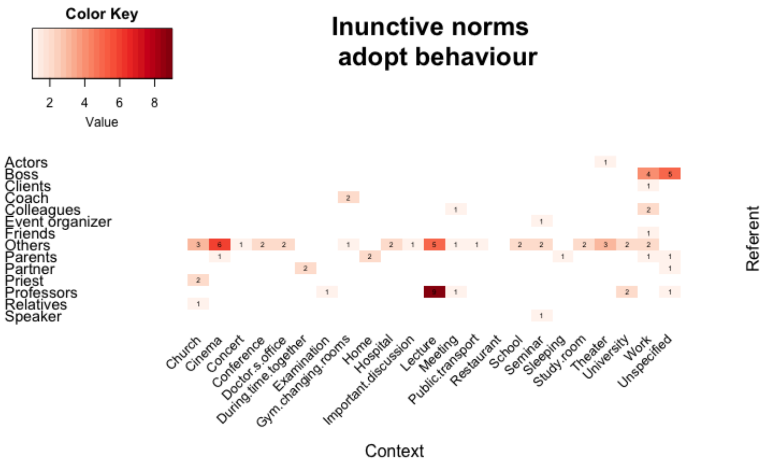
In this part of the survey, four questions aim to investigate the injunctive and descriptive norms affecting behaviour. Injunctive norms identify the referents (people) who would approve or disapprove of respondents adopting the behaviour (thereby setting social and peer-based expectations for the user). Descriptive norms identify the referents who do, or do not, also engage in this behaviour themselves (i.e. providing examples of behaviour that the user looks to mimic). For these questions, we asked respondents not only to mention the referents, but also to specify, if they could, the context in which these referents operate (e.g. *"My boss also puts her phone on silent when we are at work"*). Once more, we asked participants to identify 3-6 referent-context combinations.

Since participants responded to these completely open-ended questions, drawing from their own experience, they were able to identify a very diverse set of referent-context combinations, as will be shown next. Therefore in the analyses of this section, we pay less attention to the actual frequencies reported, but more towards the *occurrence* of any referent-context combination.

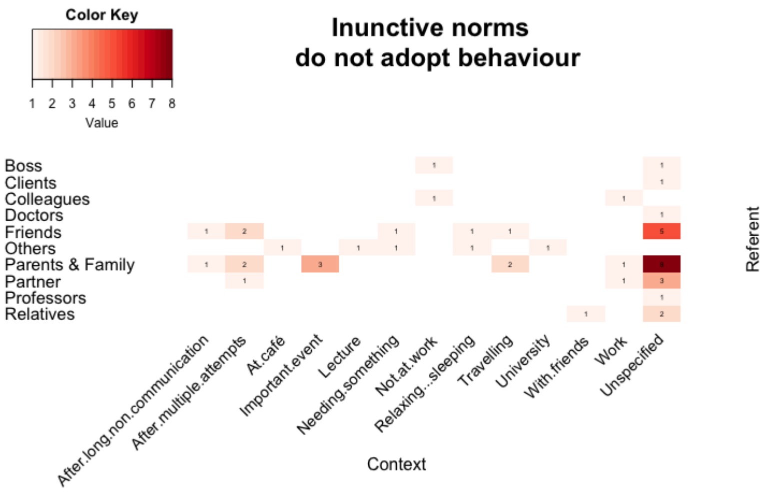
3.2.1. Injunctive norms

Injunctive norms where elicited with two separate questions about other people who approve or disapprove of users adopting the behaviour. In terms of approval, respondents identified a total of 14 referent roles and 23 context types. As seen in Figure 1a, a rich picture emerges of the respondents' awareness of appropriate contexts of switching to silent. In terms of context, we observe that many responses relate to specific location types (e.g. "cinema", "hospital"), while other responses relate to types of activity (e.g. "lecture", "during time together") irrespective of where that activity takes place. On the other hand, in terms of referents, we note that some sparsity exists, which is natural since certain referents are "roles" which only have meaning in very specific contexts (e.g. "Actors", paired with "Theater"). In these contexts, we note that the specific "role" mentioned is a person with authority over the context, which can be a tight coupling of the space's function and the current activity e.g. a professor during a class, or a priest at the church), or a loose coupling where the activity becomes the dominant factor (e.g. spending time with a partner at home or at a cafe). We note that the generic description "Others" features prominently across all context types. This finding correlates well with the previously reported advantage of "respecting others" - it seems that respondents are indeed conscious about not disturbing not just referents with authority over their current context, and believe that these other people would want them to have their mobile switched to silent.

This understanding of responsibility comes to a conflict with a range of roles under the same contexts, as can be seen in the heatmap of disapproval of switching to silent (10 referents, 13 contexts), shown in Figure 1b. Note, for example, how respondents perceive that their partners, relatives, family and friends mostly disapprove of them having their phone switched to silent in a general (unspecified) context, as well as more sensitive contexts such as work. Respondents also mention some factors that act as mediators to solve such conflicts - it could be tolerated by others to have your device silent, but not after multiple communication attempts, while an important event is occurring, or when others have a genuine reason to be looking for you. Of course, some of these contexts are outside the control of the user. Hence, the previous mentions of anxiety over missing important communications and upsetting other people, are represented here with validity.



(a) When is switching to silent is an approved behaviour?

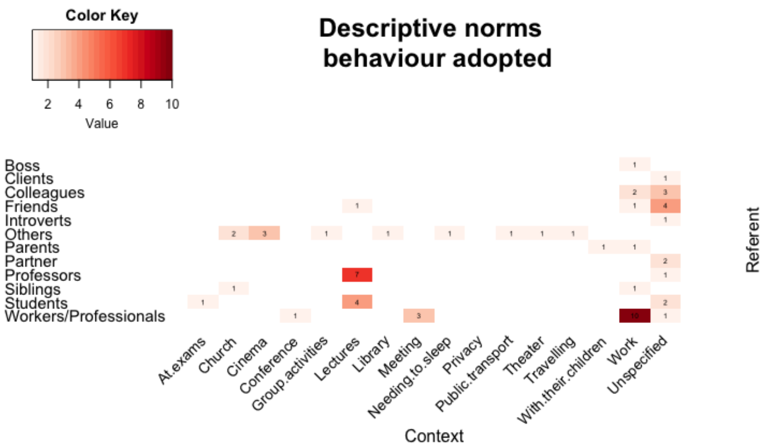


(b) When is switching to silent is a disapproved behaviour?

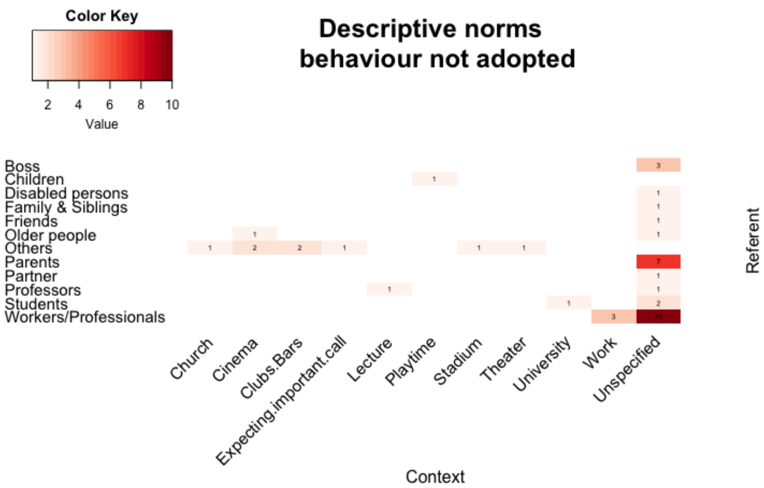
Figure 1: Referents and Contexts influencing perception of switching behaviour appropriateness.

3.2.2. Descriptive norms

Next, we identify the respondents' perception on what other people do or don't do in given contexts with two separate questions. Similar to the injunctive norms for behaviour adoption, we note again that respondents indicate a wide range of roles (12) and contexts (16) under which they assume others are switching their devices to silent (Figure 2a).



(a) When do others switch to silent?



(b) When do others not switch to silent?

Figure 2: Referents and Contexts influencing perception of behaviour adoption by others.

Paired with the injunctive norm findings towards expectations of others that this behaviour will be adopted, a consensus seems to form about appropriateness of switching to silent in specific contexts (e.g., lectures, cinema, at work). Respondents also indicated several referents (12) and contexts (12) where the behaviour is not observed to be adopted (Figure 2b). Based on the textual comments left by respondents, these represent cases where there might be genuine reasons for these people to deviate from the assumed norms (e.g., when someone has a disability, older people, people working in noisy environments, people expecting a call). However, more interestingly, we note that in the answers provided by respondents, many of the comments are rather derogatory of the referents (e.g., *"Clueless people at the cinema", "Idiots who want to show off their mobile", "anti-social people", "indifferent people"*). These findings demonstrate not just the many circumstances where a person's failure to adopt the behaviour can inadvertently place them outside the social norms, but also that there is often a very negative perception when this failure occurs.

3.2.3. Control beliefs

In this section of the survey, we queried respondents on the factors that they believe make it easy or hard for them to switch their phones to silent with two separate questions. In terms of facilitating factors, several respondents (10) mentioned that it is generally easy to achieve the switch (e.g., *"I can do it in seconds", "It's very easy to do"*). Four participants especially commented that switching to silent only requires a few actions (e.g., *"it only takes 2 moves to do this"*). Many respondents also attributed this ease of switching to having physical buttons on the device to control the switch, with 10 mentions of this theme (e.g., *"I use the volume buttons to do this, it's very easy", "It only requires pressing the button on the side of my phone"*). Only three respondents referenced the virtual buttons available for this purpose from the phone's touchscreen. Another important facilitator was proximity to the device (6), for example *"My phone is always next to me", "my phone is usually in my hands or in my pocket"*.

While respondents seem to primarily mention the use of physical buttons to switch to silent, two respondents also mentioned having sometimes difficulties in precisely controlling the switch. Since the device can be set to silent by pressing and holding the volume-down button, the two participants mentioned "overshooting" the target and thereby setting the device to completely silent, when they only intended to put it on vibrate only. In terms of factors that make it hard to switch, the most frequent responses related to "being far away from the device" (6) or that their device is placed out of sight, such as "in a bag" (4), "in a pocket" (1), or "forgotten where I put it" (1). Another important factor is that they often "forget to switch" (4). Further, we note that several respondents feel prevented from switching to silent because of feelings of fear and guilt about missing important events, both when they are indeed expecting one (3) but also even through sheer worry that one might happen (3). Other factors include a difficulty in understanding the effect of the physical or virtual controls on the device ("confusing vibrate and silent mode": 2, "confusing media and ringer volume": 2), the physical size and use of the device ("too big to use with one hand": 1, "hands are usually occupied": 1) and finally the lack of physical buttons to achieve the switch (1).

3.2.4. Discussion

The RAA approach used in our study involves examining the attitudes, injunctive and descriptive normative referents (other people who influence our decisions) as well as control factors in the decision. Our findings about attitudes confirm the findings of previous work, e.g., Chang & Tang [1], Exler et al. [25] and other literature [6–9] relating to the perceived advantages and disadvantages of switching to silent. These attitudes certainly influence the decision process for switching to silent.

The identification of referent-context combination effects on ringer mode switch intentions, is a novel insight that has only been lightly touched upon in past literature (e.g., [1]). Participants were able to identify a wide range of referent and context combinations from their personal experience. We discovered that few referents are globally important, and most are only relevant to the decision-making process under certain contextual parameters. These relate to the space in which the user is located, and the type of activity that the user is engaged in, and which potentially involves these referents. The role of the referent in each activity is also important (e.g., whether they are a peer to the user, or a person of authority over the space, or the activity taking place in the space).

Our study contributes further novel findings relating to control factors, not previously seen in literature. Prior to switching ringer modes, the user has to "remember" whether making such a decision is relevant. Considerate users may keep this question always in mind, but users are often forgetful, or even unaware that making such a decision is pertinent to their circumstances. The debilitating control factors uncovered in our survey, as well as the reported negative affect that can be caused by switching to silent (validating [6–9]), demonstrate that users could benefit from increased automatic support for this important function. As such, we assert that increased support for ringer mode switching behaviour to help users remember, or to *know* that their device should be switched to silent, is an important research direction.

To enable this automated support, we need context awareness relating to the presence of others in the user's vicinity, their attitude towards being interrupted by the user's device, the role of others in a joint activity where the user partakes, the authority of others over the common space (and even on the user), and the users' affective and mental state (e.g., anticipating an important event, or focusing on a task). These high-level contextual abstractions might be difficult, if not impossible, to acquire with present-day sensing technologies. Such attempts are additionally confounded by the difficulty to capture the dynamic properties and semantics of space and users within them.

On the other hand, it is plausible that certain types of venue have regular patterns of ringer mode use under most contexts, and that these patterns can be inferred, despite our inability to capture individual context elements (e.g., presence of others, the user's relationship with others or current mental state). For example, it is less likely that a user might find themselves completely alone in a classroom, and more likely that if they are present in a classroom, then others would also be present and that a person of authority over the space would also be present, leading to a prevalent behaviour of switching to silent by both the user and the co-located individuals.

Hence, the necessary data for such ringer mode switching predictions can be mined, not only through monitoring the user's own behaviour, but by collecting, processing and sharing *spatiotemporal* context information across many users. Given a large pool of monitored users, it can be possible to learn the ringer mode norms that pertain to a range of locations, and to impart this knowledge to users who are currently at, or may even have never visited these locations before. Even more, while information about users' behaviours at a specific location may not be available, we can likely infer the appropriate social norm from knowledge we have about semantically and geographically similar places.

4. Study 2: Empirical data from switching to silent behaviour

To determine whether spatiotemporal context can be used to predict appropriate ringer mode, we used an Experience Sampling Method, and a data logging approach, recruiting participants from the same population target as the qualitative survey (young students and professionals). We developed a simple UI-less Android application (background service), which leverages two core Android API components: the NotificationListener API allows the capture of incoming notifications, and the AudioManager API (RINGER_MODE_CHANGED_ACTION) broadcast event, which is fired whenever the device ringer mode is changed. The application logs data from users and sends it to our remote server, at regular intervals and when the user is connected to Wi-Fi. The ESM application UI is shown in Figure 3.

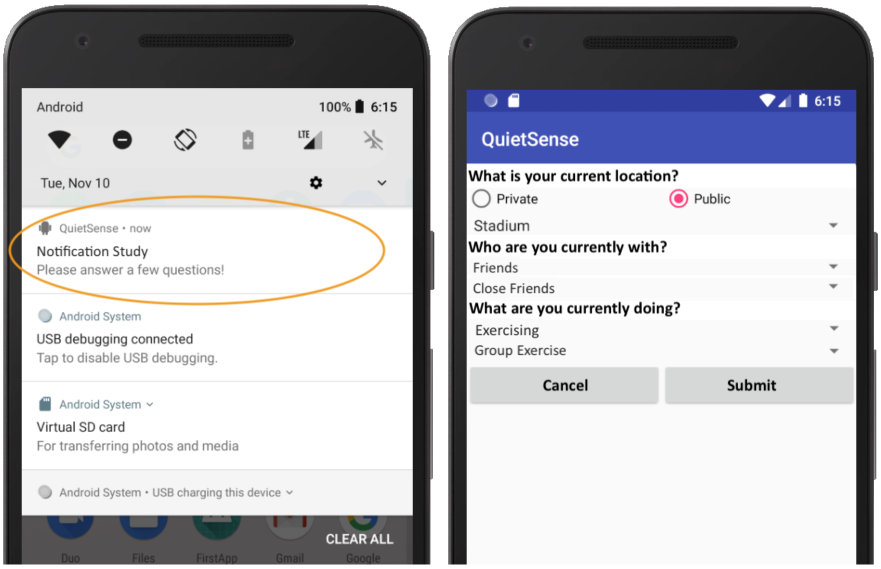
Overall, we recruited 44 participants (26 female, mean age 25.6 years, $\sigma = 3.2$) who were Android smartphone users, and installed the application on their devices. All users self-reported as using their phone daily, and being aware of how to change ringer mode on their device. All users self-reported changing ringer mode at least once a week (8), at least once per day (13) and more than once per day (23). A consent form was signed and participants were instructed that they could quit the study at any time by informing us of this decision and uninstalling the software on their device. The study software was programmed to automatically stop collecting data and uninstall itself after a maximum of 3 months of use, although no participant stayed on for that length of time (average participation duration 22.72 days, $\sigma = 10.392$, max=34, min=6). Participants were requested to leave location services enabled on their device for the duration of the study, although we did not enforce this condition. No incentive for participation was given. Particulars for each type of dataset collected from participants follow in the subsequent sections.

4.1. Results from experience sampling

4.1.1. Data collection

With our application installed, when a ringer mode change is performed by the user, a notification is issued on the device (Figure 3a). Upon clicking (accepting) the notification, a simple questionnaire appears for the user to fill in (Figure 3b). In order to prevent overloading the users with questionnaires, we enforced a condition that a questionnaire notification would not be issued until after at least 4 hours had passed from the previous issuing, and also that questionnaires would not be issued between 23:00 and 07:00. Therefore, a participant would need to respond to at most 4 questionnaires per day.

Participants were asked to provide information about their current location and the presence of other people around them, as well as their current activity. To design the available options, we took inspiration from relevant work describing situational contexts such as Saucier et al. [28] and the taxonomy described by Rauthman et al. [29]. Based on this work, we created our own context granularity taxonomy, using situational contexts reported in the Study 1 survey and enriching them with items from related work (Table 1), including "Other" as a generic response type in each context category. To assist the usability, users were asked to provide answers using drop-



(a) ESM Notification issue after the user changes ringer mode.

(b) ESM Questionnaire options. Available options change according to primary context type category selection as per Table 1.

Figure 3: ESM Application Interface

down menus (Figure 3b), first specifying the context *Category* for the response, and then providing specific details (*Options* in Table 1). For example, for location context, the generic categories were "Public" and "Private", and options included elements such as "My home", "My office", etc. Participants had a free choice of dismissing the notification when it appeared (i.e., ignoring the questionnaire), dismissing the questionnaires without fully completing it (in which case no data was logged), and finally completing and submitting the questionnaire. An example of the data collected by the application is shown in Table 2.

Table 1: Context granularity levels used in ESM application questionnaire

Context	**Category**	**Options**
Location	Public	Café, Church, Cinema, Clubs/Bars, Concert, Lecture / Seminar room, Gym / changing rooms, Hospital, Library, Public transport, Restaurant, School, Stadium, Study room, Theater, University, My workplace (company), Shop, Street, Square, Park, Other

Table 1: (continued from previous page)

Context	Category	Options
	Private	Home, Someone else's home, Doctor's office, Car, My office, Other
Social	Alone	Alone
	Colleagues	Boss, Coworkers, Clients, Other
	Family	Children, Parents, Relatives older than me, Relatives similar age as me, Relatives younger than me, Siblings, Other
	Friends	Close friends, Acquaintances, Other
	Partner	Husband / wife, Boyfriend / girlfriend, Other
	Strangers	Older than me, Similar age as me, Younger than me, Persons of respect
	Other	Animal - Pet, Other
Activity	Studying	Attending a lecture / seminar / classes, Studying, Taking an exam, Other
	Leisure time	Attending a match, Attending a play / performance, Attending group activities, Eating or drinking, Going to sleep, Having important discussion, Relaxing, Spending time together, Watching a movie, Shopping, Walking, Cycling, Other
	Social event	Attending ceremony, Attending important event / celebration, Having important discussion, Visiting, Other
	Exercising	Exercising alone, Group sports, Group exercise, Other
	Relaxing	Hanging out, Watching a movie, Reading, Doing a hobby, Watching TV, Online - surfing, Communicating, Other
	Working	Attending group activities, Having a work meeting, Having important discussion, Working, Taking a break, Other
	Commuting	Travelling on bus, Travelling on train, Travelling on airplane, Travelling by car, Driving, Travelling by ship, Walking, Cycling, Other

Table 1: (continued from previous page)

Context	**Category**	**Options**
	Personal care	Grooming, Getting dressed, Preparing to sleep, Other
	Other	Expecting important call or message, Doing house chores, Waiting, Other
Time	Hour of day	Obtained automatically from system clock

Table 2: Sample of ESM data as recorded on participant devices. The various nominal variables were stored as coded values. In this example, the values represent Handling: *Questionnaire Filled*, Ringer mode: *Normal*, Location: *Private*, Location Type: *Home*, Social: *Family*, Social Type: *Parents*, Activity: *Relaxing*, Activity Type: *Reading*

User ID	**Event ID**	**Time Posted**	**Handling**	**Time Handled**	**Ringer Mode**	**Location**	**Loc. Type**	**Social**	**Soc. Type**	**Activity**	**Act. Type**
U1	7	03 Jan 2019 15:24:00	2	03 Jan 2019 15:24:00	2	1	0	2	1	4	2

4.1.2. Ringer mode changes and user context

Overall, we recorded 5,743 ringer mode changes performed on the participant devices. Despite all participants mentioning at the start of the study that they perform ringer mode switches at least once a week, our data indicated that only 15 participants actually performed the behaviour, and did so multiple times per day (μ = 5.116 ringer switches per day, σ =6.530). The ESM prompting restrictions resulted in prompting users with questionnaires for approximately 13% of these switches (726 cases total, $\mu = 45.375$ prompts per participant, $\sigma = 26.263$), spanning a total time period of 109 days (first to last data prompt response). The majority of ESM prompts was dismissed, participants gave up the completion of 15 prompts, and we ended up with 155 completed questionnaires from 15 users ($\mu = 10.333$ completed per participant, $\sigma = 8.715$). We note that the responses in these relate to 82 switches to silent mode and 73 to normal, hence we consider the dataset to be reasonably balanced.

To begin the analysis, we transformed the available data to contain binary values where appropriate and also by one-hot encoding the multinomial features, in order to best represent the nominal nature of responses and to enable the use of classifiers that do not accept nominal variables. To begin with, we consider the binary variable (true-false) describing if the current device ringer mode is silent or not, a binary variable for location type (public or not), the hour of day as an ordinal variable, and one-hot encoding representations activity and social context at the *category* granularity level (Table 1). We did not use the *options* granularity level to create feature vectors, since this would result in very sparse sets due to the small variation in responses. We note that the distribution of responses at public and private type locations is reasonably mixed (72 private, 83 public). The majority (8) of activity categories are represented in the set (Relaxing:45, Studying:29, Working:25, Leisure:21, Commuting:16, Social event:9, Other:7, Exercising:3). All 7 social category types are represented (Alone:56, Other:8, Colleagues:35, Family:13, Friends:19, Strangers:10, Partner:14)

As a first step, we perform a correlation analysis on the resulting dataset. Since we are dealing with binary variables, the analysis reports the ϕ coefficient. Sensitivity power analysis shows that our sample size is enough to detect two-tailed correlations of 0.28 with an error probability level $\alpha = 0.05$. In the resulting analysis (Figure 4 and Table 3), we note that several contextual factors seem to be correlated with the state of the device being switched to silent or normal mode. Factors associated with the device set to normal mode include activity types (resting, leisure time), being alone or in the presence of family members, and with the hour of day (later in the day results in more likelihood of the device being set to normal mode). On the other hand, the device being set to silent is significantly correlated with some other activity types (work, study), the presence of strangers, colleagues and "other" types, but most strongly with the location being a public place. These results add to the findings by Chang & Tang [1], who investigated only location type and hour of day, however our analysis provides some different results in association with time of day and location type. Perhaps this is owed to the different populations used in our studies.

These encouraging results prompt us to model the users' reported context as a predictor for

Table 3: Statistically significant correlations of context and ringer mode, ordered by strength of correlation. Negative correlation implies a setting of ”normal” ringer mode.

Normal mode		**Silent mode**	
Context	ϕ	Context	ϕ
Hour of day	-0.416***	Location.Public	0.443***
Activity.Resting	-0.365***	Social.Colleagues	0.397***
Social.Alone	-0.313***	Activity.Study	0.287***
Social.Family	-0.199*	Activity.Work	0.238***
Activity.Leisure.time	-0.193*	Social.Other	0.195*
		Social.Strangers	0.160*

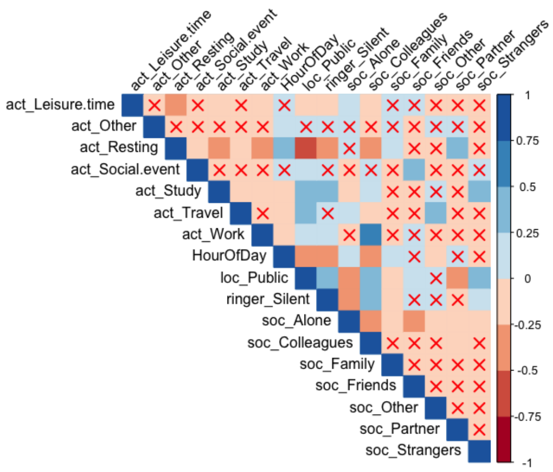


Figure 4: Correlation analysis of contextual factors and silent mode. Statistically non-significant correlations (p¿0.05) are crossed out.

ringer mode, using various machine-learning and statistical classifiers and evaluating the model performance with a $k - fold$ cross validation ($k = 10$). While our dataset is not large, simpler classifiers such as kNN, decision trees and SVM have been demonstrated to be of practical importance in the biomedical or industry domains, where small datasets are frequently encountered (e.g., [30]). Nevertheless, we present the results with some reservation. Initially, we employ a logistic regression model, achieving a respectable average metric of accuracy (appropriate since the dataset is balanced) of 71.00% ($\sigma = 13.38\%$). Further modelling using empirically derived parameters was done with decision trees (max.depth=25, minimal gain=10^{-4}, minimal leaf size=2, minimal split size=2) producing a comparable result with an average accuracy of 72.83% ($\sigma = 11.72\%$). Using ada-boosting on the decision tree model has a marginal effect on accuracy (72.21% $\sigma = 11.39\%$). Using a kNN classifier with various values for $k \in [5, 30]$ with a step of 5, we obtain best results with $k = 25$ for an accuracy of 75.54% $\sigma = 11.64\%$ (see Figure 5). Finally, using SVM classifiers we find best performance with the ANOVA kernel at 74.83% ($\sigma = 12.04\%$).

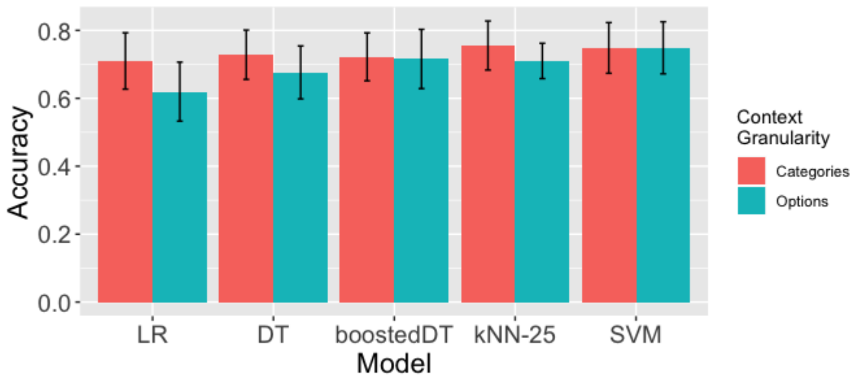


Figure 5: k-fold cross validation results with various ML algorithms, predicting ringer mode from context factors at the Category and Options granularity level (Table 1), error bars at 95%c.i.

Next, we also attempted to model context using the *options* granularity level (Table 1), rather than the context categories. As such, our dataset included a total of 52 features, of which one was a binary variable (true-false) describing if the current device ringer mode is silent or not, 13 features for Location, 21 for Activity, 16 for Social 16 as a result of the one-hot encoding all the specific context options present in the dataset (i.e. not every option in the questionnaire was represented), and one was ordinal (hour of day). We repeated the analysis in the same way as with context and location categories only. The accuracy for most algorithms declined, however, as shown in Figure 5, boosted trees and SVM maintained the same levels of performance. The overall performance decline is expected given the distribution of the data across more features (resulting in sparse feature vectors) and the larger number of possible prediction labels. In multi-label classification problems, a smaller amount of labels leads to better predictive performance, hence the evolution of hierarchial multi-label classifiers (e.g., see [31]). In our case, the number of labels is not huge (e.g., as could be in other problem domains) hence a flat label classification approach is appropriate.

Therefore, worse performance using the “options” labels is generally expected. On the other hand, we note the lack of performance difference for the boostedDT and SVM, but this is explainable since boosted trees and SVM are able to handle sparse data quite well [32, 33].

4.2. Results from automatic spatiotemporal context capture

ESM-type methods, as the one reported on previously, often meet the indifference of participants, or are able to collect limited amounts of information from each participant so as not to annoy them. As such our resulting data set was not especially large. However, in parallel with the ESM capture, our application automatically captured participant spatiotemporal context in a more unobtrusive manner, meaning that context can be captured more frequently. Obviously, such an approach cannot reliably capture rich information such as the participants’ current activity or social context, but as per the results of Study 1, we posited that much of this information might be implicitly embedded in the semantics of spatial data (i.e., user location context). In this section, we present findings from the analysis of a much larger dataset of approximately 57k spatiotemporal context samples captured automatically from all 44 participants.

4.2.1. Capturing spatiotemporal context

Regular frequent sampling of user locations (e.g., every minute), has privacy implications and can heavily drain a user’s smartphone battery. It also would result in significant volumes of data that is redundant, as users tend to stay at the same location for extended periods. To overcome these concerns, we thought to leverage the fact that user smartphones receive tens, or even hundreds of notifications per day, and that these events can be exposed programmatically through the Android APIs. When a notification is received, the device ”wakes up” from sleep, meaning that all sensors and network connectivity can become available at the time, and that battery is already being expended for other purposes by the user. Hence we considered ”piggybacking” our data collection on the incoming notification events, to capture the current user’s ringer mode and their location. For the latter, we did not log user coordinates. Instead, we obtained them from the device, and used them to query the Google Places API, in order to receive the participants’ current venue details, including name and assigned semantics (categories). In total, we captured over 170k notification events from all 44 participants. The Places API does not return a result when the user is not online, hence the samples for which we were able to obtain spatial context were 57,737.

Submitted coordinates are resolved to a venue by the Places API, with a returned confidence $\in [0, 1]$. A reported venue can belong to multiple categories. These are provided in a non-ordered list by Google, ostensibly therefore the order of appearance shows the prevalence of a category type (e.g. "Bar, Restaurant, Cafe" shows that a place is primarily of type "Bar", but also functions as a restaurant and cafe). We therefore extract and keep the primary category of a venue. In doing so, we observed that many places included the vague category "Point of Interest". Hence, where this was the primary category, it was replaced by the immediately subsequent category type.

Another note here relates to Google's list of categories, where 127 different categories are listed. Predicting on 127 category classes is possible, but presents an unnecessary complexity to the problem, as many venue categories are quite similar in nature and it can be expected that a user will exhibit similar behavioural patterns in these. For example, "Church" and "Mosque" are both places of worship, where user attitude towards ringer mode is expectedly similar. We therefore attempted to group the individual categories into larger sets, as per Table 4. Ultimately, we assigned to each place the super-category to which it belongs, based on its primary category type. An exception to this were the "Miscellaneous" and "Entertainment areas" categories, since for these the user behaviour might be quite different depending on conditions (e.g., a user probably can't notice a notification in a night club as easily as in a cafe), hence for these we used the primary categories ungrouped. As a result, we find that the user notifications were issued at 23 distinct place categories.

Summarizing the data capture process, at the time of each notification we capture raw sensor and device data, and convert it, on the user's device, into a storage form for further processing, as shown in Table 5.

Table 4: Grouped place categories from the Google Places API

Category group	**Categories**	**Samples**
Accomodation	Campground, Lodging, Room, Rv Park	1,350

Table 4: (continued from previous page)

Category group	Categories	Samples
Address	Administrative Area Level 1, Administrative Area Level 2, Administrative Area Level 3, Country, Geocode, Locality, Political, Post Box, Postal Code, Postal Code Prefix, Postal Town, Street Address, Sublocality, Sublocality Level 1, Sublocality Level 2, Sublocality Level 3, Sublocality Level 4, Sublocality Level 5, Synthetic Geocode	86
Civil Services	City Hall, Courthouse, Embassy, Fire Station, Local Government Office, Police, Post Office	89
Contractors	Electrician, General Contractor, Moving Company, Painter, Plumber, Roofing Contractor	76
Education	Library, School, University	11,996
Entertainment Areas	Amusement Park, Aquarium, Bar, Bowling Alley, Cafe, Casino, Gym, Movie Theater, Museum, Night Club, Restaurant, Stadium, Zoo	11,157
Financial Services	Bank, Atm, Finance	93
Healthcare	Dentist, Doctor, Health, Hospital, Physiotherapist	617
Miscellaneous	Establishment, Floor, Other, Point Of Interest, Premise, Subpremise	18,347
Outdoor Areas	Colloquial Area, Natural Feature, Neighborhood, Park, Parking, Route	516
Personal Care	Beauty Salon, Hair Care, Spa	1,104
Place Of Worship	Cemetery, Church, Hindu Temple, Mosque, Place Of Worship, Synagogue	758

Table 4: (continued from previous page)

Category group	Categories	Samples
Professional Services	Lawyer, Accounting, Car Dealer, Car Rental, Car Repair, Car Wash, Funeral Home, Insurance Agency, Laundry, Locksmith, Real Estate Agency, Storage, Travel Agency, Veterinary Care	659
Public Transport	Airport, Bus Station, Intersection, Subway Station, Taxi Stand, Train Station, Transit Station	580
Shopping	Art Gallery, Bakery, Bicycle Store, Book Store, Clothing Store, Convenience Store, Department Store, Electronics Store, Florist, Food, Furniture Store, Gas Station, Grocery Or Supermarket, Hardware Store, Home Goods Store, Jewelry Store, Liquor Store, Meal Delivery, Meal Takeaway, Movie Rental, Pet Store, Pharmacy, Shoe Store, Shopping Mall, Store	10,309

Raw Feature	Captured from	Converted and stored as
System Time	Device clock	Hour of day (e.g 14)
Ringer Mode	Device settings	Silent (true/false)
Coordinates	Device GPS or Network Location Provider	Place category (e.g. "Shopping")
Coordinates	Device GPS or Network Location Provider	Place confidence $\in [0, 1]$

Table 5: Data captured from automatic spatiotemporal logging

4.2.2. Predicting ringer mode from spatiotemporal context

Next, we use a similar data preparation process as in the previous analyses, selecting the features of ringer mode, hour of day and location type (this time one-hot encoding all features including hour). An example is shown in Table 6.

ringer.silent	**hour.0**	**hour.1**	*[...]*	**hour23**	**loc.accom**	*[...]*	**loc.shop**	**conf**
0	0	0	...	1	0	...	1	0.30
1	1	0	...	0	1	...	0	0.70

Table 6: Two examples of one-hot encoded samples after data pre-processing. The first example is a device on "normal mode", the user is at some "shop" with a 30% confidence and the time is between 23:00-23:59. The second is a device on "silent mode", the user is at some "accommodation" with 70% confidence, and the time is between 00:00-00:59.

In this dataset, we find a bias towards normal ringer mode (68.79% of samples), in line with self-reported use in Chang & Tang [1]. Figure 6 shows the 20 most frequently represented venues in our dataset, and how the prevalence of ringer mode use changes at these venues during the day. This demonstrates that the temporal dimension is important to consider and confirms the variability also discovered by Exler et al. [25]. We observe that some venue types exhibit patterns that appear to be intuitively correct, for example, V15 and V16 are our department's teaching room building and researcher office building respectively (hence the difference is normal). On the other hand, venues V7 and V8 belonging to the same category (Establishment) show very different patterns. This can be due to misclassification by Google, or due to the fact that the category is a very broad descriptor (in this case, V7 is an art studio, and V8 is a construction company). Assessing problems with Google's misclassification of data is an entirely different topic, so we proceed assuming that overall there is generally acceptable level of quality in this data.

We employed Tensorflow-based modelling using a lightweight deep learning 3-layer sequential model (*L1: 128 units, RELU activation, dropout 0.4; L2: 64 units, RELU activation, dropout 0.3; L3: 1 unit, sigmoid activation; loss function: binary cross entropy; optimizer: ADAM*). Initial exploration of the hyperparameters, using an 80/20 split of the dataset and a 30% of the training dataset as a hold-out validation set, resulted in an optimal batch size of 200 with 12 training epochs, providing quick model-building speed and without loss of accuracy. With these settings, we perform a k-fold cross validation ($k = 10$). The resulting accuracy is 70.06% ($\sigma = 0.49\%$).

Since the Places API reports venues with a varying degree of confidence, we wanted to explore how filtering out venues at various confidence thresholds might affect results (Figure 7). Our assumption here is that the inclusion of venues where confidence is high, might produce the best

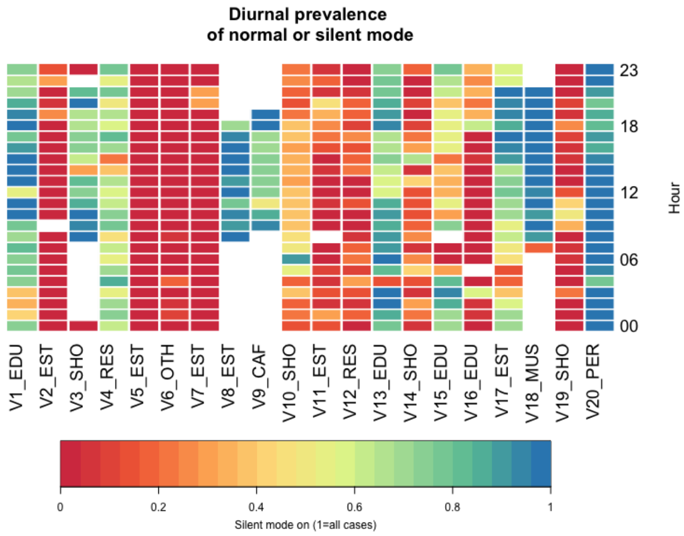


Figure 6: Diurnal variation of ringer mode prevalence for 20 venues most frequently represented in our dataset (confidence $\geq$ 0.7). Each colour block shows the average hourly ringer mode settings captured at that location (0=normal, 1=silent). A value of 0.5 indicates split prevalence. Venue categories are: EDU(cation), EST(ablishment), SHO(pping), RES(taurant), OTH(er), CAF(e), MUS(eum), PER(sonal care).

possible results in the training/testing process. Indeed, we observe best performance with venues at > 0.9 (73.01%). We note, however, that the performance curve does not trend upwards, but appears to drop after excluding venues with a confidence ≤ 0.5, which seems counter to our initial assumption. One plausible explanation might be to attribute this observation to the shrinking of the training dataset - a very small dataset would significantly affect the model's training and thus its predictive performance. However, further investigation reveals this not to be the case. We observe in Figure 8 that the dataset at confidence level $\geq$0.6 is 16,336 samples (28.3% of the whole), which is reasonably large, and that size reduction is almost linear.

Therefore, we wondered if this effect is due to another reason, and considered that earlier during analysis, we observed that not all venue categories are equally represented in our dataset (Table 4). We wondered thus if the distribution of categories changes much, when confidence filters are applied to the dataset, and if this might be having an effect on the model training and predictive ability. Indeed, this change in category distribution can be seen in Table 7, which shows the top-10 venue categories across the entire dataset, and when a confidence threshold of ≥ 0.7 is applied. Because the presence of rare categories might be diluting the performance, we repeated the analysis

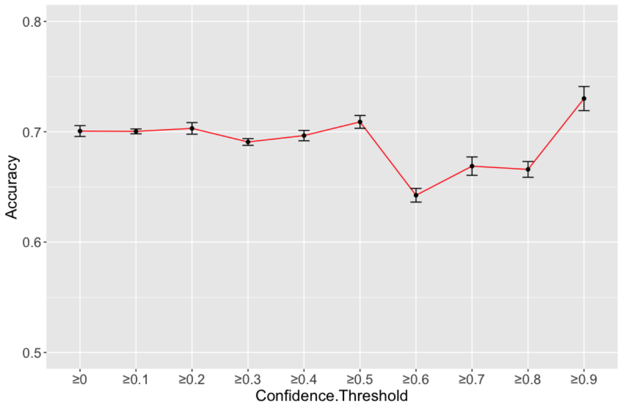


Figure 7: k-fold cross validation results at various levels of venue confidence (error bars at 95%c.i.)

Table 7: Distribution of top 10 most frequent venue types

	Entire dataset		**Confidence** ≥ 0.7	
Rank	Venue Category	Proportion	Venue Category	Proportion
1	Establishment	27.63%	Education	37.57%
2	Education	20.78%	Other	14.94%
3	Shopping	17.86%	Cafe	13.28%
4	Restaurant	8.21%	Shopping	9.72%
5	Cafe	6.87%	Restaurant	9.34%
6	Other	4.14%	Establishment	8.58%
7	Accomodation	2.34%	Gym	1.53%
8	Personal Care	1.91%	Accomodation	1.43%
9	Place Of Worship	1.31%	Outdoor Areas	0.64%
10	Gym	1.20%	Address	0.63%

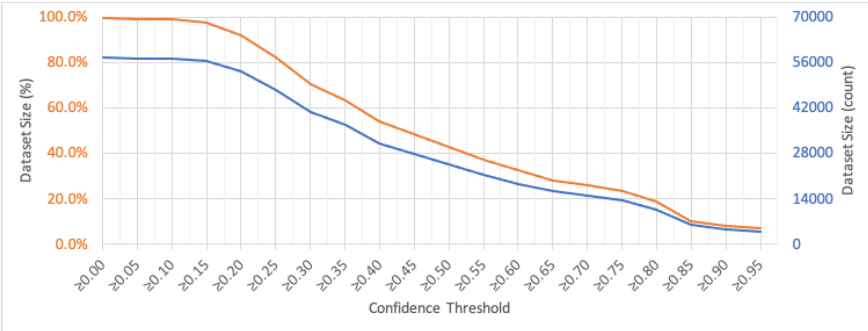


Figure 8: Dataset size after applying confidence filters (orange = percentage, blue = count)

using only the top-10 categories present in the whole dataset and also in the subsets derived at confidence thresholds ≥ 0.6 and above. The results (Table 8) demonstrate significant improvements in all cases, which means that the model benefited from training on the most prevalent categories, compensating for the reduction in the size of the training set. A final observation relates to the confusion matrices produced by the predictive process (an example is shown in Table 9, patterns for other threshold values are similar). Examining the precision of the predictions, we note that this remains very similar for predicting when the device should be in silent mode (72.53% vs. 72.74%), but we obtain a large increase in precision when only locations for which we are highly confident are considered (54.58% vs. 68.18%).

Table 8: Result comparison using all categories and using *Top-10* only

	All categories		**Top-10 only**	
Conf. Level	Accuracy	SD	Accuracy	SD
All	70.06%	0.49%	69.74%	0.82%
≥ 0.6	64.25%	0.62%	71.60%	0.79%
≥ 0.7	66.88%	0.83%	73.58%	1.24%
≥ 0.8	66.59%	0.71%	72.77%	1.89%
≥ 0.9	73.01%	1.09%	81.28%	2.39%

These results demonstrate that a predictive accuracy with a limit of $\approx 80\%$ is attainable using temporal and location context, given reasonably reliable data for the latter and performing predictions only on the top-N prevalent categories. As such, we must attribute any losses in accuracy

Table 9: Confusion matrices for predictions using the top-10 categories

Entire dataset			**Confidence** ≥ 0.7		
	Truth			Truth	
Prediction	0	1	Prediction	0	1
0	3306	1252	0	771	289
1	347	421	1	84	180

to the inability to capture other context types, such as user activity and social context. However, notably, we note that the achieved accuracy implies that location context plays an important role in the decision to switch to silent, perhaps because location semantics can work as a proxy to the other types of context (social, activity) that can be strongly tied to specific locations. Compared to the only other directly related study with a goal to predict ringer mode [2], our approach is less precise (up to 80%, compared to 90%), however [2] employ a simple training-validation approach which could be leading to model overfit, and cover only a small amount of contexts (4) compared to our study.

5. Discussion, Limitations and Future Work

Mobile users have their own habits of managing ringer mode switches on a daily basis. As seen in previous literature, switching to silent modes is not only culturally mandated in certain contexts, but voluntary switches can assist human productivity and concentration by reducing interruptions. On the other hand, prolonged periods of silent mode use can result in higher levels of anxiety, discomfort and negative consequences for users. A successful balance in managing ringer mode switches is therefore highly desirable in daily life. Even though there is no related literature on reasons for failure to switch ringer modes at appropriate contexts, from our own experience we posited that it is not always possible for a user to be aware of social etiquette regarding ringer mode use, and a range of control factors such as forgetfulness or physical distance from the device might hinder their ability for timely ringer mode switches. Although some support towards automated management of ringer mode switches is offered by third party apps or mobile operating systems, these require the manual specifications of operating rules, a process which is subject to significant trial-and-error in order to get it to work, or might be dauntingly difficult for several user groups. In any case, creating rules to cover every possible context is unlikely to be possible for any user. In our paper, we sought to understand the driving factors behind ringer mode switching (Study 1), and further, to examine whether it might be possible to predict appropriate ringer modes depending on spatiotemporal context (Study 2), so that more intelligent ringer mode switching support can be offered to users.

Is the management of ringer mode a problem worth addressing? In our work, we found that all participants indicated many of the perceived benefits in switching ringer modes that are also found in previous literature (Section 3.1.1). However, we noted that even though all participants self-reported switching ringer modes at least once a week, in reality many didn't do this at all, at least not for the duration of our study (Section 4.1.2). It is reasonable to expect that some of this behaviour (i.e., not switching) can be attributed to the impact of control factors that we uncovered in Study 1 (Section 3.2.3). Therefore, an automatic application to assist ringer mode switches, could remove some of the barriers towards control and likely be useful in users' lives.

In the rest of this section, we summarise our main contributions and limitations, starting with Study 1. In this, we contribute to the limited understanding of ringer mode switching behaviour, using, for the first time, a structured theoretical framework to elicit and understand behaviour. This has helped to both confirm previous findings, and to uncover novel insights. Our work uncovered strong user concerns for the avoidance of disruption to themselves, confirming some of the findings in previous work (Chang & Tang [1] and Exler et al. [25]). We also discovered strong concerns towards the wellbeing of co-located people, a novel finding not reported in previous literature such as [12]. Further, we contribute novel insights about the roles of referents under context. The "respect towards others" depends on the authority these persons over the current space, and on the individual or group activities taking place in it. In this sense, we bring some bad news for anyone hoping to capture social context - it seems not enough to know what the user is doing, but we must know both who else is near the user, and what *they* are (or might be) doing too.

In Study 1 we also discovered novel insights, relating to control factors placed in user habits (e.g., always leaving my phone on silent, always having my phone in a bag) and usability problems (e.g. touchscreen controls, physical button "overshooting", mental models). These open up new directions for future research. Control and attitude beliefs were also mediated by several cases of affect, not just of the user, but of co-located people, during the decision making process (e.g., fear of becoming isolated, upsetting others who are expecting the user to be reachable, worrying about missing important notifications etc.), validating previous work [6–9]. Affect is unfortunately not part of the RAA approach we used. Other frameworks might be more appropriate for future use (e.g., the Theory of Interpersonal Behaviour framework also includes roles, affect and habit [34]).

Our survey in Study 1 is limited by its narrow definition of the population target and its sample size (30), as are the core related qualitative studies in our literature review (e.g., 28 in [1] and 40 in [25]). Both these studies and ours targeted similarly aged and educated population groups, but we found both commonality and differences among the results. This is not unexpected, given the cultural diversity that exists across multiple levels of community (local, regional, national and international). There is a clear need thus, to expand the research across multiple countries and communities, to compare the global North-South and East-West cultural discrepancies.

Our contribution in Study 2 was to investigate empirical evidence linking spatial and temporal context with ringer mode changes, based on our observations in Study 1. Starting with an analysis of ESM data, we demonstrate the important correlations between real contexts and ringer mode use. Chang & Tang [1] demonstrated some relationships between ringer mode, time and user-tagged location semantics and our study adds to these findings with richer breakdowns of context types, including activity and social. In contrast with their findings, we find silent mode to be positively correlated to earlier time and public locations, demonstrating that different populations might not share the same preferences. We also carried out predictive modelling for ringer mode use based on ESM data, but report the findings with some reservation, due to the dataset size.

Next, we analysed spatial and temporal context logged automatically (57k observations), adding to the work of Exler et al. [25], who attempted the same but with a qualitative study. We confirmed their findings about the variability of predominant ringer mode use according to location type, and found that these two context factors account for much of the observed variation. Compared to the only other directly related study with a goal to predict ringer mode [2], our modelling was less accurate (up to 80%, compared to 90%), though we employ a more generalisable approach and cover many more possible contexts. We demonstrated that certain venues display variable ringer mode appropriateness patterns through the day, however, more analysis is needed, including examining more temporal aggregates (e.g., day of week, month). Future work could include the addition of sensor data as in [2] to increase performance, particularly when considering the presence of others (e.g. via Bluetooth scan or audio analysis).

Our paper has two limitations that stem from the implicit assumptions in our analysis of the automatic context capture, with regard to ground truth. First, we use Google Places API as the ground truth for location semantics. We noticed that when venues are reported with a high confidence, we attained good predictive performance. However, a large number of venues is classed by the Places API under generic categories (e.g., Establishment) or is misclassified, as we saw browsing manually through the dataset (e.g. a university faculty, classed as "school"). A better approach might be to cross-validate venue categories across several APIs such as Yelp or Foursquare. The second aspect is that we treat all ringer mode captures as correctly labelled examples, for the purposes of training our ML models. Given the findings of Study 1, it is possible that several examples in our dataset are mislabelled, as they might include a ringer mode which is not the one the user would have actually preferred. Given the size of the dataset though, we believe that such cases are mostly "noise" and that the ML algorithm can train around these issues, however, it is likely that the accuracies we report here could be improved if there was more confidence in the correct labelling of the dataset. In future work we would also like to examine if accuracy can be improved by extending our context capture to include activity detection (e.g., Google's Activity API such as in [35]) and incorporating novel approaches to detecting social context, such as in [36].

6. Conclusion

The first goal of our paper was to obtain insights about the contextual factors that drive users to switch ringer modes. To the best of our knowledge, this is the first paper to present such findings using a structured human behaviour theory as a framework to discover, and explain the role of such factors. Although it might be very hard for smartphones to capture or infer all the related factors, we find that the barriers in ringer mode switching behaviour can be attributed to control factors which can likely be mitigated by automatic inference means. More importantly, we found that the activity and role of users and other co-located persons play are significant determinants in ringer mode switching, and that these roles and activities are often deeply embedded into location semantics, which can be inferred by smartphones.

Our second goal was to take a first step towards determining whether the autonomous learning of these factors can be achieved. First, we empirically validated the findings of our survey, through experience sampling on user smartphones. Further, we explored the finding of activity and role embeddedness into location semantics, and attempted to infer appropriate ringer mode using machine learning, based solely on spatial and temporal context features. We found that prediction accuracy is affected by the degree of confidence regarding the user's current spatial context, though it can reach upwards of 80% for adequately represented venue categories.

More data from various countries, paired with our qualitative study, can begin to help us unfold the black-box process that is currently the decision to switch to silent. We end the paper with a call to other researchers who might be willing to embark on a multi-national data collection, to get in touch with us.

7. Funding and data availability

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Our empirical data are openly available at https://github.com/komis1/ami2019-notifications

References

[1] Y.-J. Chang, 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, New York, NY, USA, 2015, pp. 6–15. doi:10.1145/2785830.2785852.

[2] Y. Qin, T. Bhattacharya, L. Kulik, J. Bailey, A context-aware do-not-disturb service for mobile devices, in: Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia, MUM '14, Association for Computing Machinery, Melbourne, Victoria, Australia, 2014, pp. 236–239. doi:10.1145/2677972.2678003.

[3] A. Exler, C. Dinse, Z. Günes, N. Hammoud, S. Mattes, 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 Computers, UbiComp '17, ACM, New York, NY, USA, 2017, pp. 970–976. doi:10.1145/3123024.3124560.

[4] A. Voit, D. Weber, N. Henze, Qualitative Investigation of Multi-Device Notifications, in: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, UbiComp '18, Association for Computing Machinery, Singapore, Singapore, 2018, pp. 1263–1270. doi:10.1145/3267305.3274117.

[5] A. Komninos, E. Frengkou, J. Garofalakis, Predicting User Responsiveness to Smartphone Notifications for Edge Computing, in: A. Kameas, K. Stathis (Eds.), Ambient Intelligence, Lecture Notes in Computer Science, Springer International Publishing, 2018, pp. 3–19.

[6] C. Stothart, A. Mitchum, C. Yehnert, The attentional cost of receiving a cell phone notification, Journal of Experimental Psychology. Human Perception and Performance 41 (4) (2015) 893–897. doi:10.1037/xhp0000100.

[7] K. Kushlev, J. Proulx, E. W. Dunn, "Silence Your Phones": Smartphone Notifications Increase Inattention and Hyperactivity Symptoms, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, Association for Computing Machinery, San Jose, California, USA, 2016, pp. 1011–1020. doi:10.1145/2858036.2858359.

[8] J. Aranda, N. Ali-Hasan, S. Baig, I'm just trying to survive: An ethnographic look at mobile notifications and attention management, in: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, MobileHCI '16, Association for Computing Machinery, Florence, Italy, 2016, pp. 564–574. doi:10.1145/2957265.2957274.

[9] M. Pielot, L. Rello, Productive, anxious, lonely: 24 hours without push notifications, in: Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '17, Association for Computing Machinery, Vienna, Austria, 2017, pp. 1–11. doi:10.1145/3098279.3098526.

[10] N. Ali-Hasan, R. Garb, M. Pereira, Designing Android Marshmallow volume controls: A user experience case study, in: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, MobileHCI '16, Association for Computing Machinery, Florence, Italy, 2016, pp. 557–563. doi:10.1145/2957265.2957273.

[11] R. d. A. Maués, S. D. J. Barbosa, Keep doing what i just did: Automating smartphones by demonstration, in: Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '13, Association for Computing Machinery, Munich, Germany, 2013, pp. 295–303. doi:10.1145/2493190.2493216.

[12] C. Anderson, I. Hübener, A.-K. Seipp, S. Ohly, K. David, V. Pejovic, A Survey of Attention Management Systems in Ubiquitous Computing Environments, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2 (2) (2018) 58:1–58:27. doi:10.1145/3214261.

[13] R. Avraham Bahir, Y. Parmet, N. Tractinsky, Effects of Visual Enhancements and Delivery Time on Receptivity of Mobile Push Notifications, in: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, CHI EA '19, ACM, Glasgow, Scotland Uk, 2019, pp. LBW0288:1–LBW0288:6. doi:10.1145/3290607.3312993.

[14] L. D. Turner, S. M. Allen, R. M. Whitaker, Reachable but not receptive: Enhancing smartphone interruptibility prediction by modelling the extent of user engagement with notifications, Pervasive and Mobile Computing 40 (2017) 480–494. doi:10.1016/j.pmcj.2017.01.011.

[15] A. Visuri, N. van Berkel, T. Okoshi, J. Goncalves, V. Kostakos, Understanding smartphone notifications' user interactions and content importance, International Journal of Human-Computer Studies 128 (2019) 72–85. doi:10.1016/j.ijhcs.2019.03.001.

[16] Y.-J. Chang, Y.-J. Chung, Y.-H. Shih, I Think It's Her: Investigating Smartphone Users' Speculation about

Phone Notifications and Its Influence on Attendance, in: Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '19, Association for Computing Machinery, Taipei, Taiwan, 2019, pp. 1–13. doi:10.1145/3338286.3340125.

[17] T. Okoshi, K. Tsubouchi, M. Taji, T. Ichikawa, H. Tokuda, Attention and engagement-awareness in the wild: A large-scale study with adaptive notifications, in: 2017 IEEE International Conference on Pervasive Computing and Communications (PerCom), 2017, pp. 100–110. doi:10.1109/PERCOM.2017.7917856.

[18] B. Poppinga, W. Heuten, S. Boll, Sensor-Based Identification of Opportune Moments for Triggering Notifications, IEEE Pervasive Computing 13 (1) (2014) 22–29. doi:10.1109/MPRV.2014.15.

[19] P. Saikia, M. Cheung, J. She, S. Park, Effectiveness of Mobile Notification Delivery, in: 2017 18th IEEE International Conference on Mobile Data Management (MDM), 2017, pp. 21–29. doi:10.1109/MDM.2017.14.

[20] L. D. Turner, S. M. Allen, R. M. Whitaker, Push or Delay? Decomposing Smartphone Notification Response Behaviour, in: Human Behavior Understanding, Lecture Notes in Computer Science, Springer, Cham, 2015, pp. 69–83. doi:10.1007/978-3-319-24195-1_6.

[21] M. Pielot, B. Cardoso, K. Katevas, J. Serrà, A. Matic, N. Oliver, Beyond Interruptibility: Predicting Opportune Moments to Engage Mobile Phone Users, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1 (3) (2017) 91:1–91:25. doi:10.1145/3130956.

[22] A. Komninos, J. Besharat, V. Stefanis, G. Gogoulou, J. Garofalakis, Assessing the perceptibility of smartphone notifications in smart lighting spaces, Journal of Ambient Intelligence and Smart Environments 11 (3) (2019) 277–297. doi:10.3233/AIS-190525.

[23] B. Saket, C. Prasojo, Y. Huang, S. Zhao, Designing an Effective Vibration-based Notification Interface for Mobile Phones, in: Proceedings of the 2013 Conference on Computer Supported Cooperative Work, CSCW '13, ACM, New York, NY, USA, 2013, pp. 149–1504. doi:10.1145/2441776.2441946.

[24] D. Siewiorek, A. Smailagic, J. Furukawa, A. Krause, N. Moraveji, K. Reiger, J. Shaffer, F. L. Wong, SenSay: A context-aware mobile phone, in: Seventh IEEE International Symposium on Wearable Computers, 2003. Proceedings., 2003, pp. 248–249. doi:10.1109/ISWC.2003.1241422.

[25] A. Exler, Z. Günes, M. Beigl, Preferred notification modalities depending on the location and the location-based activity, in: Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers, UbiComp/ISWC '19 Adjunct, Association for Computing Machinery, London, United Kingdom, 2019, pp. 1064–1069. doi:10.1145/3341162.3344842.

[26] M. Fishbein, I. Ajzen, Predicting and Changing Behavior: The Reasoned Action Approach, Taylor & Francis, 2011.

[27] I. Ajzen, Constructing a theory of planned behavior questionnaire (jan 2006).
URL https://people.umass.edu/~aizen/pdf/tpb.measurement.pdf

[28] G. Saucier, T. Bel-Bahar, C. Fernandez, What Modifies the Expression of Personality Tendencies? Defining Basic Domains of Situation Variables, Journal of Personality 75 (3) (2007) 479–504. doi:10.1111/j.1467-6494.2007.00446.x.

[29] J. F. Rauthmann, D. Gallardo-Pujol, E. M. Guillaume, E. Todd, C. S. Nave, R. A. Sherman, M. Ziegler,

A. B. Jones, D. C. Funder, The Situational Eight DIAMONDS: A taxonomy of major dimensions of situation characteristics, Journal of Personality and Social Psychology 107 (4) (2014) 677–718. doi:10.1037/a0037250.

[30] I. Martin-Diaz, D. Morinigo-Sotelo, O. Duque-Perez, R. J. Romero-Troncoso, An Experimental Comparative Evaluation of Machine Learning Techniques for Motor Fault Diagnosis Under Various Operating Conditions, IEEE Transactions on Industry Applications 54 (3) (2018) 2215–2224. doi:10.1109/TIA.2018.2801863.

[31] J. Levatić, D. Kocev, S. Džeroski, The importance of the label hierarchy in hierarchical multi-label classification, Journal of Intelligent Information Systems 45 (2) (2015) 247–271.

[32] S. Si, H. Zhang, S. S. Keerthi, D. Mahajan, I. S. Dhillon, C.-J. Hsieh, Gradient boosted decision trees for high dimensional sparse output, Vol. 70 of Proceedings of Machine Learning Research, PMLR, International Convention Centre, Sydney, Australia, 2017, pp. 3182–3190.
URL http://proceedings.mlr.press/v70/si17a.html

[33] X. Li, H. Wang, B. Gu, C. X. Ling, Data sparseness in linear svm, in: Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI'15, AAAI Press, 2015, p. 3628–3634.

[34] G. D. Moody, M. Siponen, Using the theory of interpersonal behavior to explain non-work-related personal use of the Internet at work, Information & Management 50 (6) (2013) 322–335. doi:10.1016/j.im.2013.04.005.

[35] S. Böhm, H. Driehaus, M. Wick, Contextual Push Notifications on Mobile Devices: A Pre-study on the Impact of Usage Context on User Response, in: I. Awan, M. Younas, P. Ünal, M. Aleksy (Eds.), Mobile Web and Intelligent Information Systems, Lecture Notes in Computer Science, Springer International Publishing, Cham, 2019, pp. 316–330. doi:10.1007/978-3-030-27192-3_25.

[36] C. Park, J. Lim, J. Kim, S.-J. Lee, D. Lee, Don't Bother Me. I'm Socializing! A Breakpoint-Based Smartphone Notification System, in: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW '17, Association for Computing Machinery, Portland, Oregon, USA, 2017, pp. 541–554. doi:10.1145/2998181.2998189.