# TaskyApp: Inferring Task Engagement via Smartphone Sensing

**Gašper Urh**
University of Ljubljana
Slovenia
gu7668@student.uni-lj.si

**Veljko Pejović**
University of Ljubljana
Slovenia
Veljko.Pejovic@fri.uni-lj.si


*Ubicomp/ISWC'16 Adjunct*, September 12–16, 2016,
Heidelberg, Germany.

ACM 978-1-4503-4462-3/16/09 $15.00
DOI: http://dx.doi.org/10.1145/2968219.2968547

## Abstract

The knowledge of a user's mental involvement, i.e. task engagement, opens up an array of possibilities for a seamless mobile computing device – human interaction. Today's most ubiquitous personal sensing devices, such as smartphones, are equipped with an array of sensors that may be used to infer different aspects of human behavior. However, inferring task engagement using smartphone sensors remains unexplored. In this paper we present our initial work on automated task engagement inference using only smartphone sensors. We design, develop and deploy a mobile sensing application TaskyApp, and collect 216 data points of sensor readings and task engagement labels from eight users in an office setting. Using machine learning we demonstrate that with up to 67.6% accuracy, relying mostly on the movement sensors, we can correctly infer a user's task engagement.

## Author Keywords

Mobile sensing; Multitasking; Task engagement inference; Interruptibility.

## ACM Classification Keywords

H.5.2. [Information Interfaces and Presentation (e.g. HCI)]: User Interfaces; H.1.2. [Models and Principles]: User/Machine Systems

## Introduction

Mobile computing devices are nowadays more than merely communication tools, and provide a wide range of services: from navigation in a new environment, tracking our exercise routine, over restaurant recommendations, and online social networking, to name a few. Further cohesion of these devices into the Internet of things indicates that the reliance on mobile computing services is yet to grow.

In his 1991 manifesto Mark Weiser outlined the need for the "stealth" ubiquitous computing device – the one that quietly blends with the lifestyle of its user [7]. However, such devices are far from reality, and the unobtrusive understanding of a user remains a major challenge of ubiquitous computing. Human behavior is complex and encompasses different aspects, including our routines, tasks we execute while working, our movement patterns, and many others. A particular aspect describing human behavior – a user's current level of task engagement – is of paramount importance in a number of ubiquitous computing scenarios. For example, a device knowing that a user is highly engaged in a task could defer the delivery of an unimportant message, and notify the user only when the level of task engagement is lowered, thus reducing frustration and improving a user's receptivity to messages.

Up to recently, inference of a user behavior has been outside of the scope of mobile computing. However, two factors, the increasingly personal use of devices and the ever-increasing sensing capabilities of devices, have opened up the opportunity for the automatic inference of certain aspects of human behavior, including mobility, physical activity, and even emotional state of a user. However, the automatic detection of task engagement, to the best of our knowledge has not been explored, yet. Thus, *the goal of our work is to explore the possibility of automated task engagement inference using smartphones*. To fulfill the goal, we have to overcome the following challenges:

- Provide a measurable definition of task engagement levels.
- Collect sensor readings from users' mobile devices at moments pertaining to different task engagement levels.
- Label the collected data with the user-perceived task engagement levels.
- Apply machine learning algorithms to uncover a potential link between the sensed data and the task engagement quantifiers.

In this paper, we tackle the above problems experimentally. We develop a smartphone sensing application that collects data from built-in sensors and at the same time interacts with the user to obtain the task engagement label – i.e. whether a user is engaged in an easy or a difficult task at the moment when sensors are read. We concentrate on the office setting and distribute our app among 8 users, who have collected a total of 216 data points. We then perform machine learning modeling of task engagement, and in our preliminary findings show that a task engagement level can be predicted with 67.6% accuracy. Finally, we present our guidelines for improved task engagement inference using mobile sensors.

## Related Work

Task engagement may be profoundly influenced by a user's multitasking ability. In their paper, Salvucci and Taatgen proposed the idea of threaded cognition – an integrated theory of concurrent multitasking. The theory provides explicit predictions of how multitasking behavior can result in interference for a given set of tasks [5]. The perceived complexity of a task, which can be lowered through memory

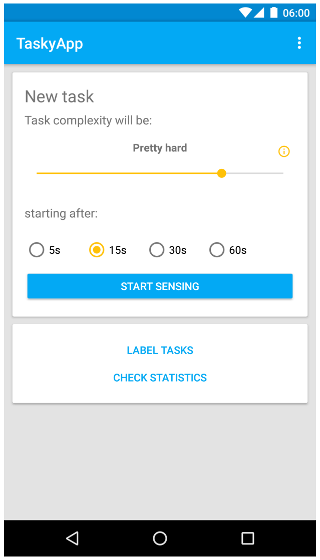


**Figure 1: Screenshot of the TaskyApp manual task labelling page.** The sensing is started after the provided timeout and the collected data is stamped with the provided label.

rehearsal, is critical for (concurrent) task performance [6]. In the highly interactive mobile realm, the offline task engagement directly impacts a user's sentiment and reaction towards an incoming notification. This has been shown both implicitly, where a detected boredom of a user served as a basis for timing recommendation messages [4], as well as explicitly, where a task engagement was queried via the experience sampling method and correlated with a user's sentiment towards a notification [3, 2]. Our goal is to infer a user's task engagement level using only sensors in a smartphone. This is related, but orthogonal to Wiese et al. effort to infer phone placement using proximity and light sensors of a smartphone [8].

## Data Collection

A task is a rather broad term, and for the purpose of our study we limit it to (mental) tasks performed in an office setting. This restriction is well aligned with mental models, such as ACT-R that have been explored in the area of multitasking analysis [6]. We are further interested in a user's task engagement, a metric that may be obtained using physiometric sensors (e.g. measuring pupil dilation [1]). However, to devise a practical, scallable task inference system, our goal is to explore whether commodity smartphones can be used for this purpose. Furthermore, we are interested in the subjective experience of the task. Thus, we rely on explicit task engagement labels provided by a user who answers a question about the perceived difficulty of the current task.

We develop an Android mobile application – TaskyApp[1] – that performs background sensing and then allows the user to explicitly label the task difficulty on a five-point Likert scale ("very easy" to "very hard"). The sensing is initiated automatically, but we also allow a user to start the sensing manually. In any case, a ten-second window of gyroscope, accelerometer and microphone amplitude data is sampled. In addition, we also capture time, current location, activity as reported by the Google Activity Recognition API, nearby Bluetooth and WiFi devices, battery charging status, volume settings, ambient light, active calendar events and the status of the screen. The automatic sensing is triggered at moments when a change in the user's context is detected. That change may stem from the location change or the physical activity change (e.g. from "walking" to "cycling", as reported by the Google Activity Recognition API). A sensing session only starts if a minimum time has passed to the last sensing session, in order to preserve battery charge. Furthermore, a user can select their usual office hours, so that irrelevant sensing is minimized.

Collecting explicit labels is done either at the time of sensing, in case of a user-initiated sensing session, or retroactively, in case of automatic sensing sessions. In the former case, a user is presented with a screen (Figure 1) on which a simple Likert scale is presented. The user selects the task engagement level and the time when the task commences (e.g. "pretty hard task in 15 seconds"). The sensing takes place after the given timeout, and the data is labeled with the provided task engagement label. The latter case, retroactive labeling, is supported by notifications that at the end of the day remind the user to label the sensed tasks, and an application menu option that enables a user to scroll through the list of sensed tasks and provide labels for non-labeled tasks. Each entry in the list contains a map showing the location and the time of when the task was sensed (Figure 2). Finally, to encourage users we introduce a gamification aspect to our app, providing the statistics of an individual's task engagement in comparison to the engagement of other people who participate in the study.

[1] https://play.google.com/store/apps/details?id=si.uni_lj.fri.taskyapp

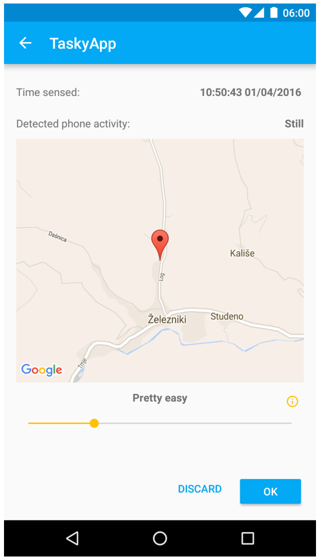


**Figure 2: Screenshot of the TaskyApp retroactive task labelling page.** Users are able to label automatically sensed tasks.

We deployed TaskyApp among ten Android users, office workers between 23 and 55 years old, and collected 216 labeled tasks, as well as more than 1000 non-labeled tasks. At the end we received labeled data from eight different users (Table 1). As an incentive to provide the data, a 50€ voucher was raffled among active participants of the app (i.e. those who label the data daily). The consent form explaining the purpose of the study was presented to the users at the time of installation, and is also available on the app web page[2]. All the collected data was stored at a secure server hosted at our institution.

| User | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Total |
|---|---|---|---|---|---|---|---|---|---|
| Num. of labels | 7 | 50 | 54 | 14 | 8 | 3 | 3 | 77 | 216 |
| Average label | 3.43 | 3.50 | 2.67 | 1.93 | 2.38 | 1.67 | 2.00 | 2.47 | 2.73 |

**Table 1: Per user task label distribution.** Task labels are between 1 (very easy) and 5 (very hard). Most of the labels were provided by three users (83.8%), whereas two did not provide any.

## Analysis and Results

The goal of our analysis is to identify potential links between smartphone sensor readings and the reported level of task engagement. We first preprocess sensor data to extract suitable features, and then apply machine learning methods to identify the relationship with the task engagement labels. In terms of feature extraction, we extract the mean intensity, its variance, mean value and the mean crossing rate of all three axes for both gyroscope and accelerometer. The latter two are also extracted out of microphone amplitude. Other features include a user's activity as reported by the Google Activity Recognition API, the number of WiFi and Bluetooth devices sensed, the charging and screen status, hour of day and the number of currently active Google calendar events.

*Linear Regression*

We first attempt a fine-grain inference of task engagement using linear regression. Information on task engagement, reported on the five-point Likert scale, is encoded as a numeric value from 1 to 5 (very easy to very hard). Using WEKA[3] data mining tool we fit a linear regression, and find the following significant features whose values determine the outcome, i.e. task engagement: accelerometer mean Y-axis and mean intensity values, gyroscope mean crossings for X-axis and Z-axis values, charging status and hour of day. The results are shown in Table 2.

Model provided us with correlation coefficient of 0.4667 and mean absolute error of 0.8379, which may be acceptable for many practical purposes, having in mind that the target values vary between 1 (very easy) to 5 (very difficult). Further, the model indicates that when a phone moves a lot (accelerometer intensity is higher) the task engagement label value leans towards easier tasks. This result is intuitive – as we are limited to the office settings, we expect that the phone is mostly on the table or in user's pocket while the user is working. A phone movement indicates that user has some free time to interact with the phone (e.g. checking phone calls, social media, etc.).

[2] http://193.2.72.121/

[3] www.cs.waikato.ac.nz/ml/weka

| Variable | Coefficient |
|---|---|
| Accel. Y-axis mean | -.04 |
| Accel. mean intensity | -.65 |
| Gyro. X mean crossing rate | .003 |
| Gyro. Z mean crossing rate | .003 |
| Charging = no | .26 |
| Hour of day | .06 |
| *(Regression Constant)* | 7.59 |
| $R^2 = 0.22$ | |

**Table 2: Linear regression model built using extracted features.** Phone movements (accelerometer mean intensity feature) infer at easier tasks in an office setting.

*Classification*
Although the above regression analysis points out to a link between task engagement levels and sensor data reading, fine-grain distinction among engagement levels is difficult. Therefore, we decide to re-encode our labels into only two classes: *easy* and *difficult*. Previously labeled "very easy" and "pretty easy" tasks are now labeled as "easy" and all the other label values are labeled as "difficult". This classification gives us a balanced set of data, having 102 and 114 labeled tasks as "easy" and "difficult", respectively. We again use WEKA, and perform a ten-fold cross-validation using the whole data set, testing of different classifiers.

First we ran ZeroR classifier, which always predicts the majority class, classifying all tasks as difficult, thus resulting in 52.8% accuracy. Now we have a baseline to compare against. We got highest accuracy of correctly classified instances with Decision Stump algorithm, reaching 67.6%. Tasks were classified based only on values of gyroscope's X-axis mean crossing rate. The confusion matrix shows that the classifier fails in correctly labeling difficult tasks – they are inferred as easy in 25% of predictions. For practical purposes, such mistakes are expensive – a messaging app based on the above classifier would predict that a highly engaged user is free for interruption, and would disturb the user at an inappropriate moment.

| **easy'** | **difficult'** | |
|---|---|---|
| 55 (25%) | 47 (22%) | **easy** |
| 30 (14%) | 84 (39%) | **difficult** |

**Table 3: Confusion matrix obtained with the Naive Bayes classifier.** The classifier lowers the rate of false negatives (difficult tasks predicted to be easy), compared to the decision stump classifier.

That helped us to remove attributes, which looked to be informative in our case at first glance, such as microphone features and number of Bluetooth and WiFi devices. This brought us to Naive Bayes classifier. We got slightly lower prediction accuracy (64.4%) for the benefit of lower rate of difficult tasks being predicted as easy, dropped by 11%. The confusion matrix can be seen in Table 3. Naive Bayes is much better at predicting tasks labeled as difficult compared to Decision Stump, which predicts easy tasks much better.

## Conclusion
Inferring task engagement is of paramount importance for seamless integration of mobile communication devices with our everyday lives. In this paper we are the first to examine the potential for automated task inference using only mobile phone sensors. We have shown that data collected from phone's sensors is correlated with an office task complexity, even though the link at this point is still fairly weak. A

simplified problem, where momentarily task complexity is inferred at the easy/difficult granularity, can be solved using off-the-shelf classifiers with the accuracy of more than 65%, which is significantly higher than the baseline classifier accuracy. There are some apps, which could take advantage of the classifier's usage. Users tend to read news more when they are bored [4]. Detecting that a user is likely engaged in a hard task could help news delivery apps (e.g. Flipboard and Google Now) reduce the news update frequency, thus preserving battery and some data traffic. Further analysis shows that accelerometer and gyroscope features proved to be the most informative for inferring task complexity. Others, such as the environment sound amplitude and the number of close by WiFi and Bluetooth devices, did not improve the inference model. Office workers often remain in the same office, therefore the environment does not change, nearby devices remain the same and the ambient noise level most likely remains the same.

Our analysis represents the initial effort in automated task engagement inference using smartphone sensors. The phone was not originally envisioned for such inference, nor is used in a manner that makes such an inference straightforward – a device is not constantly kept at the same place in the office (e.g. it can be at a desktop, in a bag, etc.), users have different habits (e.g. one uses phone a lot, another keeps it away during working hours), the environment may differ for different users (e.g. different offices, different numbers of coworkers). Hence, the next step in data analysis should concentrate on individual task inference models. Due to the lack of data, such an analysis has been omitted in this paper. Finally, the importance of features coming from accelerometer and gyroscope, as well as the previously investigated physiological measurements, points us towards our next step – task engagement inference using automated sensing of wearable devices.

## REFERENCES


1. Shamsi T Iqbal, Xianjun Sam Zheng, and Brian P Bailey. 2004. Task-evoked pupillary response to mental workload in human-computer interaction. In *ACM CHI'04 (extended abstracts)*. Vienna, Austria.

2. Abhinav Mehrotra, Veljko Pejovic, Jo Vermeulen, Robert Hendley, and Mirco Musolesi. 2016. My Phone and Me: Understanding People's Receptivity to Mobile Notifications. In *ACM CHI'16*. San Jose, CA, USA.

3. Veljko Pejovic, Mirco Musolesi, and Abhinav Mehrotra. 2015. Investigating The Role of Task Engagement in Mobile Interruptibility. In *Smarttention, Please! Intelligent Attention Management on Mobile Devices Workshop (with MobileHCI'15)*. Copenhagen, Denmark.

4. Martin Pielot, Tilman Dingler, Jose San Pedro, and Nuria Oliver. 2015. When attention is not scarce-detecting boredom from mobile phone usage. In *ACM UbiComp'15*. Osaka, Japan.

5. Dario D. Salvucci and Niels A. Taatgen. 2008. Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review* 115, 1 (1 2008), 101–130.

6. Dario D Salvucci and Niels A Taatgen. 2010. *The multitasking mind*. Oxford University Press.

7. Mark Weiser. 1991. The computer for the 21st century. *Scientific american* 265, 3 (1991), 94–104.

8. Jason Wiese, T. Scott Saponas, and A.J. Bernheim Brush. 2013. Phoneprioception: Enabling Mobile Phones to Infer Where They Are Kept. In *ACM CHI '13*. New York, NY, USA.