

# Balancing Awareness and Interruption in Mobile Patrol using Context-Aware Notification

*Jan Willem Streefkerk\*, TNO, the Netherlands*

*D. Scott McCrickard, Virginia Polytechnic Institute and State University, USA*

*Myra P. van Esch-Bussemakers, TNO, the Netherlands*

*Mark A. Neerincx, Delft University of Technology, the Netherlands*

## ABSTRACT

*In mobile computing, a fundamental problem is maintaining awareness of the environment and of information presented as messages on a mobile device. In mobile police patrol, officers need to pay attention to their direct environment and stay informed of incidents elsewhere. To prevent unwanted interruption, a context-aware notification system adapts the timing and appearance of incident messages, based on user activity (available, in transit or busy) and message priority (high, normal or low). We evaluated the benefits and costs of adaptive notification compared to three uniform notification styles (presenting full messages, postponing messages or presenting indicators). Thirty-two trained student participants used a prototype notification system in a controlled mobile patrol task. The results were validated in a follow-up study with twenty-four police officers. We found that full messages elicited a quick, but sometimes incorrect response to incident messages, whereas with adaptive notification responses were slower but only for lower priority messages. The results are discussed in view of notification systems' design for mobile professionals.*

*Keywords: Mobile devices; context-aware computing; notification systems; interruption; awareness.*

*\*) Corresponding author, e-mail: [j.w.streefkerk@tno.nl](mailto:j.w.streefkerk@tno.nl)*

## INTRODUCTION

In mobile professional domains, such as the police domain, increasingly more operational information becomes available. In addition, more and more interaction with mobile devices is required, straining users' cognitive resources. Consider mobile police officers on foot patrol. They work in a dynamic environment characterized by large variations in time pressure and workload (Sørensen & Pica, 2005). They need to focus their attention on their direct environment to be able to detect criminal behavior. At the same time, they need to be informed about incidents occurring elsewhere which may require their presence. Thus, while on patrol, officers must divide their attention to ensure awareness of their direct environment and of incidents elsewhere.

Current notification systems in the police domain broadcast all incident messages to all officers as a central dispatcher does not know the current activity of each officer in detail. While this maintains officers' awareness of incident messages, it can diminish awareness of the environment due to unwanted interruption. This causes officers to focus their attention inappropriately (e.g., on the device instead of on the environment) and can result in decision errors, longer response times and potentially dangerous situations. For example, a message about an illegally parked car (low priority) might be irrelevant and distracting for an officer who is just apprehending a suspect (high priority). However, to a high priority message about a colleague in danger, even officers engaged in an incident need to respond quickly. So, depending on two important context factors (message priority and officer activity), an incident message might constitute an unwanted or an appropriate interruption.

This illustrates a fundamental problem in mobile human-computer interaction: the cost-benefit trade-off that exists between awareness and interruption. Awareness of incident messages on a mobile device may be more important than the need to focus on the environment, requiring an interruption. On the other hand, avoiding interruption (e.g. by postponing messages) comes at the cost of delayed awareness of the message (Horvitz, Apacible & Subramani, 2005; McCrickard & Chewar, 2003). Depending on the context (i.e. priority of the message), delayed awareness might not be a problem at all. Hence, to balance this awareness trade-off, notification systems should determine *when* a particular interruption is appropriate (appropriate timing) and *how* it should be presented (appropriate appearance) (Bailey & Konstan, 2006; McCrickard & Chewar, 2003; Streefkerk, van Esch-Bussemaekers & Neerincx, 2006). Previous research has shown that postponing, scheduling or deferring interruptions until appropriate moments mitigates the negative effects of these interruptions (Adamczyk & Bailey, 2004; Iqbal & Bailey, 2008; McFarlane, 2002). Also, the presentation modality (e.g. visually, auditorially) and salience of the message influences its interruptiveness (Kern & Schiele, 2003; Nagata, 2003; Streefkerk, van Esch-Bussemaekers & Neerincx, 2007).

So in short, mobile users want to stay aware of incoming messages, but do not want to be disturbed when they are busy, unless the message is important. The level of interruption is determined by when and how a mobile device presents a message. To address this awareness-interruption trade-off, we design a context-aware notification system that adapts the *notification style*; i.e. the timing (e.g. postpone message) and appearance (e.g. use an indicator icon) of an incident message. The system takes into account users' activity (available for a new incident, in transit to an incident or handling an incident) and relative priority of the message (higher, equal or lower than the current incident) at the moment of notification to determine which notification style is appropriate. This is expected to balance the awareness-interruption trade-off: limiting unwanted interruption while maintaining awareness of the environment. In this paper, we take the police domain as application domain using the following approach. First, based on previous research and context modeling in the police domain, we demonstrate that the awareness trade-off is indeed problematic in this domain. Next, we test the effects of different notification styles on the awareness trade-off in a controlled mobile experiment with non-professional participants. Finally, we validate the results of the first study with a follow-up study (previously presented at a conference) in which police officers use the same context-aware notification system in a realistic task setting.

Designing for mobile professional domains, such as the police domain, requires an iterative approach in which design solutions are incrementally improved (Neerincx & Lindenberg, 2008). The first study evaluates the benefits and drawbacks of four intermediate notification style designs on task performance and the user experience. Intermediate designs may not yet be suitable to use in actual task-relevant settings with police end-users (see also “Evaluating context-aware notification”). For example, postponing all incident messages for police officers will certainly interfere with their task performance. Hence, we first employ trained non-professional student participants in a mobile patrol task. The evaluation setting captures core task features of police patrol relevant to the awareness trade-off (observation, navigation, notification and incident handling). The goal is not to reflect actual police work literally, but to create relevant divided attention situations to do controlled measurements of task performance. Trained participants have to notice targets and handle incidents while their notification system presents incident messages in one of four notification styles (full message, postpone, indicator or adaptive). Compared to the other three styles, we expect that adaptive notification will improve the effectiveness (e.g. decrease decision errors on messages) and efficiency (e.g. improve response time to messages) of responding to incident messages. Adaptive notification is expected to prevent unwanted interruption of incident handling, leading to a positive user experience of the system. The follow-up study focuses on the difference between adaptive notification and full messages, where we expect to find similar results with police officers.

In the remainder of this paper, the related work section shows approaches to realize context-aware notification in other domains (e.g. office-based tasks) and how they relate to the current study. Then, we describe how interruption affects mobile computing and how a context-aware notification system can help, taking police patrol as an example. The evaluation method of the first experiment is described next, focusing on the operationalization of the experimental setup. We present the results and validate them with a follow-up study involving police officers. The main results and limitations of both studies are addressed in the discussion and implications are presented for notification systems’ design for mobile professionals.

## **RELATED WORK**

To manage interruption, notification systems must have knowledge about the user (e.g. activity) and task (e.g. priority) factors to subsequently adapt the notification presentation in a meaningful way (Bailey & Iqbal, 2008; Gievska & Sibert, 2005; Horvitz, Kadie, Paek & Hovel, 2003; Iqbal & Bailey, 2008; McCrickard & Chewar, 2003; Streefkerk et al., 2006). These context-aware notification systems use sensor information from users’ context — such as location, activity, or task phase — as input to predict appropriate moments of interruption. Interruptions unrelated to the primary task negatively influence task performance and affective state. Longer task completion times, higher task switching costs, higher error rates, and increased frustration and anxiety have been demonstrated (Adamczyk & Bailey, 2004; Bailey & Konstan, 2006; Cutrell, Czerwinski & Horvitz, 2001; Nagata, 2003). Based on these results, researchers argue for an attention management system that gathers knowledge about users’ context to decide when to interrupt (Adamczyk & Bailey, 2004; Bailey & Konstan, 2006; Fogarty et al., 2005).

## Timing of Interruptions

These negative effects can be mitigated by timing interruptions at appropriate points in task execution (Bailey & Konstan, 2006; Fogarty et al., 2005; Gievska & Sibert, 2005). A study on instant messaging interruptions concluded that interruptions presented during the evaluation phase of a task were more readily accepted than during planning or execution phases (Cutrell et al., 2001). Adamczyk and Bailey (2004) predicted the best (e.g., between coarse breakpoints of a task) and worst (e.g., during subtasks) interruption moments based on an a priori task model. They demonstrated significantly lower mental effort, frustration and anxiety for interruptions at the predicted best moments. Following up on this line of research, Iqbal and Bailey (2008) showed that deferring notifications to task breakpoints reduces response time and user frustration. A related study demonstrated that this is due to lowered workload at breakpoints (Bailey & Iqbal, 2008). Furthermore, previous work showed that in mobile environments, predicting interruptibility could be done reliably based on location or activity transitions. Sensor-based modeling of the use context (in this case location and ambient sound) could predict user interruptibility with up to 94% accuracy (Kern & Schiele, 2003). In addition, user acceptance of interruptions was found higher just before or after location transitions (Kostov, Tajima, Naito & Ozawa, 2006) or physical activity transitions (Ho & Intille, 2005) compared to other interruption moments.

However, these studies in the mobile domain did not consider the notification *content* or *priority* and how it related to the primary task. The priority of the notification should be considered with respect to the priority of the ongoing task in determining the timing and style of notification. Relative to task priority, a lower priority message needs to be postponed, whereas a higher priority message needs to be presented immediately. Furthermore, identifying breakpoints in task execution and postponing notifications until such breaks will result in performance benefits and increased user acceptance. The present study will identify task priority and breakpoints based on user actions (e.g., finished with an incident) and use this knowledge to appropriately time incident messages.

## Notification Presentation

Context-aware notification systems can adapt the presentation modality (e.g., visual, auditory and tactile signals), salience and information content of notifications to limit interruption. For example, Kern and Schiele (2003) adapted the modality (auditory or tactile) and salience (beeping or ringing) of a notification to personal interruptibility in a social context. Other work by Sawhney and Schmandt (2000) resulted in the mobile *Nomadic Radio* prototype, which presented more salient auditory signals and more elaborate information content as message importance increased. While tactile cues are used to limit disruption, especially to relieve visual attention (Hopp, Smith, Clegg & Heggstad, 2005), these require the device to be in close contact with the body. Finally, in multi-device environments, notification messages can be presented on different devices or platforms influencing their interruptiveness (e.g., presenting information as a text message on a cell phone or as an e-mail message on a desktop computer) (Ebling,

Hunt & Lei, 2001; Horvitz et al., 2003). For the police domain, adapting the visual and auditory salience of notifications seems the most promising approach.

Related work focused on evaluating different notification styles (notification salience and information density) for a mobile notification system (Streefkerk et al., 2007). Using adaptive notification styles based on message priority and location, users felt less interrupted. This slightly improved their task performance in high workload situations. The current study follows up on this line of research, by defining the design space of possible notification styles (timing and visual appearance).

Another approach to limit the disruption of notifications is creating anticipation of interruptions (Andrews, Ratwani & Trafton, 2009; McFarlane, 2002; Nagata, 2003). In mobile computing tasks, providing prior knowledge of when an interruption will occur has been shown to improve performance compared to unanticipated interruptions. This approach is difficult for the mobile police domain, as interruptions are inherently unexpected. Instead, we attempt to prevent unwanted interruption away from the environment by using specific, subtle user interface designs (e.g., an indicator icon). This icon is used for notification messages users need to be aware of to anticipate future actions.

### **Evaluating Context-Aware Notification**

In evaluating context-aware systems, the evaluation setting, participants and metrics should be chosen carefully (Streefkerk, van Esch-Bussemakers, Neerincx & Looije, 2008b). For the two studies in this paper both the fidelity and realism of the evaluation setting are important (Smets et al., 2010). The fidelity of our evaluation is determined by how well it captures the awareness trade-off (e.g., divided attention situations) and core task features of police patrol (e.g, notification, navigation and incident handling). The realism of the evaluation regards how well it resembles real-life police work. Consequently, the first study is high in fidelity, but low in realism, employing a controlled mobile experiment (analogous to *mobile quasi-experimentation* ; Oulasvirta, Tamminen, Roto & Kuorelahti, 2005). The follow-up study is higher in realism, employing a police patrol task in a virtual city environment. Although the added value of field evaluation is contested (e.g, Kjeldskov & Graham, 2003), evaluating context-aware applications in a real-life mobile setting lets users experience the adaptive system within the use context and task flow. This allows users to judge the appropriateness of adaptive system behavior in relation to changes in the use context.

Regarding participants, end-users may be employed in all stages of the development process, depending on their availability (Streefkerk et al., 2008b). However, access to police end-users is limited, making it more cost-effective to only employ them at select moments. Previously, a focus group with police officers helped to define the rules, criteria and task features of mobile police patrol (Streefkerk et al., 2006). The current study focuses on the awareness-interruption trade-off, which depends on general cognitive abilities, instead of domain-specific police knowledge. Also, intermediate designs may not yet be suitable to evaluate with end-users in the actual domain, as they may give a wrong impression of the final design. Because of this, the notification designs are evaluated with trained student participants in a simulated, relevant task setting. To

increase ecological validity, the results are validated in a follow-up study with police officers.

In the current mobile experiment, evaluation metrics are based on criteria from the police domain (Streefkerk et al., 2008b). For example in police patrol, fast responses to high priority incidents are important. The notification system should thus be assessed on how well it facilitates this response (e.g. by measuring response time). Furthermore, adaptive system evaluation should capture a specific set of user experience metrics, such as controllability, predictability and affective responses (Kort & De Poot, 2005). Such metrics determine whether a system is accepted and used.

Concluding, earlier work demonstrated that appropriate timing of interruptions mitigates distraction and that context information (such as location, task priority or activity transitions) determines when it is appropriate to present or postpone a notification. Furthermore, adapting the salience and information density of notifications limits interruption. Still, a lack in empirical work on context-aware notification in critical mobile work domains (such as military or police work environments) is apparent. It is not clear what the trade-off in terms of task performance is between awareness of the environment and awareness of messages. What are the effects of (in)appropriately timed notifications on effectiveness, efficiency and user experience in these domains? Furthermore, is a system that presents different notification styles within the task flow understandable and easy to use? To address these gaps, we design a controlled mobile experiment that captures situations relevant to the awareness-interruption trade-off and test the effects of different notification styles in these situations.

## **CONTEXT MODELING**

This section demonstrates how the awareness-interruption trade-off influences work in the police domain, resulting in a task-relevant scenario for our controlled mobile experiment. Based on previous research in the police domain, we argue that message priority and user activity are two relevant context factors that determine which notification style is appropriate in which situations. Finally, we describe how rules on notification styles are implemented in an experimental notification system prototype.

### **Priority and Activity in Mobile Police Patrol**

Knowledge on the typical tasks in police patrol comes from a focus group with police professionals, as well as participatory observation of police patrol during a field study (Streefkerk et al., 2006; Streefkerk, van Esch-Bussemaekers & Neerincx, 2008a). Police officers on patrol need to focus their attention on their direct environment to detect criminal behavior. They may be on the move toward an incident (*in transit*), or already handling an incident. At the same time, they receive incident messages informing them of incidents elsewhere. The priority of new incidents is relative to the priority of the current activity (*lower, equal, or higher*), indicating which incident is more important to handle first and how quickly police officers should respond (Streefkerk et al., 2008a).

The scenario below shows that relative priority and officer activity are two important context factors to determine whether an incident message is relevant.

Police officer Jason is on patrol in the city centre on a busy Friday night. He receives a high priority incident message about a domestic violence incident, and proceeds to the incident location (in transit). While navigating, he receives two low priority incident messages about an unpaid fine and about an illegally parked car. Distracted, he takes a wrong turn and has to backtrack to reach the right address. He manages to talk to the perpetrator to calm him down. While speaking, he suddenly receives a high priority message about a colleague in danger. As he is nearby, he decides to rush to the scene.

Jason has to make the right decision in responding to incident messages; i.e. ignore the message about the fine, but respond quickly to the message about the colleague in danger. Similarly, an incoming low priority incident message may not be directly relevant and cause unwanted interruption. For example, handling a domestic violence incident must not be interrupted by a new incident message about a fine that needs to be collected. Postponing all messages when busy might mitigate the problem of unwanted interruption, but diminishes the officers' awareness of incident messages that are relevant. For example, Jason still needs to receive a high priority message about a colleague in danger. Or when moving towards the domestic violence incident, he needs to be aware of any equal or higher priority messages to decide if a switch to another incident is necessary. So, balancing awareness of the environment with awareness of an incident message hinges on the interplay between how important the message is (relative priority) given what the officer is currently doing (officer activity). Based on these two factors, we can distinguish nine notification situations (see also Table 2). The next section specifies appropriate notification styles (timing and appearance) for each of these situations, based on notification rules.

**Table 1. The notification design space (timing and visual appearance) with the three notification styles used in this study.**

<b>Timing</b>	<b>Visual appearance</b>	
	<i>Full message</i>	<i>Indicator</i>
<i>Direct</i>	Presenting full message directly, with sound (F)	Presenting indicator directly, with sound (I)
<i>Postpone</i>	Postponing full message, without sound (P)	N/A

**Table 2. Notification matrix matching the notification styles to relative priority and officer activity.**

<b>Relative priority</b>	<b>Officer activity</b>		
	<i>Available</i>	<i>In transit</i>	<i>Handling incident</i>
<i>Higher</i>	F	F	I
<i>Equal</i>	F	I	P
<i>Lower</i>	F	P	P

## Notification Styles and Rules

The design space of our notification styles is defined by notification timing (directly or postponed) and visual appearance (full message or indicator) (see Table 1). Auditory signals are coupled to timing; sounds are used for directly presented notifications, whereas no sound is used for postponed messages. As in previous work, the salience of

the sound conveys the priority of the message (Streefkerk et al., 2007). Presentation timing is either direct (when message becomes available) or postponed (until a change in officer activity). Presenting full messages is a salient form of visual appearance, creating immediate awareness of incident messages and allowing a fast response. Postponing messages limits interruption of ongoing work, but also limits awareness of these messages. Alternatively, a less distracting, subtle notification can be presented in the form of an indicator icon. This creates awareness of a new incident message, without overly disrupting the current activity. Postponing an indicator (the fourth cell in Table 1) is not considered a useful notification style.

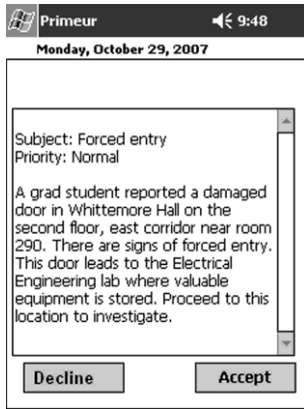
Based on the police patrol task characteristics in the previous section, we can now specify the following notification rules for an adaptive notification system. These rules dictate for each notification situation which style is appropriate. The result of this process is the notification matrix in Table 2. The notification rules are:

1. If the officer is *available* (i.e. not handling an incident), then a full message is presented directly, regardless of the incident priority.
2. If the officer is *in transit* to an incident and a *higher* priority incident occurs, then a full message is presented. This aids awareness of the incident message and facilitates a switch to the new incident.
3. If the officer is *in transit* to an incident and an *equal* priority incident occurs, then an indicator is directly presented.
4. If the officer is *handling an incident* and a *higher* priority incident occurs, then again an indicator is directly presented.
5. In all other cases, the messages are considered not directly relevant and are postponed until the officer is available, to avoid unwanted interruption.

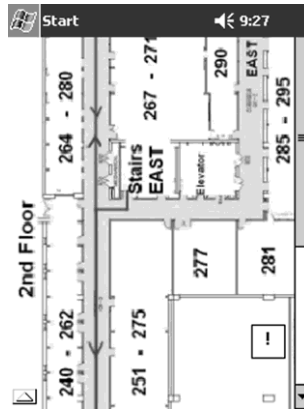
## Implementation

An experimental prototype of this context-aware notification system was implemented on a PDA (Personal Digital Assistant) handheld computer, similar to the handheld device police officers used in an earlier field study (Streefkerk et al, 2008a). Based on the notification matrix in Table 2, the prototype system presented notification messages in different styles. Full messages (see Fig. 1) were shown as text messages in the interface. Users could “Accept” or “Ignore” a message with two buttons below the message text. Indicators (see Fig. 2) were shown as a small icon (!) in the lower right corner of the screen. By clicking on this icon, the full message could be read. Postponed messages were presented as full message when the user was available again. Sounds were used to convey the message priority; a loud sound repeated three times for high priority messages, a softer sound repeated twice for normal priority messages, and an even softer sound repeated once for low priority messages. Users could review and check off messages in the message list (see Fig. 3).

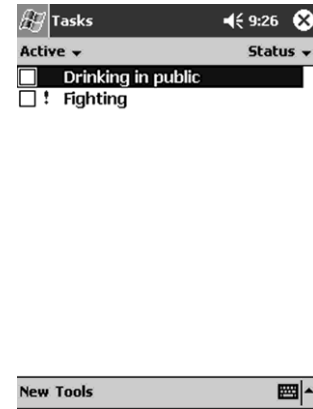




**Fig. 1.** Screenshot of the full message.



**Fig. 2.** Screenshot of the indicator (“!” in lower right corner).



**Fig. 3.** Screenshot of the message list with two incident messages.

The system determines message priority from standard incident categorization in the police domain. Officer activity can be recognized from communication signals, common in police work. Officers usually acknowledge receiving an incident message, arriving at the incident location and finishing an incident. Based on these communication signals and priority categorizations, the system can determine relative priority and officer activity. In this experimental prototype, the context-awareness of the system was simulated by having the test leader send the notification messages. When participants were “available” (i.e. there was no current incident), relative priority of a new incident was always higher than walking the patrol round. User activity was determined by the following user actions: accepting a message, arriving at the scene and finishing an incident. Based on these actions (“acknowledge”, “on scene”, “finished”), user activity was classified as “available”, “in transit” or “handling incident”. While this prototype employs a Wizard-of-Oz setup, it is important to note that the information used by the prototype (priority and activity) is readily available in the police domain and that there are no technical constraints to fully implement this functionality. In fact, handheld computers for police officers on patrol are becoming more common to complement the information exchange via radio transceivers (Streefkerk et al., 2008a). In addition, in an earlier focus group, police officers commented positively on such a context-aware notification system and expected it to improve their patrol (Streefkerk et al., 2006).

In summary, we described the design of an adaptive notification system that estimates the importance of a message (relative priority) given the current activity of the user (user activity). The system chooses one of three different notification styles (full message, indicator or postpone) based on a set of notification rules (Table 2).

## EVALUATION METHOD

To systematically assess how different notification styles affect the trade-off between awareness of the environment and of incident messages, a mobile patrol task was constructed for the purpose of this first experiment. The task was based on the police scenario described above and required walking a predetermined route through a

university office building while looking for targets (cf. awareness of the environment). Trained student participants carried out the patrol task with the prototype notification system, which presented messages on current incidents (cf. awareness of incident messages). When a message was presented, participants suspended the patrol, read the message, moved to the incident location and handled the incident. Either during navigation to or during handling this incident, *an interrupting message* about a second, new incident was presented. The presentation moment and priority of these messages was systematically varied, at unexpected moments for the participants.

## Hypotheses

To capture the awareness trade-off, notification timing and appearance were manipulated between four different experimental conditions. In three conditions, *uniform* notification styles presented the interrupting message always as “full message”, “postpone” or “indicator”, regardless of message priority or officer activity. The fourth, *adaptive* condition followed the notification matrix in Table 2 to determine timing and appearance of notification presentation. We investigated the effects of these notification styles on effectiveness (decision errors, number of targets) and efficiency (response time, incident handling time) of task performance as well as user experience measures (message interruptiveness, workload, user preference). The following hypotheses on task performance and user experience specify the awareness trade-off for each of the notification styles (see also Table 3):

**Table 3. Hypothesized effects of the notification styles on awareness of the environment and awareness of incident messages.**

Notification styles	Awareness of environment			Awareness of incident messages	
	<i>Number of targets</i>	<i>Message interruptiveness</i>	<i>Incident handling time</i>	<i>Decision errors</i>	<i>Response time</i>
Full message (F)	Low	High	Long	Intermediate	Short
Postpone (P)	High	Low	Short	High	N/A
Indicator (I)	Intermediate	Intermediate	Intermediate	Low	Short
Adaptive (A)	High	Low	Short	Low	Short

1. *Full messages* will maintain awareness of incident messages, resulting in a short response time. However, using full messages will sometimes cause users to inappropriately attend to the messages, resulting in decision errors. Furthermore, this will also decrease awareness of the environment, causing a low number of targets noticed, high interruptiveness of messages and long handling times.
2. *Postponing* all messages will maintain awareness of the environment (a high number of targets noticed, low message interruptiveness and short handling times). However, postponing will limit awareness of incident messages, resulting in a high number of decision errors. Because messages are postponed to a moment when users are available, response time is less relevant.

3. Providing an *indicator* will maintain awareness of incident messages, resulting in a low number of decision errors and short response time. But presenting indicators for messages that are not directly relevant still creates unwanted interruption, resulting in intermediate number of targets noticed, intermediate message interruptiveness and intermediate handling times.
4. The *adaptive* notification style will balance awareness of the environment (high number of targets noticed, low message interruptiveness and short handling time) with awareness of incident messages (low number of decision errors and short response time).

In addition, this study will explore whether different notification styles impact workload and user preference differently. For example, maintaining awareness of both the environment and incident messages may come at the cost of increases in workload.

## Participants

Thirty-two undergraduate and graduate Computer Science students participated in this study (24 male, 8 female). Their mean age was 22.8 years ( $SD = 2.8$ ). All of them had extensive experience with computers, software and computer programming. 72% had never before or only occasionally used a PDA, and 15% used a PDA on a daily basis. None of them was familiar with the use of navigation software on mobile devices or with the layout of the building. They were compensated for participation in this study.

## Patrol Task

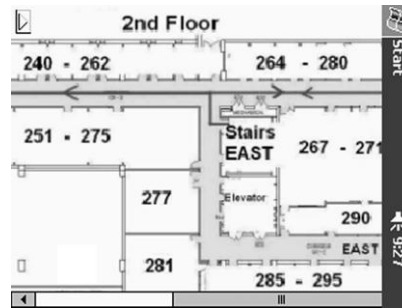
The patrol task consisted of walking a predefined route along four floors through a university office building. Participants were accompanied by the test leader during this task. To focus their attention on the environment, participants were required to notice 14 *targets*, consisting of 4-inch yellow paper disks, placed on the walls at various locations throughout the building (see Fig. 4). When they noticed a target, participants gave verbal confirmation. The test leader counted the number of targets participants noticed. Participants were instructed to perform this task as fast as possible without navigation errors while noticing all targets. To aid navigation, the PDA showed a map of the route on each floor (see Fig. 5). Participants could scroll and switch between these floor plans.

Participants were equipped with the notification system that presented in total twelve incident messages (five with high priority, four with normal priority and three with low priority) during the entire patrol. All messages specified the incident, its priority and location, as well as instructions to the participant (e.g. “proceed to room 435 to investigate”; see also Fig. 1). Examples of incidents were a fight between students (high priority), forced entry into a lab (normal priority) or interviewing a burglary victim (low priority). Incident handling consisted of four stages:

- Reading the incident message and deciding to “Accept” or “Ignore” the incident.
- Moving to the incident location (in transit) after having accepted the incident.
- Handling the incident by listening to an audio / video narration of incident details.
- Checking incident off (available) and returning to the patrol route.



**Fig. 4.** Targets in the patrol task consisted of yellow paper disks at random places on the wall (arrow added).



**Fig. 5.** Floor overview on the PDA (rotated 90 degrees). The light gray area represents the hallway, while the dark gray line indicates the route.

Incident messages were presented in sets of two. The first message of the set (i.e. M1, M3, M5, etc.) was presented when participants were “available”. These messages were always presented as full message. Shortly after that, an *interrupting incident message* signaling a second incident (i.e. M2, M4, M6, etc.) was presented, either during “in transit” to or during “incident handling” of the first incident. By systematically varying the presentation moment and priority of these interrupting messages, six distinct interruption moments were created (see Table 4). Participants always finished the message set before receiving the next set.

Participants were required to make a correct decision to attend or ignore the incident message and handle or ignore the incident. When the interrupting message had higher priority than the current incident (in message sets 3 and 6), the correct decision for participants would be to pause their activity, read the interrupting message and switch to this incident *as fast as possible*. The wrong decision would be not to attend to the message. When the interrupting message had lower priority (in message sets 2 and 5), participants could ignore the interruption and attend to the message when they were available again. The wrong decision would be to immediately attend to the message, or to switch to the incident. In case of equal priority (in message sets 1 and 4) participants could decide for themselves which incident to handle first. The observer noted the correctness of the decisions.

**Table 4.** Presentation order of the twelve messages (M1 to M12) during the patrol task.

Message Set	First message (when “available”)	Interrupting message	Relative priority	Interrupted activity
1	M1 (normal)	M2 (normal)	Equal	<i>In transit to incident M1</i>
2	M3 (high)	M4 (low)	Lower	<i>Handling incident M3</i>
3	M5 (low)	M6 (high)	Higher	<i>In transit to incident M5</i>
4	M7 (normal)	M8 (normal)	Equal	<i>Handling incident M7</i>
5	M9 (high)	M10 (low)	Lower	<i>In transit to incident M9</i>
6	M11 (normal)	M12 (high)	Higher	<i>Handling incident M11</i>

## Experimental Design and Manipulation

This experiment employed a 4 (notification style; between subjects) x 3 (relative priority; within subjects) mixed design. *Notification style* was manipulated between the four experimental conditions (see Table 1). In the “Full message” condition (F), the prototype presented the second, interrupting message of the set directly as full message, when it became available. In the “Indicator” condition (I), all interrupting messages were directly presented as indicators. In the “Postpone” condition (P), all interrupting messages were postponed until the participant was available again and then presented as full messages. In the “Adaptive” condition (A) however, relative priority of the interrupting message and user activity were used to determine notification presentation according to the notification matrix in Table 2. The same set of messages and incidents was used in all conditions, to accurately compare the notification styles between conditions. The presentation order of the route and message sets was reversed for half of the participants to avoid order effects. Each participant participated in one experimental condition (6 male and 2 female participants per condition). A between-subjects design had to be employed, because the patrol route could only be followed once without knowing the route and location of the targets.

## Measures

In this experiment, individual characteristics, performance measures on the patrol task and subjective measures were collected (see Table 5).

*Before the experiment* individual characteristics (gender, age, mobile and desktop computer usage and computer game experience) were assessed using a questionnaire. To check whether participants in each condition differed in task switching and memory ability, two tests were administered. First, the trail making test (TMT) is a paper-based test of “connecting the dots” (Miner & Ferraro, 1998). The percentage difference in completion time between the first part (only numbered dots) and the second part (dots alternating with numbers and letters, i.e. 1, A, 2, B, 3...) is taken as a measure for *task switching ability*. Second, a computerized memory test was administered, consisting of a 6 x 4 grid of cards placed facedown. By turning the cards over, matching pairs had to be found as fast as possible. The task completion time is measured as the *memory score* (Neerinx, Pemberton, Lindenberg & van Besouw, 1999).

*During the experiment* effectiveness of the patrol task was measured as two types of decision errors: inappropriately attending to or ignoring a message (*read errors*) and inappropriately handling or ignoring an incident (*handling errors*). The observer noted and counted these decision errors. In addition, the observer also counted the *number of targets* noticed by the participant. Efficiency of the task was measured as the *response time* to the second, interrupting message, timed from presentation of the notification to accepting or declining the message. *Incident handling time* was calculated by subtracting the time spent on navigation from the total time on task to compensate for differences in walking speed. After every message, participants rated *message interruptiveness* on a scale from 1 (not interruptive) to 7 (highly interruptive) on the PDA.

*After the experimental session* participants rated their *experienced workload* using the NASA Task Load Index (TLX; Hart & Staveland, 1988). Participants filled out the *user*

*experience questionnaire* containing 16 statements about working with the prototype (e.g. “the notification system interrupts me too much” or “the notification system is easy to use”). In addition, four rating scales were filled out, concerning the disruption and supportiveness of the system, the extent to which the system aided awareness of messages and participants’ satisfaction with the system. Finally, four open questions about improvements to the prototype concluded the experiment.

**Table 5. Measures and variables in the experiment.**

<b>Phase</b>	<b>Measure</b>	<b>Variable</b>
<i>Before</i>	Individual characteristics	Age, Gender, Computer experience, Task switching ability, Memory score
<i>During</i>	Effectiveness of patrol task	Number of read errors, Number of handling errors, Number of targets noticed
	Efficiency of patrol task	Response time, Incident handling time
	Subjective judgments	Message interruptiveness
<i>After</i>	Subjective judgments	Workload, System disruption, System supportiveness, Awareness of messages, Satisfaction

## **Apparatus**

The prototype notification system was programmed using the Microsoft .NET framework and implemented on a HP IPAQ handheld computer. This device had a stylus-based touch-screen with a resolution of 320 x 240 pixels. The test leader accompanying the participant used a Tablet PC and a peer-to-peer wireless connection to send the messages to the handheld computer at predefined intervals as unobtrusively as possible. For the NASA TLX and the memory test, a laptop computer was used. All questionnaires and the TMT test were administered on paper.

## **Procedure**

The experiment was performed individually by all participants and took between 90 and 120 minutes to complete. Participants were told they had to perform a patrol task through the building, while using a prototype notification system. They then signed an informed consent form and the individual characteristics questionnaire and tests were administered. Participants familiarized themselves with the floor plans on the PDA and followed the patrol route once, accompanied by the test leader. Subsequently, they were trained on recognition of the targets, incident locations and notification styles depending on the experimental condition. They then performed the patrol task as quickly and accurately as possible, accompanied by the test leader. Hereafter, they filled out the NASA TLX and the questionnaires.

## Statistical Analyses

All data were checked for normality and significant outliers ( $> 2.5$  SD from the mean) were omitted from the data set. Multivariate ANOVA was performed on all performance variables and interruptiveness scores, with “condition” as a four-level between subjects factor and “priority level” as a three level within subjects factor. Post-hoc Bonferroni comparisons between conditions and between priority levels were performed for a detailed analysis. The questionnaires and rating scales were analyzed using non-parametric Kruskal-Wallis H-tests.

## RESULTS

Results are presented separately for patrol task effectiveness and efficiency, workload and subjective measures. An overview of means for all variables per condition (full message (F), postpone (P), indicator (I), and adaptive (A)) is presented in Table 6. No significant differences were found between participants in the four conditions for age, computer experience, task switching ability and memory score.

**Table 6. Mean results per condition on task performance variables and message interruptiveness (MI).**

Condition	Patrol task effectiveness and efficiency					MI
	Read errors (#)	Handling errors (#)	Targets (#)	Response time (s)	Incident handling time (s)	
Full message (F)	1.5	0.3	8.6	10.2	178	4.6
Postpone (P)	3.1	2.0	10.8	12.6	172	3.1
Indicator (I)	1.4	0.4	6.8	15.2	176	3.7
Adaptive (A)	0.5	0.1	8.5	17.2	181	3.5

### Patrol Task Effectiveness

Effectiveness of the patrol task was measured as the number of read errors (errors in ignoring or attending to a message), handling errors (errors in deciding to handle an incident) and number of targets noticed along the route. The total number of *read errors* showed a significant effect of condition ( $F(3, 28) = 14.3, p = 0.000008$ ; see Fig. 6). Postponing messages resulted in 3.1 errors on average, significantly more than in the adaptive condition ( $M_A = 0.5; p = 0.000004$ ) and in the indicator condition ( $M_I = 1.4; p = 0.001$ ). The full message condition counted 1.5 read errors, intermediate to (but not significantly different from) the other three conditions.

Similarly, the total number of *handling errors* showed a main effect of condition ( $F(3, 28) = 27.8, p = 0.000001$ ; see Fig. 7). Again, participants in the postpone condition made 2.0 errors on average, significantly more than in the adaptive ( $M_A = 0.1; p < 0.000001$ ), full message ( $M_F = 0.3; p < 0.000001$ ) and indicator ( $M_I = 0.4; p = 0.000001$ ) conditions. These last three conditions did not differ significantly. As expected, postponing messages

resulted in a high number of read and handling errors, while the full message condition showed an intermediate number of read errors. Adaptive condition showed the lowest number of both read errors and handling errors.

For number of *targets* noticed, an overall significant difference between conditions was found ( $F(3, 28) = 3.48, p = 0.03$ ; see Fig. 8). Post-hoc analysis showed that significantly more targets were noticed in the postpone condition ( $M_P = 10.8$ ), compared to the indicator condition ( $M_I = 6.8$ ) ( $p = 0.02$ ). The full message and adaptive conditions resulted in a similar number of targets noticed (8.5 and 8.6 respectively) but not significantly different from the other conditions. Thus, as expected, postponing messages maintained awareness of the environment, resulting in a high number of targets noticed.

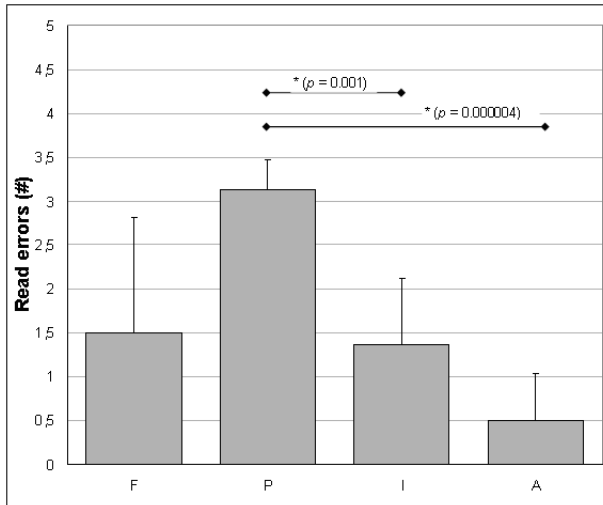


Fig. 6. Mean number of read errors per condition.

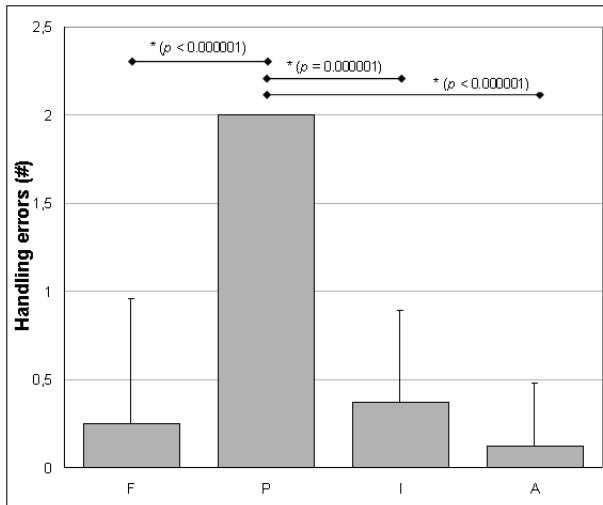
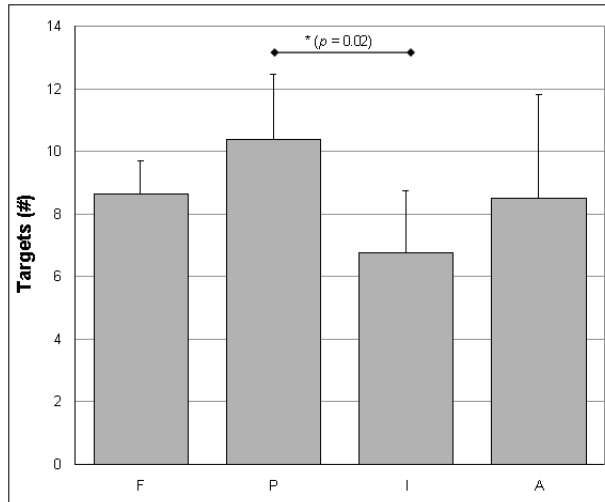


Fig. 7. Mean number of handling errors per condition.



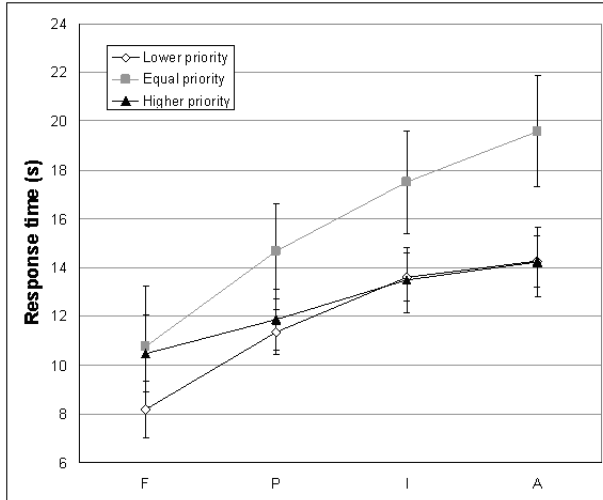


**Fig. 8. Mean number of targets noticed per condition.**

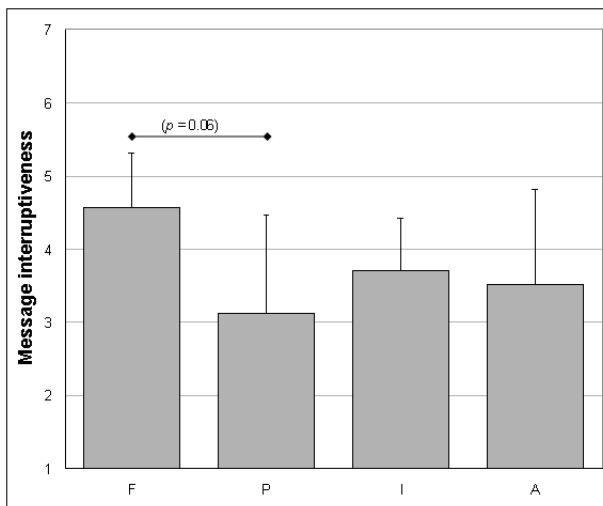
### Patrol Task Efficiency

Efficiency was measured as the response time to interrupting messages and the incident handling time. Response time was analyzed with repeated measures ANOVA per condition and per priority level (lower, equal and higher priority). A significant main effect of condition was found ( $F(3, 22) = 3.90, p = 0.02$ ; see Fig. 9). Post-hoc analysis showed response time to be significantly longer in the adaptive condition ( $M_A = 16.0$  s), compared to the full message condition ( $M_F = 9.8$ ) ( $p = 0.02$ ). No significant differences between the other conditions were found. In addition, a significant main effect of priority was found ( $F(2, 44) = 11.95, p = 0.00007$ ). Overall, people responded faster to lower (12.0 s) and higher (12.5 s) priority messages than to equal (15.8 s) priority messages. Presumably, the decision to attend or ignore a message was harder for equal priority messages, thereby increasing response time. The interaction effect between condition and priority was not significant ( $F(6, 44) = 0.78, p = 0.59$ ). Using different notification styles did not make people respond faster or slower to different priority messages. Overall, adaptive notification increases response time more than the uniform notification styles in the other three conditions.

*Incident handling time* means were very similar in the four conditions, around 170-180 seconds. The differences between conditions were not significant ( $F(3, 28) = 0.397, p = 0.76$ ; see Table 6). When incident handling time was analyzed per priority level, again no significant differences were found. This was contrary to what was hypothesized.



**Fig. 9. Mean response time to interrupting message per condition. Separate lines indicate priority level.**



**Fig. 10. Mean message interruptiveness scores per condition.**

### Message Interruptiveness

Message interruptiveness scores showed a trend that approached significance ( $F(3, 28) = 2.63, p = 0.07$ ; see Fig. 10) between the conditions. Participants in the full message condition rated the messages as more interruptive compared to those in the postpone condition, which had the lowest rating ( $M_F = 4.6$  vs.  $M_P = 3.1$ ;  $p = 0.06$ ). The adaptive and indicator conditions resulted in intermediate interruptiveness ratings ( $M_A = 3.5$  and  $M_I = 3.7$ ) and not significantly different from the other two conditions. Although the differences in message interruptiveness are not strictly significant, the p-values of 0.06 and 0.07 do represent a strong trend in the hypothesized direction.

When analyzed per priority level, the data on the interruptiveness scale showed a significant main effect of priority ( $F(2, 52) = 28.8, p < 0.000001$ ). The interrupting higher

priority messages were rated as significantly more interruptive than equal priority ( $p = 0.00002$ ) or lower priority messages ( $p < 0.000001$ ).

## Workload

NASA TLX scores were lower in the postpone condition ( $M_P = 47.8$ ) compared to the other conditions ( $M_F = 56.4$ ,  $M_I = 59.8$  and  $M_A = 59.4$ ). However, this difference in workload scores between the conditions was not significant ( $F(3, 28) = 1.35$ ,  $p = 0.28$ ).

## User Experience

The data on four of the 16 statements from the user experience questionnaire showed overall significant differences between conditions (all  $p < 0.05$ ; see Table 7, upper part). These four were further analyzed with multiple comparisons of mean ranks (see Table 7). The full message condition was considered significantly more interruptive ( $M_F = 4.0$ ) than the postpone condition ( $M_P = 2.6$ ) or adaptive condition ( $M_A = 2.8$ ) ( $p = 0.004$ ). None of the four rating scales on disruption, support, awareness and satisfaction showed significant differences between conditions. Remarkably, the full message condition scored highest on the satisfaction ratings ( $M_F = 102$ ; not significant), probably because participants were able to recognize the messages better in this condition compared to the adaptive condition.

**Table 7. Mean scores on the questionnaire items and rating scales per condition. A higher score (from 1 to 6) represents more agreement with the statement. A higher score on the rating scales (from 0 to 120) represents a more positive rating.**

Statement	F	P	I	A
The notification system is easy to use	5.4	5.4	4.3	4.8
The notification system prevents interruption	1.5	3.6	2.6	2.4
The notification system interrupts me too much	4.0	2.6	3.3	2.8
I can recognize message priority by the sound	5.8	4.4	4.3	4.0
Rating scale	F	P	I	A
How disruptive was the notification system?	55	70	55	63
How supportive was the notification system?	88	74	92	87
How aware were you of notifications?	108	93	103	99
How satisfied were you with the notification system?	102	85	80	92

After the experimental session, participants were asked how the system could be made less interruptive and whether message priority or activity should be taken into account for notification presentation. Their answers corresponded with the design decisions on which the prototype system was based. Participants in the full message condition would like equal or lower priority messages postponed until they were finished with an incident. Their solutions would be to “use icons” or “just play a sound” to minimize disruption. However, participants in the indicator condition were not satisfied with this design solution. Indicators were easily overlooked or forgotten and required more interface

actions (clicking the icon). Participants in the postpone condition were concerned about missing high priority messages and would like to be notified of these messages with an auditory signal. Finally, participants in the adaptive condition indicated that they were satisfied with the presentation moment and the interruptiveness of the notifications. Two participants indicated that trying to understand the adaptive system behavior caused higher workload. In conclusion, remarks made by participants in post-experimental questionnaires supported the design solutions to postpone notifications based on availability and match notification salience to message priority and user activity.

### **Comparing the Notification Styles**

When the different notification styles are compared across all results, the hypothesized costs and benefits of each notification style become apparent (see also the hypotheses in the “Evaluation” section). As expected, full message presentation maintained awareness of messages, resulting in fast responses to messages. However, this fast response is not always appropriate (e.g. attending to a low priority message when engaged in a high priority incident) thereby leading to an intermediate number of decision errors. Full messages increased message interruptiveness more than the other conditions.

Postponing messages maintains awareness of the environment, demonstrated by the highest number of targets noticed and lowest message interruptiveness. However, postponing messages comes at the cost of high error rates in attending to messages and handling incidents. There was a trend towards lowest workload in the postpone condition (not significant).

Presenting incident messages as indicators maintained awareness of messages, resulting in low error rates. However, indicators still caused unwanted interruption away from the environment, resulting in the lowest number of targets to be noticed. In addition, participants did not prefer indicators as they were forgotten or overlooked.

Adaptive notification causes the lowest number of decision errors and message interruptiveness was rated as low as in the postpone condition, demonstrating that adaptive notification provides appropriate interruption and does not decrease awareness of the environment. This comes at the cost of slightly higher response time to incident messages.

## **FOLLOW-UP STUDY**

To validate the results from the first study and increase external validity, we investigated whether results obtained with trained student participants in any way reflect results obtained with experienced police officers. This section describes a summary of a follow-up study, relevant for the current research question. In this study, police teams used an identical context-aware notification system, focusing specifically on the effects of adaptive versus full message notification on task performance. As we needed a way to reliably compare specific notification situations and collect accurate task performance measures, the follow-up study took place in a synthetic task environment. For a detailed description, please see Streefkerk, van Esch-Bussemakers and Neerincx (2009).

The task setup in the follow-up study was similar to the first study, requiring police teams to find targets in their vicinity (cf. awareness of the environment) and handle

incidents in a virtual city environment (see Fig. 11). When an incident occurred, their notification system presented an incident message and police officers decided who should handle the incident. In the adaptive condition, their notification system adapted the notification style of incident messages to user activity and message priority. When a team member had to handle an incident, the full incident message was presented with a salient sound. When he was busy and a new incident was waiting for him, the system presented an indicator with a less salient sound. When he did not have to handle an incident, an indicator was presented without sound. In the control condition, all messages were presented as full messages (uniform notification).



**Fig. 11. Police officer participating in the follow-up study.**

## **Method**

The experimental manipulation focused on the difference between adaptive and full message notification. Eight teams of three experienced police officers (20 male, 4 female, mean age = 33.0 years, SD = 9.9) participated in both conditions. Two experimental scenarios with equal duration and number of incidents (six high and six low priority) were established in close cooperation with two experienced police officers. The patrol task required officers to collect a maximum of 30 targets, represented by barrels that appeared at random locations throughout the environment. Participants were seated behind two 17" monitors, one above another (see Fig. 11). The top monitor displayed the virtual environment and the incident details. The notification system prototype was implemented using a simulated Personal Digital Assistant (PDA) on a touch screen monitor. Task performance was measured as the number of targets collected, response time to incident messages, errors in decision making on incident handling, and incident handling time. In addition, workload measures were collected using the Rating Scale Mental Effort (Zijlstra, Roe, Leonora & Krediet, 1999) and subjective ratings were collected with a preference questionnaire after each condition. In total, the experiment took about three hours to complete; the two experimental sessions took about twenty minutes each.

## Results

Data on all performance variables was averaged and compared per condition using dependent samples t-tests and repeated measures ANOVA. The results are remarkably similar to the results obtained in the first experiment with non-professionals. On average, more *targets* were collected in the adaptive condition ( $M = 18.5$ ) compared to the control condition ( $M = 17.4$ ). However, this difference was not significant ( $t(7) = -0.44, p = 0.67$ ). Adaptive notification caused a slightly (but not significantly) longer response time to messages than full message notification. When response time was analyzed for high and low priority messages separately, the interaction effect of condition and priority approached significance ( $F(1, 7) = 4.32, p = 0.076$ ; see Fig. 12). With full message notification, response time to low and high priority incidents was almost identical, while using adaptive notification, police officers' response time was appropriate for the message priority (longer for low priority, shorter for high priority messages). Adaptive notification lead to less decision errors on incident handling ( $M = 3.4$ ) than full message notification ( $M = 5.0$ ), this effect approached significance ( $t(7) = 2.09, p = 0.07$ ; see Fig. 13) which is an even stronger result than in the first study. Similar to the first study, adaptive notification did not decrease incident handling time or influence workload measurably. Importantly, the majority of police officers (76%) preferred this adaptive support in their daily work. Again similar to the first study, more than half of them (58%) commented negatively on the use of indicators without sound (see Streefkerk et al., 2009, for a full report of the results).

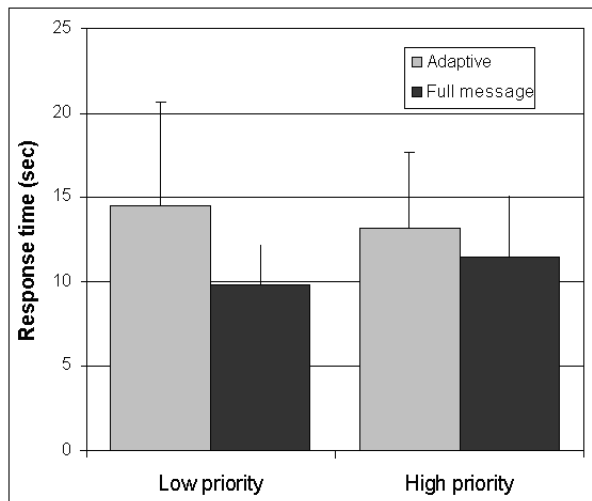


Fig. 12. Mean response time to low and high priority messages per condition.

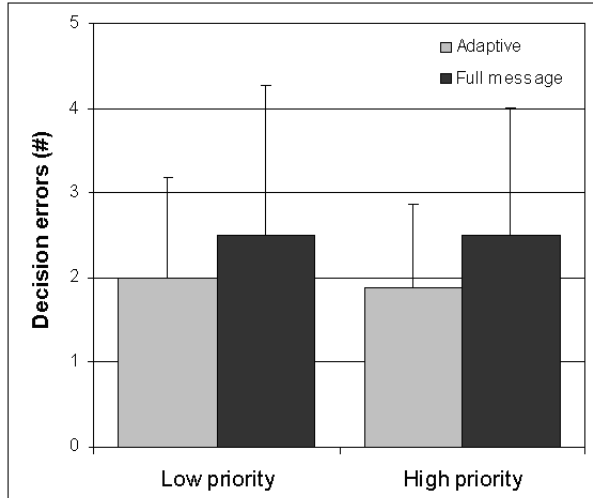


Fig. 13. Mean number of decision errors on low and high priority messages per condition.

## GENERAL DISCUSSION

This paper investigated the effects of different notification styles on awareness of the environment and awareness of incoming messages on a mobile device. To this end, a mobile notification system adapted the timing and appearance of incident messages, based on user activity and message priority. As a first step, a controlled mobile experiment with trained student participants measured task performance, workload and the user experience with this system. Four different notification style conditions (full message, postpone, indicator or adaptive) were compared. We found partial support for each of the four hypotheses, and the direction of the observed effects corresponds to the hypotheses (with two exceptions). Table 8 summarizes the observed effects of each notification style in relation to the two goals of the notification system: maintaining awareness of the environment and of incident messages.

Table 8. Observed effects of notification styles on awareness of environment and awareness of incident messages (ns = no significant effect).

Notification styles	Awareness of environment			Awareness of incident messages	
	<i>Number of targets</i>	<i>Message interruptiveness</i>	<i>Incident handling time</i>	<i>Decision errors</i>	<i>Response time</i>
Full message (F)	ns	High	ns	Intermediate	Short
Postpone (P)	High	Low	ns	High	N/A
Indicator (I)	Low <sup>a</sup>	ns	ns	Low	ns
Adaptive (A)	ns	ns	ns	Low	Long <sup>a</sup>

<sup>a</sup> this effect is different than hypothesized.

The results from the first study show that presenting incident messages as full messages facilitates a quick response, but increases interruption: they are considered interruptive and people respond incorrectly to lower priority messages. Postponing all messages to a moment when users are available maintains awareness of the environment, but decreases awareness of messages, leading to significantly more decision errors than other styles. Indicators decrease awareness of the environment more than expected, resulting in fewer targets to be noticed than the other styles. This is presumably due to more actions required from users. However, indicators keep people informed of messages, leading to a low number of decision errors. Adaptive notification maintains awareness of incoming messages without decreasing awareness of the environment. This comes at the cost of longer response time, presumably due to unfamiliarity with the adaptive behavior of the system (e.g. varying the notification styles). User preference for this adaptive behavior corresponds with the design choices implemented in the prototype system.

These results are corroborated by a follow-up study, employing experienced police officers in a similar setup. The follow-up study found that adaptive notification caused increased response time (but appropriate for the message priority) and less decision errors than presenting full messages. In addition, police officers preferred the adaptive notification system over a non-adaptive system. Taking the results from the two studies together, it seems adaptive notification is appropriate for improving the *right* response to messages, and full messages are good for *faster* response to messages. However, this comes at the cost of higher interruption and more inappropriate responses to messages (e.g. reading a low priority message while busy with a high priority incident). This seems logical as adaptive notification provides more information cues (saliency, information density) on which to base the decision whether a message is relevant at the moment of notification.

Both studies address the gap in empirical work on mobile, context-aware notification systems in real world tasks. We demonstrated that a set of notification rules could determine appropriate timing and appearance of notification messages. Adaptive notification has slightly positive effects on task performance and the user experience in a (mobile) patrol task for both non-professionals and professionals. Results from these studies emphasize the positive influences of appropriate timing of interruptions found in other domains (e.g. desktop computing) (Bailey & Konstan, 2006; Cutrell et al., 2001; Iqbal & Bailey, 2008). They provide further evidence that postponing or deferring interruptions until users are available helps mitigate negative influences of interruptions (Iqbal & Bailey, 2008). Additionally, the decrease in number of task errors found in earlier work is replicated here (e.g. Bailey & Konstan, 2006). Concerning the awareness trade-off, our results implicate that designers of context-aware notification systems should use full messages when awareness of messages needs to be high and a fast response is required. They should postpone messages when users' attention needs to be focused on the environment. Adaptive notification seems less suited to be used when time pressure is high (cf. increased response time). Drawbacks to the use of icons on mobile devices are that they are sometimes overlooked, forgotten and require more display manipulations.

In the first experiment, as well as in the police follow-up study, we did not find positive effects of adaptive notification on time on task (incident handling time) as



reported elsewhere (Bailey & Konstan, 2006; Iqbal & Bailey, 2008). Nor did we find effects of adaptive notification on workload. The absence of significant effects might be explained by the manipulation: notification presentation in the adaptive condition necessarily has some overlap with the uniform conditions (see also the notification matrix in Table 2). In addition, relatively long task durations (over 170 s) and the required between-subjects setup of the first study could have masked differences between the conditions. Hence, this paper leaves a number of questions still open, specifically regarding the influence of notification styles on workload and time on task.

An important limitation of the first study was that the patrol task was necessarily a simplification of actual police work, to systematically investigate the awareness trade-off. This was a first step in the iterative design approach of our notification system, as explained in the introduction. The use of audio and written descriptions of incidents might have influenced the level of engagement of the participants in the patrol task. In real police patrol, emotional state and danger would certainly influence how notifications are received. In addition, professional end-users might be more experienced in dividing their attention between the environment and incoming messages. However, the follow-up study provides evidence that the same effects of adaptive notification hold for professional end-users as well as non-professionals. As such, we believe our current results can be applied to the police domain with care. We must stress the need to further test the concept of adaptive notification in actual domains with professional end-users.

A practical implication of this work is that notification presentation in operational contexts (such as police patrol, military patrol, and Urban Search and Rescue) can benefit from taking into account user activity and message priority. The current work shows how location-based notification in these domains can be made less interruptive by considering additional factors such as officer activity and message priority (Streefkerk et al., 2008a). Mobile notification systems can be implemented that estimate these factors based on readily available information in the domain (location sensing, priority categorization, communication signals and user actions). In the future, such systems provide appropriately timed interruptions via the appropriate modality, reducing the risk of unwanted interruption for police officers and other mobile professionals.

## **CONCLUSION**

Staying aware of your direct environment and incoming messages on a mobile device is a fundamental challenge in mobile HCI. The current paper contributes to a solution to this challenge, by stating a design rationale on how appropriate timing and visual appearance of notifications can be realized, based on message priority and user activity. Four notification styles were compared in a mobile, task-relevant setting to assess their effects on task performance and user experience. The results of this first study demonstrate the benefits and drawbacks of the different notification styles (see also “Comparing the notification styles”) that were validated in a follow-up study with police officers. Full messages facilitate a quick response to the message at the cost of unwanted interruption, while postponing messages diminishes interruption but also diminishes awareness of messages. An adaptive notification system supports effectiveness of mobile patrol in terms of errors and the user experience. Although adaptive notification increased response time to messages, this was only for lower priority messages. These results

provide a foundation for further design and field evaluation of these systems with end-users. Based on these results, employing context-aware notification systems in operational police contexts is expected to support the effectiveness of patrol tasks.

## ACKNOWLEDGEMENTS

This research within the MultimediaN project is sponsored by the Dutch Ministry of Economic Affairs, the Fulbright Scholarship Program and the TNO Defence Research Scholarship Program. We would like to thank Bert Bierman and the staff at the Computer Science Department at Virginia Tech for their support. We extend our sincere gratitude to Woodrow Winchester III, Brad Davis and coworkers at Virginia Tech for the use of their facilities and their help in the experiment.

## REFERENCES

- Adamczyk, P.D., & Bailey, B.P. (2004). If Not Now, When?: The Effects of Interruption at Different Moments Within Task Execution. In *Proceedings of CHI 2004* (pp. 271-278). Vienna, Austria.
- Andrews, A.E., Ratwani, R.M., & Trafton, J.G. (2009). The Effect of Alert Type to an Interruption on Primary Task Resumption. In *Proceedings of the HFES Annual Meeting 2009*. Santa Monica: HFES.
- Bailey, B.P., & Iqbal, S.T. (2008). Understanding Changes in Mental Workload during Execution of Goal-Directed Tasks and Its Application for Interruption Management. *ACM Transactions on Computer-Human Interaction* 14(4), Article 21.
- Bailey, B.P., & Konstan, J.A. (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, 685-708.
- Cutrell, E., Czerwinski, M., & Horvitz, E. (2001). Notification, disruption and memory: effects of messaging interruptions on memory and performance. In M. Hirose (Ed.), *Proceedings of Interact 2001 Conference* (pp. 263-269). Tokyo, Japan: IOS Press.
- Ebling, M., Hunt, G. D. H., & Lei, H. (2001). *Issues for context services for pervasive computing*. Paper presented at Middleware 2001 Workshop on Middleware for Mobile Computing, Heidelberg.
- Fogarty, J., Hudson, S. E., Atkeson, C. G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J. C., & Yang, J. (2005). Predicting human interruptibility with sensors. *ACM Transactions on Computer-Human Interaction* 12(1), 119-146.
- Gievska, S., & Sibert, J. (2005). Using task context variables for selecting the best timing for interrupting users. In *Proceedings of the 2005 Joint Conference on Smart Objects and Ambient Intelligence: Innovative Context-Aware Services: Usages and Technologies* (pp. 171-176). New York: ACM Press.
- Hart, S., & Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). North Holland, Amsterdam.
- Ho, J., & Intille, S.S. (2005). Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. In *Proceedings of CHI'05 Conference* (pp. 909-918). New York: ACM Press.
- Hopp, P.J., Smith, C.A.P., Clegg, B.A., & Heggestad, E.D. (2005). Interruption management: the use of attention-directing tactile cues. *Human Factors* 47, 1-11.
- Horvitz, E., Apacible, J., & Subramani, M. (2005). Balancing awareness and interruption: Investigation of notification deferral policies. In L. Ardissono, P. Brna, & A. Mitrovic (Eds.), *User Modeling 2005: Proceedings of the 10th International Conference (UM 2005)* (pp. 433-437). Berlin: Springer-Verlag.
- Horvitz, E., Kadie, C., Paek, T., & Hovel, D. (2003). Models of attention in computing and communication: From principles to applications. *Communications of ACM* 46(3), 52-59.
- Iqbal, S.T., & Bailey, B.P. (2008). Effects of Intelligent Notification Management on Users and Their Tasks. In *Proceedings of CHI2008* (pp. 93-102). Florence, Italy.

- Jameson, A., Schäfer, R., Weis, T., Berthold, A., & Weyrath, T. (1999). Making systems sensitive to the user's time and working memory constraints. In *Proceedings of the 4th international conference on Intelligent User Interfaces* (pp. 79-86). New York: ACM.
- Kern, N., & Schiele, B. (2003). Context-aware notification for wearable computing. In *Proceedings of 7th International Symposium on Wearable Computers (ISWC)* (pp. 223-230).
- Kjeldskov, J., & Graham, C. (2003). A review of mobile HCI research methods. In L. Chittaro (Ed.), *Proceedings of Mobile HCI 2003 Conference* (pp. 317-335). Berlin: Springer-Verlag.
- Kort, J., & De Poot, H. (2005). Usage analysis: Combining logging and qualitative methods. In *Proceedings of CHI 2005 extended abstracts* (pp. 2121-2122).
- Kostov, V., Tajima, T., Naito, E., & Ozawa, J. (2006). Analysis of appropriate timing for information notification based on indoor user's location transition. In *Proceedings of the Fourth Annual IEEE International Conference on Pervasive Computing and Communications (PerCom'06)* (pp. 245-250). Los Alamitos: IEEE Computer Society.
- Nagata, S. (2003). Multitasking and Interruptions During Mobile Web Tasks. In *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 1341-1346). St. Louis, Missouri: Mira Digital Publishing.
- McCrickard, D.S., Catrambone, R., Chewar, C., & Stasko, J. (2003). Establishing tradeoffs that leverage attention for utility: empirically evaluating information display in notification systems. *International Journal of Human-Computer Studies* 58, 547-582.
- McCrickard, D.S., & Chewar, C.M. (2003). Attuning Notification Design to User Goals and Attention Costs. *Communications of the ACM* 46(3), 67-72.
- McFarlane, D.C. (2002). Comparison of Four Primary Methods for Coordinating the Interruption of People in Human-Computer Interaction. *Human Computer Interaction* 17, 63-139.
- Miner, T., & Ferraro, F.R. (1998). The Role of Speed of Processing, Inhibitory Mechanisms, and Presentation Order in Trail-Making Test Performance. *Brain and Cognition* 38(2), 246-253.
- Nagata, S.F. (2003). Multitasking and interruptions during mobile web tasks. In *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 1341-1345). Santa Monica: Human Factors and Ergonomics Society.
- Neerincx, M.A., & Lindenberg, J. (2008). Situated Cognitive Engineering for Complex Task Environments. In J.M.C. Schraagen et al. (Eds.), *Naturalistic Decision Making and Macrocognition* (pp. 373-390). Aldershot: Ashgate Publishing.
- Neerincx, M.A., Pemberton, S., Lindenberg, J., & van Besouw, N.J.P. (1999). *U-WISH Attuning Network User Interfaces to individual capacities: User requirements and support concepts* (Tech. Rep. TM-99-C056). Soesterberg, the Netherlands: TNO Human Factors.
- Oulasvirta, A., Tamminen, S., Roto, V., & Kuorelahti, J. (2005). Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In *Human Factors in Computing Systems: Proceedings of CHI 2005* (pp. 919-928). New York: ACM Press.
- Sawhney, N., & Schmandt, C. (2000). Nomadic Radio: Speech and Audio Interaction for Contextual Messaging in Nomadic Environments. *ACM Transactions on Computer Human Interaction* 7(3), 353-383.
- Sørensen, C., & Pica, D., (2005). Tales from the police: Rhythms of interaction with mobile technologies. *Information and Organization*, 15(2), 125-149.
- Smets, N.J.J.M., van Diggelen, J. Neerincx, M.A., Bradshaw, J.M., Jonker, C.M., de Rijk, L.J.V., Senster, P.A.M., ten Thijs, O. & Sierhuis, M. (2010). Assessing Human-Agent Teams for Future Space Missions. *IEEE Intelligent Systems*, 25, 2-9.
- Streefkerk, J.W., van Esch-Bussemaekers, M.P., & Neerincx, M.A. (2006). Designing Personal Attentive User Interfaces in the Mobile Public Safety Domain. *Computers in Human Behavior* 22, 749-770.
- Streefkerk, J.W., van Esch-Bussemaekers, M.P., & Neerincx, M.A. (2007). Context-aware notification for mobile police officers. In D. Harris (Ed.), *Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562* (pp. 436-445). Berlin: Springer-Verlag.
- Streefkerk, J.W., van Esch-Bussemaekers, M.P., & Neerincx, M.A. (2008a). *User Evaluation Attentive services [Gebruikersevaluatie Attenderingsservice]* (Tech. Rep. TNO-DV 2008 D290 [in Dutch]). Soesterberg, the Netherlands: TNO Defence, Security and Safety.
- Streefkerk, J.W., van Esch-Bussemaekers, M.P., & Neerincx, M.A. (2009). Context-aware Task Allocation and Notification to Support Police Team Performance. In *Foundations of Augmented Cognition, Proceedings of the HCI International 2009 Conference* (pp. 88-97). Berlin: Springer Verlag.

- Streefkerk, J.W., van Esch-Bussemakers, M.P., Neerinx, M.A., & Looije, R. (2008b). Evaluating Context-Aware Mobile Interfaces for Professionals. In J. Lumsden (Ed.), *Handbook of Research on User Interface Design and Evaluation for Mobile Technology* (pp. 759-779). IDEA group.
- Zijlstra, F.R.H., Roe, R.A., Leonora, A.B., & Krediet, I. (1999). Temporal factors in mental work: Effects of interrupted activities. *Journal of Occupational and Organizational Psychology* 72 (2), 163–185.