Reducing Interruptions at Work: A Large-Scale Field Study of FlowLight

Manuela Züger[1], Christopher Corley[2], André N. Meyer[1], Boyang Li[2], Thomas Fritz[1], David Shepherd[2], Vinay Augustine[2], Patrick Francis[2], Nicholas Kraft[2], Will Snipes[2]

[1]University of Zurich, Zurich, Switzerland
[2]ABB Corporate Research, Raleigh, United States

{zueger, ameyer, fritz}@ifi.uzh.ch, {christopher.corley, boyang.li, david.shepherd, vinay.augustine, patrick.francis, nicholas.a.kraft, will.snipes}@us.abb.com

ABSTRACT
Due to the high number and cost of interruptions at work, several approaches have been suggested to reduce this cost for knowledge workers. These approaches predominantly focus either on a manual and physical indicator, such as headphones or a closed office door, or on the automatic measure of a worker's interruptibilty in combination with a computer-based indicator. Little is known about the combination of a physical indicator with an automatic interruptibility measure and its long-term impact in the workplace. In our research, we developed the FlowLight, that combines a physical traffic-light like LED with an automatic interruptibility measure based on computer interaction data. In a large-scale and long-term field study with 449 participants from 12 countries, we found, amongst other results, that the FlowLight reduced the interruptions of participants by 46%, increased their awareness on the potential disruptiveness of interruptions and most participants never stopped using it.

ACM Classification Keywords
H.5.3. Information Interfaces and Presentation: Group and Organization Interfaces

Author Keywords
Interruption cost; knowledge worker; awareness; physical indicator; automatic interruptibility measure; field study

INTRODUCTION
Knowledge workers are frequently interrupted by their co-workers [13, 10, 29]. While many of these interruptions can be beneficial, for instance to resolve problems quickly [20], they can also incur a high cost on knowledge workers, especially if they happen at inopportune moments and cannot be postponed [4, 22, 8, 24].

CHI 2017, May 6-11, 2017, Denver, CO, USA.

ACM 978-1-4503-4655-9/17/05...$15.00
http://dx.doi.org/10.1145/3025453.3025662

Due to the high cost and the high number of interruptions that knowledge workers experience every day (e.g., [10, 13]), several approaches have been proposed that can roughly be categorized by the interruptions they address: computer-based and in-person. Studies have shown that the cost of computer-based interruptions can successfully be mitigated by automatically detecting a knowledge worker's interruptibility and mediating interruptions by deferring them to more opportune moments (aka. defer-to-breakpoint strategy) [19, 2, 16]. Another strategy to reduce the cost of computer-based interruptions is to indicate a person's interruptibility to co-workers in a contact-list style application on the computer [30, 21, 5]. While these approaches have also been suggested for addressing in-person interruptions, they did not show to have any effect on them, probably since the contact-list style applications can easily be hidden behind other applications and thus forgotten at communication initiation [21, 5, 12, 14].

For in-person interruptions—one of the most costly kind of interruptions due to their high frequency and immediate nature [29, 13, 24]—approaches predominantly rely on manual strategies to physically indicate interruptibility, such as wearing headphones, closing the office door, or using busy lights that have to be set manually [29, 1]. Since manual approaches are cumbersome to maintain, users generally don't update them on a regular basis and their accuracy and benefits are limited [25]. Only very few approaches have looked at a combination of a physical interruptibility indicator with an automatic interruptibility measure to reduce the cost of in-person interruptions [15, 7] and there is no knowledge on the long-term effects of such approaches.

In our research, we developed the FlowLight approach, an approach to reduce the cost of in-person interruptions by combining a physical interruptibility indicator in the form of a traffic-light like LED (light emitting diode) with an automatic interruptibility measurement based on a user's computer interaction. In a large-scale and long-term field study with 449 knowledge workers from 12 countries and 15 sites of a multinational corporation, we evaluated the FlowLight and its effects in the workplace. Over the course of the study, we collected a rich set of quantitative and qualitative data, including self-reported interruption logs of 36 participants, survey responses of 183 participants that used the FlowLight for at least 4 weeks, and in-depth interviews of 23 participants.

Our analysis of the data shows, amongst other results, that the FlowLight significantly reduced the number of interruptions of participants by 46%, while having little impact on important interruptions. Further, the FlowLight increased the awareness on the cost of interruptions within the workplace, participants felt more productive using the FlowLight and most participants continued using the light for up to 13 months by now. Overall, the gained insights on the long-term usage of the FlowLight provide strong support for the benefits of combining a physical interruptibility indicator with an automatic interruptibility measure in the workplace and its significant impact on reducing in-person interruption costs.

RELATED WORK

Related work on managing interruptions can broadly be grouped into strategies for reducing interruptions and disruptiveness, and ways of measuring and indicating interruptibility.

Reducing Interruptions and their Disruptiveness

Knowledge workers have long recognized the detrimental effects of interruptions and have sometimes developed their own techniques for managing them. These techniques include the use of instant messaging to negotiate availability for an interruption beforehand and reduce the disruptiveness for the interrupted person [26], as well as the use of manual and physical indicators, such as headphones or a closed office door to either signal unavailability or tune out distractions [29].

In addition to these informal means, researchers have developed approaches to reduce the negative effects of interruptions. One strategy to reduce the disruptiveness of interruptions is to defer them from moments when the interruptee is in the middle of a task to naturally occurring breakpoints—aka. ‘defer-to-breakpoint’ strategy. This idea is based on studies finding that the cognitive load drops at task boundaries, and that interruptions at lower cognitive load are less harmful [3, 8]. Iqbal and Bailey developed a system that implements a defer-to-breakpoint policy to reschedule notifications to more opportune moments and found that they caused less frustration and shorter reaction times [19]. Ho and Intille used accelerometers to detect activity transitions and found that messages on mobile devices were better received during transitions compared to random times [16]. While these approaches have been successful at mitigating interruptions from the computer and mobile devices, they do not address the frequent and costly in-person interruptions in workplaces that the FlowLight targets.

A second strategy that builds upon the idea of deferring interruptions to more opportune moments is to indicate a knowledge worker’s interruptibility to potential interrupters and thereby implicitly help negotiate the timing of the interruption. In the following, we discuss approaches to measure and to indicate a knowledge worker’s interruptibility.

Measuring Interruptibility

Previous research has explored various features to measure a person’s interruptibility. For instance, Hudson et al. simulated sensors by coding audio and video recordings into features related to the person’s current context, such as the number of people present or the phone being on the hook [18]. While their approach showed promise in measuring interruptibility, the chosen features are difficult to capture automatically.

To automatically detect a person’s interruptibility, Stern et al. developed an approach that is based on the person’s location and calendar information [28]. Fogarty et al. used speech sensors, location and calendar information and activity on the computer to measure presence and availability [11]; Tani and Yamada measured interruptibility using the pressure applied on the keyboard and mouse [31]; and Coordinate by Horvitz et al. uses user activity and proximity of multiple devices to forecast presence and availability [17].

More recently researchers have also started to use biometric data to measure interruptibility. For instance, Kramer classified interruptibility during a US military training with an electroencephalography (EEG) sensor that captures the electrical activity of the brain [23]. Chen et al. calculated interruptibility based on an electromyography (EMG) sensor that captures heart rate variability and muscle activity [9]. In our previous work, we used various biometric sensors (EEG, electrodermal activity (EDA), skin temperature, and photoplethysmography (PPG)) to predict interruptibility [32]. Overall, research has shown that biometric sensors can be valuable in automatically measuring interruptibility, however, at this point the biometric sensors required to accurately measure interruptibility are generally still too invasive for long-term usage.

The FlowLight builds upon previous research in this area by automatically measuring interruptibility based on a combination of computer activity, calendar information and log-in state. It thereby utilizes a minimally invasive set of features that performs well without compromising the users’ privacy or requiring additional body-worn biometric sensors. It further extends previous research in this area by combining the automatic measure with a physical indicator.

Indicating Interruptibility

To indicate a knowledge worker’s interruptibility to co-workers, most prior research focused on contact list-style tools that are installed on the user’s computer and vary mostly in the data that is used to determine availability/interruptibility. For instance, the ConNexus tool has a contact list view that provides awareness information on a person’s device idleness, log-in state and activity history and thus indicates a person’s availability to facility communication for the integrated communication channels, such as IM [30]. Awarenex and Lilsys build on ConNexus, adding mobile location tracking and physical presence sensors, respectively. An evaluation of these tools found a qualitative improvement in interruption awareness but no reduction in the number of interruptions [5]. Lai et al.’s MyTeam approach uses information on presence, network connection and mouse and keyboard activity to indicate availability in a contact list. In a small user study, they found that the approach decreased the number of phone calls and voice mails but increased the face-to-face interruptions [21]. Fogarty et al. developed MyVine that integrates with a phone, IM and an email client and uses context information from speech sensors, computer activity, location and calendar information. A four week study revealed that the context information was mainly used as presence indicator and did not prevent interruptions via IM [12]. Overall, study results for these computer-based interruptibility indicators suggest that they can help increase awareness on the disruptiveness of interruptions, which could be a first good step as stated by Beyea[6]. However, the results also suggest that these approaches do not reduce in-person interruption costs, which is what the FlowLight is addressing.

Since in-person interruptions are one of the top causes for interruptions in the workplace and their immediate nature makes them particularly disruptive [29], researchers found that knowledge workers use physical indicators, such as headphones or office doors to indicate interruptibility and reduce interruptions and distraction [29].

Only few researchers examined indicators that are not just visible on a knowledge worker's computer monitor. InterruptMe projects availability cues of possible contacts onto a wall at the time when the interrupter is about to initiate a communication [15, 14]. The MoodLight uses an ambient display connected to an electro-dermal activity (EDA) monitor that indicates the excitement level of one or two individuals [27]. Bjelica et al. developed an automatic interruptibility indicator that displays the status through ambient lighting effects and found in a small and short study that the indicator reduced the number of interruptions [7]. The FlowLight presented in this paper uses a physical traffic-light like LED placed at the desk of each person, such that the person's interruptibility status can be seen by anyone approaching. Thereby, our approach is more direct and prominent than subtle ambient lighting and different to previous research, our large-scale field study examines the long-term effects of such physical indicators.

APPROACH AND IMPLEMENTATION

The FlowLight consists of a computer application to automatically determine a user's interruptibility state and a physical LED light to indicate this state to co-workers. The FlowLight was developed iteratively over more than a year and improved continuously based on feedback from a small developer team that we used for testing, and later on, also based on feedback from study participants.

Physical LED Light. FlowLight uses a physical traffic-light like LED to indicate the interruptibility status to co-workers. This light has evolved throughout the pilots[1]. The first model, which was designed and soldered in-house, is shown in Figure 1a. In Figure 1b the same model light is shown encased in plastic and deployed in an open office space. Finally, Figure 1c shows the blink(1)[2] LED light that we adopted to avoid installation issues with certain drivers immediately after the first major pilot, which was also the first of two pilots in India (denoted as India #1 in Figure 2). Typically, we mounted the LED light on a user's cubical wall or outside a user's office.

The light uses different colors to indicate four states: *Available* as green, *Busy* as red, *Do Not Disturb (DnD)* as pulsating red, and *Away* as yellow. Note that these states and colors mimic the ones used by prominent instant messaging services, in particular the one used by the company under study.

[1]We use the term *pilot* to refer to each individual field study trial with a separate team.

[2]https://blink1.thingm.com/

Application. The application features three main components: a *Tracker* to capture events relevant for calculating the interruptibility state, a *Status Analyzer* to analyze the captured events and calculate the user's interruptibility state on the fly, and a *Status Manager* to manage the user's current status, propagating it to the LED light and other applications, in particular instant messaging (IM) clients. The application was implemented to be compatible with the Windows operating system, Skype for Business, an IM and video-conferencing system, and Office 365, a software suite that provides email and calendaring services, amongst others. We chose to tailor our application to these systems and applications due to the IT setup at the target company for our study.

The *Tracker* logs a user's mouse and keyboard interaction. In particular, it collects mouse clicks, movements as pixels moved, scrolling as pixels scrolled and keystrokes (without recording the specific key). This component also logs calendar events to determine meetings and the Skype status.

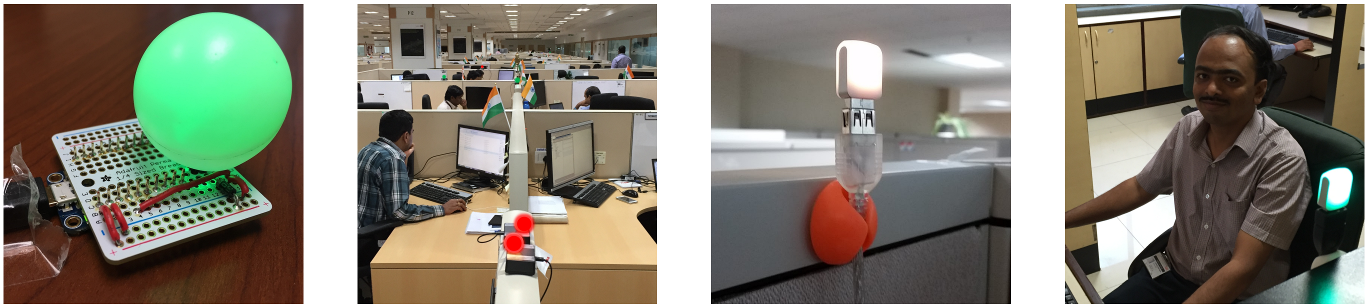
The *Status Analyzer* uses the tracked keyboard and mouse events to calculate the user's interruptibility status on the fly, i.e., whether the user is available, busy, highly busy (DnD) or away. The algorithms used to calculate the interruptibility status are described below.

The *Status Manager* is notified by the Status Analyzer at every change in the user's interruptibility, and then propagates the updated status to the physical LED light and the user's presence status in Skype for Business. The presence status in Skype for Business can also be changed manually by the user, or automatically by the Office 365 calendar, in case a meeting is scheduled. In case the presence status is changed manually, the Status Manager updates the interruptibility state of the application and the physical LED light.

Algorithms for Status Updates. Over the course of this study, we used three different algorithms to determine and update the interruptibility status automatically, improving them based on critical user feedback as discussed below.

FlowTracker. This algorithm sums up the computer interaction in the past three minutes according to heuristic weights assigned to each type of event, which were tuned based on feedback from early alpha and beta users of the FlowLight. If the value of the sum is in between the top 9% and the top 4% of their activity range—we captured averages over the past days—the user is considered busy. If it is within the top 4%, the user is considered highly busy. In our first pilot study in Bangalore, India (India #1 in Figure 2), we used different thresholds at first, namely 13% and 5% based on a prior study that indicated that knowledge workers are not interruptible for approximately 18% of their day. However, several technical writers (and others) involved in that pilot gave strong feedback that the light switched to the busy state too easily, which is why we lowered the thresholds to the mentioned 9% and 4%.

Smoothing. While the FlowTracker showed promise, many early users complained that it was too sensitive to certain input. For instance, a twenty second burst of typing may cause a user to temporarily be shown as busy. Therefore, the Smoothing algorithm marks users as busy if they were active in each of the



(a) July 2015 - Original Prototype

(b) August 2015 - First Major Pilot (Lights Emphasized with Overlays)

(c) October 2015 - Blink(1) Version with Adhesive Clip

(d) April 2016 - Deployment in Second Pilot in India (India #2)

Figure 1: Evolution of the Physical Indicator of the FlowLight Over Time

last three minutes and exceeded a threshold of 100 combined mouse clicks and key presses in the recent past (between 4 and 7 minutes ago). This algorithm reduces frequent changes by requiring over three minutes of activity to become busy and, once busy, by requiring only one above-threshold minute in the recent past to remain busy. To achieve the highly busy status, users had to be busy at the current point in time and had to be above-threshold for fifteen of the last thirty minutes.

Smoothed FlowTracker. While the Smoothing algorithm leads to fewer status changes, since it relied on a static threshold (i.e., 100 combined mouse clicks and key presses), it did not adapt to individual users' work patterns. For instance, designers working on drawings tended to use mouse clicks almost exclusively, which makes it difficult to exceed the threshold. Thus, we finally combined the FlowTracker algorithm with the Smoothing algorithm to achieve the advantages of both approaches. This algorithm, currently in use, operates as the Smoothing algorithm, but instead of using a static threshold, it utilizes the FlowTracker algorithm to determine above threshold values. This algorithm eliminated all of the most common complaints reported by pilot users. Further refinement of the algorithm is left for future work.

Although our main intent was to use an algorithm to infer interruptibility, we offered participants a "Manual Only" mode since it was requested by some participants, especially those with management roles that needed to be available to others most of the time, and we noticed (and our study confirmed) that our algorithms might not be accurate for everyone or for all activities requiring focus, such as reading or thinking.

EVALUATION

To evaluate the FlowLight, in particular the combination of the physical indicator and the automatic interruptibility measure as well as its effect on knowledge workers, we conducted a long-term and large-scale field study with 449 knowledge workers. For this study, we installed the FlowLight at over 15 locations in 12 countries of one multinational corporation. Over the course of the study, we collected a rich set of data using a combination of experience sampling, a survey, an interview and computer interaction monitoring. Figure 1 illustrates a few pictures of the FlowLight in use in different pilots. Figure 2 indicates the increasing and continuous number of participants

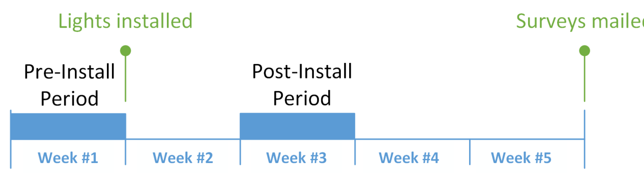


Figure 3: Timeline of Study Procedure

and the major pilots of this study since its beginning and up to September 2016.

Study Procedure

For each team participating in our field study we conducted the same five-week pilot procedure as illustrated in Figure 3. Prior to the start of a pilot, we asked the participants to install the FlowLight application in 'data collection only'-mode.

In *Week #1* of the pilot, users were instructed to use the FlowLight application to manually log the time and severity of each interruption during the five day work week. Our application allowed participants to log interruptions by a click on the taskbar menu or a hotkey combination for minimal invasiveness. As soon as an interruption had been logged, a single modal dialog appeared that asked participants to specify the severity of the interruption on a 5-point Likert scale.

At the beginning of *Week #2*, the physical indicator of the FlowLight was installed and the automatic status update feature for the interruptibility status was activated. To minimize Hawthorne-type effects and have participants and co-workers get used to the FlowLight, we then waited for one week before we gave further instructions.

At the beginning of *Week #3*, we again asked participants to manually log their interruptions for 5 work days. We also reminded participants about the manual logging in Week #1 and #3 to ensure they would not forget.

During *Week #4 and #5* users continued using the FlowLight. Throughout these 5 weeks the application collected anonymized usage data. At the end of *Week #5*, after participants had the FlowLight for four weeks, our application prompted them to complete a survey. The survey took an average of 14.2 minutes to complete and had questions on the

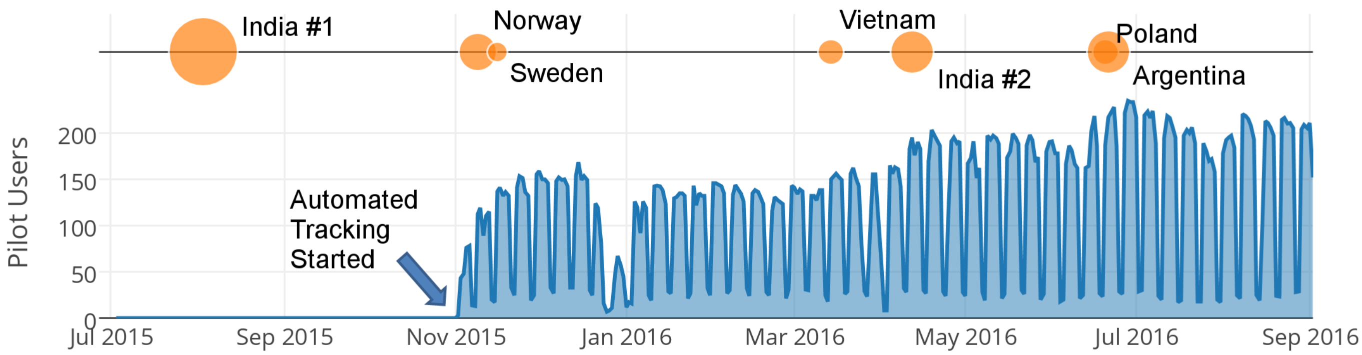


Figure 2: FlowLight Users Over Time (Size of Orange Circles Indicates the Number of Participants; Regular Dips in the Number of Pilot Users Represent Weekends and the Prolonged Dip in December/January 2016 represents the Christmas Break)

FlowLight approach and its impact, in particular on participants' interest in continuing using the approach, its impact on interruption costs, productivity and interaction behavior, on the accuracy of the automatic state detection and manual setting, as well as on general feedback and demographics. After completing the survey, users were asked on the last page of the survey to upload their data collected by the FlowLight application, which included the usage data logs and the logs of the manually captured interruptions.

For a deeper understanding of the long-term usage, experience and effect of the FlowLight, we conducted in-depth interviews with a subset of participants approximately two months after they installed the FlowLight. Interview participants were selected semi-randomly, based on accessibility, availability and willingness to participate in the interview. The interviews were on average 19.5 minutes long and the questions focused on the benefits and limitations participants observed with the FlowLight approach, as well as on how it impacted their own behavior and interactions in the team over the course of the two months since the installation. For instance, we asked participants whether they felt that their colleagues respected their FlowLight or if they noticed situations in which the status was not accurate. Note that the interview and survey questions can be found on our supplemental materials site [3].

Independent of the timeline of the study procedure, we also started to anonymously log the number of people running the FlowLight application each day. For privacy reasons, we only keep track of the number of unique active FlowLight users in the online log.

Participants

Since the beginning of our study 13 months ago, we installed the FlowLight approach with a total of 449 participants from 15 sites, located in 12 different countries, of one multinational corporation. From these 449 participants, we were able to gather:

Survey responses from 183 participants (IDs: S1-S183), 144 male and 39 female, with an average age of 36.0 years (standard deviation, in the following denoted with ±, of 8.7), an average professional experience of 12.0 years (± 8.0), from a variety of work areas, including 77 participants in development, 56 in other engineering, 24 in project management, 15 in other non-engineering, and 11 in testing, and with various job roles, including 70 individual contributors, 36 other, 32 leads, 31 managers, 8 executives, and 6 architects;

Interview transcripts (conducted by us) from 23 participants (IDs: I1-I23), 22 of which were male, 1 female, average age of 36.9 years (± 5.8), average experience of 13.2 years (± 4.7), and with various job roles, including 9 managers, 11 software developers, 1 researcher, 1 product owner, and 1 tester;

Interruption logs (self-reported) from 36 participants across six different countries, 13 from Argentina, 6 from Norway, 5 from Poland, 5 from Switzerland, 5 from Sweden, and 2 from the USA;

Usage data logs from 47 participants (IDs: D1-D47) 20 from Argentina, 18 from India, 4 from Poland, and 5 from Vietnam.

Online logs from all 449 participants that installed the approach (each one had the application running for at least one day after we integrated the logging feature).

Note that due to privacy concerns with the collected data, we did not require participants to identify themselves in each step and/or fill in their demographics, except for the survey, which is why we can only report some demographics for each round and are not able to track the participants across the different methods, for instance the survey and the self-reported interruption logging.

Data Collection and Analysis

Survey and Interview. In total, we collected survey responses from 183 participants after they had been using the FlowLight for at least four weeks, and interview transcripts from the 23 participants after they had been using the FlowLight for approximately two months. To analyze the textual data of the survey and interview responses, we used techniques based on Grounded Theory, in particular open coding and axial coding to determine higher level themes. To establish a common set of codes and themes, two of the authors applied open axial coding to the same subset of interview transcripts and then established a common understanding and defined a structure for the most commonly mentioned concepts. As the topics of the survey and interviews overlap, we used and extended the same coding scheme to analyze the textual survey responses.

[3] https://sites.google.com/site/focuslightonline

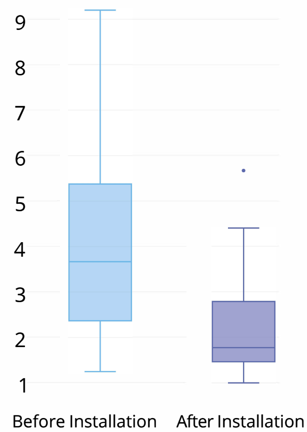
To validate the analysis of the survey results, two additional authors extracted their main findings from a subset of the responses independently.

Interruption Logs. Interruption logs capture the self-reported interruptions per participant logged with the FlowLight software. We collected interruption logs with at least two logged interruptions from 102 participants. We down-selected these to 36 logs by applying strict filtering criteria to ensure data validity as follows. We excluded all interruptions in all logs that were accidentally logged during the first five days after the installation of the FlowLight, as interruptions in the period right after the installation are not representative due to Hawthorne-type effects, such as participants getting used to the FlowLight, and co-workers asking curiosity questions. We then excluded all participants, that logged interruptions for fewer than three days in the pre- or the post-installation period. We chose three days as the threshold for each period to ensure a representative sample of work days for comparison without a too strong bias by individual outlier days. Each of the 36 interruption logs captured a combined average of 9.0 work days ($\pm$ 2.2) for pre- and post-period, and contained an average of 28.9 total logged interruptions ($\pm$ 17.0) per participant for the combined time period. We used these interruption logs to compare the impact of the FlowLight on the number of interruptions rated as disruptive by participants.

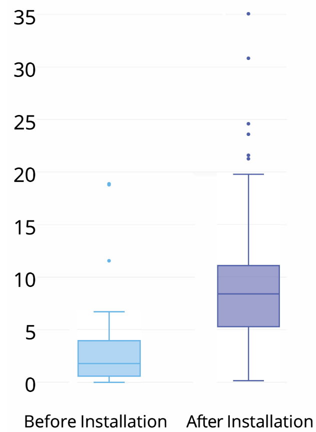
Usage Data Logs. We captured usage data logs from a total of 179 participants. These logs consist of computer interaction logs, such as mouse and keyboard events, and FlowLight usage data. Since we wanted to analyze user behavior before and after installing the light, we removed any logs that did not include at least two days before and after installing the light. We also excluded logs older than January 2016, as key usage messages were not yet logged by our software, making the analysis infeasible. We ended up with 47 usage data logs containing a total of 1560 work days. These logs consisted of an average of 7.3 work days ($\pm$4.2) prior to light installation and 25.9 work days ($\pm$14.0) after light installation per participant.

We analyzed usage logs in two ways. First, we counted the number of status change events recorded in the log per day per user for the period before and after the light installation event. It is worth noting that we only included usage logs within the five work days and not on weekends. Second, we used the intervals between status change events detected by one of our algorithms to determine how much time was spent in each status, again for before and after light installation. To eliminate inappropriate intervals (e.g., a user did not turn off the workstation after work), we only accumulated the duration within 12 hours per day.

Online Logs. We collected online logs for a total of 305 days from November 2, 2015 until September 2, 2016 and from 449 participants. These logs were used to determine how many users were using the FlowLight on a given day (as shown in Figure 2). We analyzed these logs by summing up the number of unique identifiers that appeared in the log on a given day, which represents the number of active users for that day. Based on participants' feedback during the period of the field study, we deployed the three main variations of the algorithm described earlier to set the status of the FlowLight. We analyzed differences between the data sets gathered with the three main variations of the algorithms and found no significant differences between the data collected with any two variations, neither in the collected survey items, nor the interruption logs. In the following, we will therefore present the results aggregated over all variations.



(a) Interruptions per Participant and Day.



(b) Number of State Changes per Participant and Day.

Figure 4: Logged Interruptions and State Changes Before and After Installing the FlowLight.

RESULTS

In this section we present the primary findings of our field study. We first examine the effect of the FlowLight on the cost of interruptions before we examine how the FlowLight changed participants' interruption awareness, their interruption-related behavior, and their perception of productivity. Subsequently, we present insights on the costs of the approach, on the influence of its accuracy, on its continued usage by participants and on professional differences.

Reduced Cost of Interruptions

Figure 4a is based on the 36 collected interruption logs and illustrates the distribution of the number of interruptions per day and participant in the period before and the period after participants had been using the FlowLight for one week.

Overall, the number of interruptions decreased after the installation and one week usage of the FlowLight by an average of 1.9 ($\pm$1.6) interruptions (49%) per participant and day, from 4.1 ($\pm$2.1) to 2.2 ($\pm$1.1). A Wilcoxon signed-rank test showed that this reduction is statistically significant (Z =-5.0, p <.000001).

A second Wilcoxon signed-rank test only on the number of severe interruptions (disruptiveness rating of 4 or 5) per day and participant further showed that there is also a statistically significant reduction with p <.001 and Z = -3.2.

An analysis of the survey results (see Figure 5 for more detail) further supports that installing the FlowLight reduced the cost of interruptions. 55.0% of the 182 survey participants that answered the question stated that they either strongly agree, agree or agree somewhat that they were interrupted less than

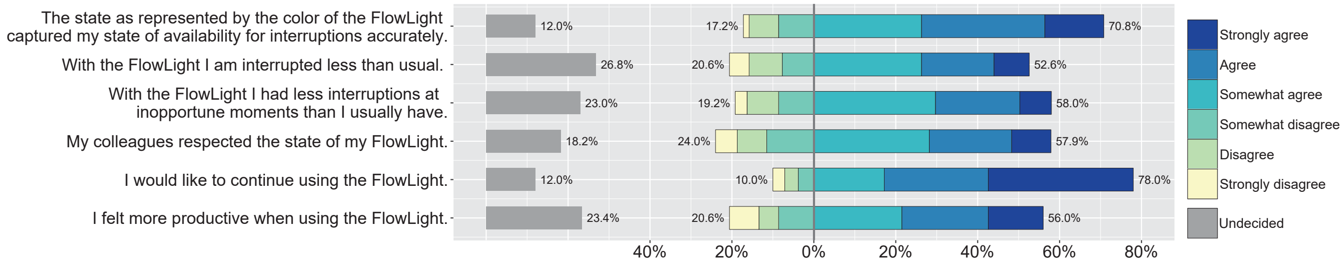


Figure 5: Results of a Subset of the Survey Questions.

usual during their work, while only 20.3% disagreed with it. Even more participants, 59.3%, agreed that they had less interruptions at inopportune moments than usual, whereas only 19.8% disagreed with this statement.

During interviews participants echoed this quantitative evidence. In interview excerpts (full quotes listed in subsequent subsections) participants consistently mentioned that interruptions were reduced. They claimed that the pilot *"..resulted in less interruptions.." (S126)* , eliminated interruptions from colleagues (e.g., *"When [the light]'s red I think they don't interrupt." (I11)*), and *"..didn't stop [interruptions] completely but they surely reduced." (S16)* .

Overall, our findings from the interruption logs, survey questions, and interview questions show strong support that the introduction of the FlowLight reduced the cost of interruptions in terms of the overall number as well as their severity.

Increased Awareness of Interruption Cost

After using the FlowLight for some time, participants developed a high degree of awareness for the cost of interruptions:

"It brings more awareness to what people are doing. Sometimes people take it for granted that people are always interruptible. But there is actually a cost or a penalty when you interrupt someone. So, I think just the concept is good because it reminds people that there is sometimes a good time and a bad time to interrupt people. So, I think just from an awareness campaign, it's valuable as well." (I20)

"The pilot increased the sensitivity to interruption. Team members think more about whether an interrupt is necessary and try to find a suitable time." (S45)

The FlowLight thereby served as a physical reminder for the interruptibility of co-workers in the moment and participants generally respect it and its state:

"It's kind of a like a mood indicator ... so it tells people the state ... of the owner of the light. And then it helps people be more aware or attentive to what my current situation is." (I18)

"I think what really changed is ... a different consciousness about interruptions in our team and also with my colleagues ... I think ... they really respect the light. When it's red I think they don't interrupt." (I11)

Overall, 70% of the 23 interview participants explicitly stated that the FlowLight is respected in their offices and 59.6% of 183 survey respondents agreed that colleagues respected the state of their FlowLight vs. 23.0% that did not (Figure 5).

The increased awareness and respect also triggered participants to change their behavior in a variety of ways, ranging from thinking twice before asking, to deferring the interruption, asking before interrupting and changing to a different communication channel, such as email or instant messaging:

"People ask each other if they are available, even when the light is green, even to people with no light. When I see the colleague I want to ask a question ... has a red light, then I wait a while, or write an email." (S77)

"If it's red, I'll send them a message so that when they're no longer busy or something like that, they'll see the message and they can respond to it then ... so it doesn't require an immediate response" (I19)

Fortunately, participants used common sense when working with FlowLights. If a light was red or red blinking participants would still interrupt if the request was urgent:

"Once I go up there [to the person] and I see the light and then I also see that they're pretty intense then I'll push it off unless I really need to get answered to." (I17)

Feeling of Increased Productivity and Self-Motivation

As a further effect of the FlowLight, 58.5% of the survey respondents felt more productive using it, while only 20.1% disagreed with this (Figure 5). This feeling of increased productivity often stemmed from the fewer interruptions:

"I definitely think it resulted in less interruptions both in person and via Skype. This resulted in more focus and ability to finish work." (S126)

Another reason for the increased productivity is that the FlowLight serves for some participants as a self-monitoring device that motivates them to become or stay focused, which, however, can also be distracting at times:

"Mostly it has helped as a personal monitor only for me. If I see the light red, I sense I am in the flow and I keep working." (I2)

"When I notice that my light is turning yellow, and I'll feel like, 'Oh yeah, I've been idle' and then I do something ... I think the other way, yeah, there's some effect there too. Like, if I see that it's red, or even flashing red, then I'm like, 'Yeah, I've been very active, or productive, I should keep that going.' At the same time, I think it's also a little bit distracting too. Sometimes just because the light is there, I turn around to check it." (I12)

Costs of Using the FlowLight

While people experienced reduced interruption costs and increase in productivity, there are also costs when starting to use the FlowLight. Especially right after installing it the curiosity of co-workers can lead to an increase of interruptions, which, however, diminishes after a few days:

“People walk by, they see it, they ask me questions, ’What’s that? How does it work? What’s going on?’ like this.” (I19)

“Initially there were many people just curious to know what the light is about. This increased the number of interruptions but after few days, people started to respect [it].” (S16)

A few participants also experienced situations in which the FlowLight provoked interruptions, as the green color of the light might be misunderstood as an invitation (observed by 26% of the interviewed participants):

“What I definitely notice is that green is more inviting. So it actually encourages people to come by and say, “Hello” for me at least.” (I20)

In some cases, changing the interaction culture might require a mandate from higher up or can even be too expensive:

“The more important issue is for it to work, you have to have people committed to following the light rules, which probably requires engagement of some higher management ... and requires introducing the lights to a wider audience.” (I6)

“For us ... the main cost of introducing [it is] that you have to change how you are used to interact with people, that you first have to remember to take a look at the light. That’s something that’s probably too much for the team. [In] our environment .. it’s easier to look at the people than at lights.” (I8)

If colleagues choose to ignore the light, especially for unimportant interruptions, it can lead to negative emotions:

“So, for us, what we also heard sometimes is that people have the light red, and others still interrupt them, and they’re like, ’Oh no, I have this light red, why did they?’ Like it bothers them, and it creates negative emotions almost more than it creates positive emotions...” (I17)

Finally, the public disclosure of the interruptibility status might make people feel exposed at times (8% of survey participants agree, 6% strongly agree) or lead to negative feelings:

“Oh, do other people see that my light is yellow? And are they thinking that I’m not working?” (I12)

Like any new technology, there is a cost to adopting the FlowLight. However, most of the identified costs diminish quickly or can be mitigated by clear direction from management. Overall though participants predominantly stated that the colors of the light were interpreted appropriately and were mostly not concerned about being observed.

Automatic State Changes and Accuracy

The algorithm of the FlowLight caused automatic state changes to indicate a user is, for instance, available for interruptions or busy and not interruptible. Figure 4b illustrates the change in distribution of the number of state changes per participant per day before and after installation. A Wilcoxon signed-rank test with Bonferroni-adjusted alpha levels of .01 per test (.05/5) showed a statistically significant change in the number of state changes (Z =-5.5337, $p \leq$ .01) with an increase in state changes from 1.8 before to 8.4 after. This

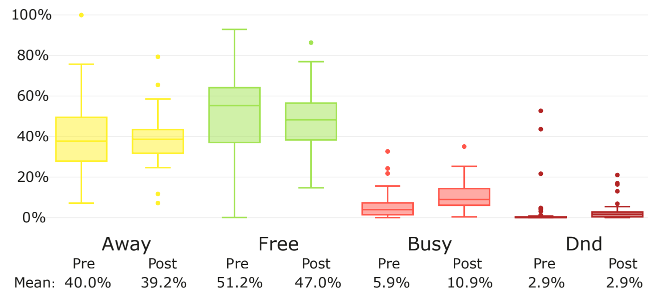


Figure 6: Time Spent in each State Before (Pre) and After (Post) Installation.

increase shows that the automated algorithm is affecting users’ availability status in Skype. Figure 6 presents the time spent in each state before and after light installation. Analysis of this data shows a small insignificant decrease in time spent in the available state from 51.2% to 47.0% (Wilcoxon signed-rank test Z =-1.7143, p = .043) and a significant increase in the time spent in the busy state from 5.9% to 10.9% (Z =-3.6403, $p \leq$ .01), a very small yet significant difference in do not disturb state (Z =-3.2093, $p \leq$ .01), and no significant changes to time spent in the away state. Note that during the before-light period the status was already affected by meetings entered in the calendar, which caused the status to change to busy.

Participants generally agreed that the FlowLight captured their state of availability for interruptions accurately:

“I think it [state representativeness] was actually quite good, because what I found is, if I’m not working on a critical task, for example, responding to email which usually isn’t critically mind provoking. The light would be green and then people would take that opportunity to stop by and see what they needed to talk about. Whereas if I was in the middle of a meeting or if I was more involved in my work, it would turn red and then at that point they might wait for it to turn green. That’s my impression.” (I20)

Overall, 71.0% of survey respondents agreed that the FlowLight captures their state accurately while only 15.8% disagreed (Figure 5). This shows that even an interruptibility measure based on a simple algorithm might be accurate enough to be accepted by users and provide value.

At the same time, interview participants and 64% of our survey respondents mentioned that there are situations where the FlowLight was not representative and accurate, partly stemming from limitations in measuring interruptibility solely with computer interaction data:

“The light was mostly green while debugging code. During debugging, I think interrupts hurt a lot. On the other hand, the light was sometimes red when working on documents / e-mails that do not require too much focus.” (S45)

“[The] light captures the movements of the mouse and keyboard, and actually, there are times, which I think of a solution separate from the time, which I implement [it] so ... I’m the most occupied when I think something and usually, I write it on a paper or just keep it on my mind.” (I4)

In several cases, participants just changed to setting the state manually when it was not accurate and they wanted to indicate to others that they are available or do not want to be disturbed:

> *"There was a case when I was reading an article, and I needed a 100% concentration on that, so I just manually changed my status to busy. It was helping me a lot. I think my colleagues are also doing the same when they are engrossed in an article and they want free time, they'll just keep their light busy." (I4)*

In fact, 32% of our survey respondents reported to have changed their Skype status (which is linked to the FlowLight) more often after the light was installed, 23% less often and 45% had no changes. With the FlowLight installed, 17% of participants reported to change their status at least once a day, 37% one to several times a week, and 46% rarely or never. The job role can also affect the accuracy of the FlowLight, especially for managers, administrative assistants, and sales people. For instance, several managers mentioned that interaction was such a core part of their role that they felt they should always be available and turned off the automatic feature.

Continued Usage of FlowLight

Most participants, 82.6% of the 23 interview participants and 79.1% of the 183 survey participants, stated their intention to keep using the FlowLight even after the pilot period. This sentiment is reflected in actual usage data: two months after installing the FlowLight application 85.5% of users remained active (384/449).

Based on online logging of application instances that we started in November 2015, Figure 2 shows the number of active FlowLight users per day. The Figure also depicts the start date and relative size for the major pilots (e.g., India #1 started in August '15 and had 80 participants, Norway started in November '14 and had 44).

Note that due to holidays in different locales, vacation, sick days, and travel the number of active users per day is consistently about 70% of the number of unique users over the last month (e.g., a measure of 200 active users per day indicates about 315 number of unique users in the last month).

In spite of most users continuing to use the FlowLight, about 20% of users discontinued usage. There were several reasons that we identified from the interviews and surveys that decreased the benefit of the FlowLight, including the office layout and the visibility of the LED light, the company culture and people ignoring the lights, the initial willingness to use such a system, and the accuracy of the state indicated by the FlowLight. In some cases, the decreased benefit also resulted in participants ceasing to use the FlowLight:

> *"From my perspective that was something I was against from the first day but as I said I decided to join the pilot because I am a team member. ... From time to time I was looking at it but it was a little bit discouraging because the color of the light didn't reflect what I was doing and maybe after one week of using it I gave up totally." (I9)*

Professional Differences in Using the FlowLight

An analysis of the survey responses with respect to professional roles shows that developers (including testers) and project managers stated more frequently than participants from other working areas that they wanted to continue using the FlowLight , even though not significantly (82% vs 70% on average) and perceived their state to be significantly more accurate (77% vs 60%, t =2.51, p =.01). For project managers, these differences might be explained by the fact that they also reported more often (but not significantly) to manually change their FlowLight status on a daily basis than participants from other work areas (24% vs 16%) and by our experiences gathered during the installation phase, in which managers often asked to disable the automatic mode completely as they wanted to be available for most of their work time. For developers, the differences might be explained by their extensive computer interaction, but future research is needed to confirm this.

DISCUSSION

The results of our large-scale and long-term study show that the FlowLight can reduce the interruption costs for knowledge workers and can increase the awareness, amongst other benefits. In the following, we discuss implications of our findings, in particular with respect to the combination of the physical indicator with the automatic interruptibility measure, the accuracy of the measure, and the cost of not interrupting. Finally, we discuss threats to validity and limitations of our study.

Reasons for FlowLight's Positive Effects

The FlowLight uses a combination of a physical LED light with an automatic measure based on computer interaction to update the user's interruptibility status. The findings show that the approach was well adopted and successfully reduced in-person interruption costs. This poses the question if these effects might after all stem solely either from the automatic interruptibility measure or the physical LED light. With respect to the sole use of an automatic interruptibility measure, prior related work that used an automatic measure to update computer-based contact-list style tools, did not find any or the same level of positive effects as our study on both, cost reduction and awareness [30, 5, 21]. On the other hand, manually maintaining the interruptibility state incurs a high cost as shown by previous research [25] and only very few of our users switched to the manual option in cases the algorithm was not accurate enough or they wanted to ensure some undisrupted time. In addition, our findings show that while participants have a high tolerance for the accuracy of the automatic interruptibility status updates, when inaccuracies happen too often, participants also stop using the approach altogether. Overall, this indicates that the combination of the physical LED light and the automatic interruptibility measure is important to provide significant benefits to knowledge workers to use it in the long-term and that it led to the positive impact on awareness and interruption cost found in our study.

Accuracy of Automatic Interruptibility Measure

Participants' high tolerance for the accuracy of the automatic interruptibility measure of the FlowLight poses the question of how accurate the underlying measure has to be to provide sufficient benefit to the user. Over the course of our field study, we adapted the automatic measure two times to account for early user feedback, yet we did not find any significant differences in the effects on interruption cost and behavior. However, we intend to study the relation between accuracy and the effects on interruption cost further in the future.

Also, while participants had a high tolerance, they reported numerous situations in which they observed the status to be set incorrectly. The most frequent situation in which the status is incorrect occurs when participants "think" about something and experience a high cognitive load, yet do not interact with the computer at all. In future work and with the continuously decreasing invasiveness of biometric sensors, we plan to extend our approach to integrate biometric sensors, to cover these situations more accurately. We further plan to improve our algorithm by integrating application data, which we were not able to collect in this study due to privacy constraints. Knowing the current application might improve the algorithm's accuracy, e.g. one might be less interruptible while working in a development related program and more while being in an email client. As the nature of work and interactions vary across work areas and job roles, tailoring the algorithm accordingly could further improve its accuracy.

Cost of Not Interrupting

As related work has shown, not all interruptions are bad and some are definitely needed, for instance, to unblock co-workers. By physically indicating knowledge workers as not interruptible (Busy and DnD state), the FlowLight might prevent co-workers from interrupting them for important issues, reducing overall team productivity. The findings of our study on the FlowLight provides evidence that this cost is minimal at best for two reasons. First, a data analysis of the usage logs collected for our study shows that the FlowLight ends up having a significant yet small effect on the time that a knowledge worker is indicated as not interruptible (+5% per day). Second, while the FlowLight increases the awareness of the cost of interruptions, participants still interrupt their co-workers regardless of the FlowLight state if they have an important concern to discuss, as also stated by 35% of our interview participants, without being explicitly asked.

Threats and Limitations

A major threat to the validity of our study is the completeness of the collected data. For instance, we were not able to identify participants across different data sets. While we encouraged participants to share their data and ensured them that we only use it for research purposes, we could not demand it due to privacy concerns. We were also not able to collect geographic data due to privacy concerns and thus were not able to analyze geographic differences.

Similarly, the accuracy of the interruption logs might be incomplete or not completely accurate. Since interruption logs are based on self-reports, participants might have forgotten to log some interruptions. Also, the work patterns and habits of the days on which they logged interruptions before and after the installation of the FlowLight might have been significantly different, which makes it more difficult to compare the effect of the FlowLight. We tried to mitigate this risk by only including the logs of participants who logged interruptions for more than three days before and three days after and by regularly reminding them to log their interruptions. Furthermore, different participants might have different criteria and judgement standards for logging interruptions. We tried to mitigate this fact by instructing participants to only log external in-person interruptions at work. In addition, by using a paired test that only compares within subject (Wilcoxon signed rank), we mitigate this effect as long as participants did not change their definition of an interruption over time.

We limited the validity threats related to generalizability across individuals and teams by collecting data from 449 participants from twelve countries and with a variety of job roles. As not all participants are native English speakers, there might be a response bias. We tried to mitigate this risk by providing sufficient instructions, opportunity for contacting us if participants had any questions, and also by visiting each major pilot site to introduce and explain the study. Based on the large number and diversity of participants, we observed that responses were not dominantly distributed to extremes, which would indicate that these knowledge workers were particularly biased based on such difficulties. From our in-person experience we can report that with very few exceptions we perceived similar acceptance, respect and in general a very positive perception of the FlowLight across all locations.

Another threat is the influence of the various algorithms on the study results. Since we wanted to ensure that participants are satisfied with the FlowLight and that we take their feedback serious, we evolved the algorithm two times. To mitigate the risk of a certain bias in the data, we looked for significant differences between populations where we might expect to find them and did not find any.

CONCLUSION

In-person interruptions at the workplace can incur a high cost and consume a lot of a knowledge worker's time, if they happen at inopportune moments. While there are several approaches to possibly reduce the interruption costs, little is known about the impact of a physical and automatic interruptibility indicator. In this paper, we presented FlowLight—an automatic interruptibility indicator in the form of a physical traffic-light like LED—and reported on results from a large-scale and long-term field study with 449 participants from 12 countries. We found that the FlowLight significantly reduced the number of interruptions by 46%. We also observed an increased awareness of the potential disruptiveness of interruptions at inopportune moments, which impacts the interaction culture in a positive way, and that our approach can motivate knowledge workers and make them feel more productive. We discuss the importance of combining the physical indicator with the automatic interruptibility measure and the high tolerance of participants to the accuracy of the approach. Overall, our study provides deep insights and strong evidence on the very positive effects of the long-term usage of the FlowLight, and the continued usage of the approach by most participants indicates the success of the approach.

ACKNOWLEDGMENTS

The authors would like to thank all study participants. This work was funded in part by SNF.

REFERENCES

1. 2016. http://www.embrava.com/. (2016).
2. Ernesto Arroyo and Ted Selker. 2011. Attention and intention goals can mediate disruption in human-computer interaction. In *IFIP Conference on Human-Computer Interaction*. Springer, 454–470.
3. Brian P Bailey and Shamsi T Iqbal. 2008. Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction (TOCHI)* 14, 4 (2008), 21.
4. Brian P Bailey and Joseph A Konstan. 2006. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in human behavior* 22, 4 (2006), 685–708.
5. James Bo Begole, Nicholas E Matsakis, and John C Tang. 2004. Lilsys: sensing unavailability. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. ACM, 511–514.
6. Suzanne C Beyea. 2007. Distractions, interruptions, and patient safety. *AORN journal* 86, 1 (2007), 109–112.
7. Milan Z Bjelica, Bojan Mrazovac, Istvan Papp, and Nikola Teslic. 2011. Busy flag just got better: Application of lighting effects in mediating social interruptions. In *MIPRO, 2011 Proceedings of the 34th International Convention*. IEEE, 975–980.
8. Jelmer P Borst, Niels A Taatgen, and Hedderik van Rijn. 2015. What Makes Interruptions Disruptive?: A Process-Model Account of the Effects of the Problem State Bottleneck on Task Interruption and Resumption. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. ACM, 2971–2980.
9. Daniel Chen, Jamie Hart, and Roel Vertegaal. 2007. Towards a physiological model of user interruptability. In *IFIP Conference on Human-Computer Interaction*. Springer, 439–451.
10. Mary Czerwinski, Eric Horvitz, and Susan Wilhite. 2004. A diary study of task switching and interruptions. In *Proc. SIGCHI*. ACM, 175–182.
11. James Fogarty, Andrew J Ko, Htet Htet Aung, Elspeth Golden, Karen P Tang, and Scott E Hudson. 2005. Examining task engagement in sensor-based statistical models of human interruptibility. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 331–340.
12. James Fogarty, Jennifer Lai, and Jim Christensen. 2004. Presence versus availability: the design and evaluation of a context-aware communication client. *International Journal of Human-Computer Studies* 61, 3 (2004), 299–317.
13. Victor M González and Gloria Mark. 2004. Constant, constant, multi-tasking craziness: managing multiple working spheres. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 113–120.
14. Juan David Hincapié-Ramos, Stephen Voida, and Gloria Mark. 2011a. A design space analysis of availability-sharing systems. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, 85–96.
15. Juan David Hincapié-Ramos, Stephen Voida, and Gloria Mark. 2011b. Sharing availability information with InterruptMe. In *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, 477–478.
16. Joyce Ho and Stephen S Intille. 2005. Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 909–918.
17. Eric Horvitz, Paul Koch, Carl M Kadie, and Andy Jacobs. 2002. Coordinate: Probabilistic forecasting of presence and availability. In *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 224–233.
18. Scott Hudson, James Fogarty, Christopher Atkeson, Daniel Avrahami, Jodi Forlizzi, Sara Kiesler, Johnny Lee, and Jie Yang. 2003. Predicting human interruptibility with sensors: a Wizard of Oz feasibility study. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 257–264.
19. Shamsi T Iqbal and Brian P Bailey. 2008. Effects of intelligent notification management on users and their tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 93–102.
20. Ellen Isaacs, Steve Whittaker, David Frohlich, and Brid O'Conaill. 1997. Informal communication re-examined: New functions for video in supporting opportunistic encounters. *Video-mediated communication* 997 (1997), 459–485.
21. Jennifer Lai, Sachiko Yoshihama, Thomas Bridgman, Mark Podlaseck, Paul B Chou, and Danny C Wong. 2003. MyTeam: Availability Awareness Through the Use of Sensor Data.. In *INTERACT*.
22. Gloria Mark, Daniela Gudith, and Ulrich Klocke. 2008. The cost of interrupted work: more speed and stress. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 107–110.
23. Santosh Mathan, Stephen Whitlow, Michael Dorneich, Patricia Ververs, and Gene Davis. 2007. Neurophysiological estimation of interruptibility: Demonstrating feasibility in a field context. In *In Proceedings of the 4th International Conference of the Augmented Cognition Society*.
24. Daniel McFarlane. 2002. Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human-Computer Interaction* 17, 1 (2002), 63–139.

25. Allen E Milewski and Thomas M Smith. 2000. Providing presence cues to telephone users. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. ACM, 89–96.

26. Bonnie A Nardi, Steve Whittaker, and Erin Bradner. 2000. Interaction and outeraction: instant messaging in action. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. ACM, 79–88.

27. Jaime Snyder, Mark Matthews, Jacqueline Chien, Pamara F Chang, Emily Sun, Saeed Abdullah, and Geri Gay. 2015. Moodlight: Exploring personal and social implications of ambient display of biosensor data. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 143–153.

28. Hermann Stern, Viktoria Pammer, and Stefanie N Lindstaedt. 2011. A preliminary study on interruptibility detection based on location and calendar information. *Proc. CoSDEO* 11 (2011).

29. Edward R Sykes. 2011. Interruptions in the workplace: A case study to reduce their effects. *International Journal of Information Management* 31, 4 (2011), 385–394.

30. John C Tang, Nicole Yankelovich, James Begole, Max Van Kleek, Francis Li, and Janak Bhalodia. 2001. ConNexus to awarenex: extending awareness to mobile users. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 221–228.

31. Takahisa Tani and Seiji Yamada. 2013. Estimating user interruptibility by measuring table-top pressure. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1707–1712.

32. Manuela Züger and Thomas Fritz. 2015. Interruptibility of software developers and its prediction using psycho-physiological sensors. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2981–2990.