DOING THE RIGHT TASK

Context-Aware Notification for Mobile Police Teams

DOING THE RIGHT TASK

Context-Aware Notification for Mobile Police Teams

PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Technische Universiteit Delft,

op gezag van de Rector Magnificus prof.ir. K.C.A.M. Luyben,

voorzitter van het College voor Promoties,

in het openbaar te verdedigen op vrijdag 20 mei 2011 om 10:00 uur

door Jan Willem STREEFKERK

doctorandus in de Cognitieve Psychologie

geboren te Heemskerk.

Dit proefschrift is goedgekeurd door de promotor:
Prof.dr. M.A. Neerincx

Samenstelling promotiecommissie:
Rector Magnificus, voorzitter
Prof.dr. M.A. Neerincx, Technische Universiteit Delft / TNO, promotor
Prof.dr. G.C. van der Veer, Vrije Universiteit Amsterdam
Prof.dr. J.M.C. Schraagen, Technische Universiteit Twente
Prof.dr. C.M. Jonker, Technische Universiteit Delft
Prof.dr. H. de Ridder, Technische Universiteit Delft
D.S. McCrickard, PhD, Virginia Polytechnic Institute and State University, USA
Dr. M.P. van Esch-Bussemakers, TNO

Het onderzoek in dit proefschrift is uitgevoerd binnen het MultimediaN project, gesubsidieerd door het ministerie van Economische Zaken.
The research in this thesis is conducted within the MultimediaN project, sponsored by the Dutch Ministry of Economic Affairs.

Tekst en lay-out: Jan Willem Streefkerk
Illustraties: Anna Verhoeven © 2011
Omslagontwerp en druk: Uitgeverij BOXPress, Oisterwijk | proefschriftmaken.nl
ISBN: 978-90-8891-266-5

TABLE OF CONTENTS

In this thesis, wherever “he” refers to mobile professionals, this can also be read as “she”.

List of Figures

List of Tables

1 INTRODUCTION

1.1 Problem statement

Mobile professionals, such as police officers, are challenged by limited means for mobile information exchange and difficulties in team collaboration across distances. Current mobile technology is not properly designed to address these challenges. First, because the technology does not support the dynamic task goals and demands in these high-risk, critical work environments. Second, because the technology does not adapt to the dynamic mobile use context, the momentary workload and attention and individual characteristics. We propose a context-aware mobile support system that takes into account operational demands and Human Factors. It subsequently adapts information presentation (e.g. notification styles) and task allocation based on this knowledge. This is expected to result in optimized (team) task performance, specifically for mobile information exchange and team collaboration. However, as these innovative systems are not yet widespread, we must find out how to design and validate these systems in mobile professional settings and whether the expected benefits are realized. Consequently, the main research question of this thesis reads:

> *Which features of context-aware mobile systems can support team task performance and optimize the user experience, specifically for mobile information exchange and team collaboration in professional domains?*

1.2 Mobile professional domains

Mobile professionals such as police officers, first responders and security officers maintain order and safety in public spaces. They are often on patrol as a physically distributed group, with team members at different geographical locations. When an incident occurs, they proceed to the incident location to collaborate with team members and handle the incident (Nulden, 2003). To do this efficiently, information exchange is required such as communication with team members and an emergency room. Currently, mobile information and communication technology, such as portable radios and mobile computers, is applied to support these work activities (Baber, Haniff, Sharples, Boardman, & Price, 2001; Ioimo & Aronson, 2004; Kun, Miller III, & Lenharth, 2004; Sorensen & Pica, 2005).

However, often these systems do not meet the high operational demands posed by the dynamic and critical environments they are used in (Kun et al., 2004; Marcus & Gasperini, 2006; Bouwman, Haaker, & de Vos, 2008). Occurring incidents cause tasks goals and information needs of mobile professionals to change over time and place. Geographical distribution of team members makes it difficult to keep track of their location and activities, thereby decreasing team situation awareness (Ferscha, 2000; Endsley, Bolte, & Jones, 2003; Carroll, Rosson, Convertino, & Ganoe, 2006). In addition, time pressure and high workload situations influence the amount of attention these professionals can dedicate to interaction with a mobile device (McFarlane, 2002; McCrickard, Catrambone, Chewar, & Stasko, 2003a; Wickens, 2008; Bailey & Iqbal, 2008). Finally, user characteristics such as memory, spatial ability and expertise influence how these systems are used (Carroll, 1993). As these operational demands are not taken into account, human-system interaction is hampered by system designs that are not optimized for these tasks or use context. This causes decision errors, slower response time and potentially dangerous situations for mobile professionals (Bailey, Konstan, & Carlis, 2000; Marcus et al., 2006; Iqbal & Horvitz, 2007).

Table 1-1. Mobile police surveillance problem scenario.

Mobile police surveillance *Youth officer Jason is on surveillance on a busy Friday evening. He already had to handle a case of domestic violence and a car theft. Via his portable radio, he hears that an aggressive juvenile shoplifter has been detained by a shopkeeper in the main shopping street. This is a high priority call requiring two youth officers. Although he is not in the direct vicinity of this street, Jason isn't sure where his colleagues Paul and Bob are or who else is available with juvenile delinquent expertise. He broadcasts via his radio that he is on his way. Navigating to find the shop, he receives a call concerning the stolen car he handled earlier. Listening to this call, he takes a wrong turn and has to backtrack. Annoyed, Jason turns down his radio to focus on finding the right place. This causes him to miss a call from Paul and Bob acknowledging that they already arrived at the shop. When Jason reaches the shop, he finds that his two colleagues have taken the suspect into custody. As he is not needed anymore, he leaves the scene.*

In the problem scenario in Table 1-1, police officer Jason had to coordinate his actions together with his colleagues in a high workload situation using auditory information provided by a radio while navigating towards the incident location. This demonstrates that appropriate design of the human-system interaction in these domains is not trivial. The challenge in designing these systems is dealing with the mutual dependencies between operational demands (information exchange, team collaboration), the mobile use context (dynamicity, distributed actors), technological constraints (limited interaction possibilities), user cognition (momentary workload and attention) and individual characteristics (memory, spatial ability, expertise) (see Figure 1-1). Our statement is that a mobile system that adapts its information presentation to the operational demands and Human Factors provides better support than a non-adaptive system. Such *context-aware* mobile support systems provide specific operational information to the appropriate team member(s) when and where they need it, via appropriate modalities (Shafer, Brumitt, & Cadiz, 2001; Dey, Abowd, & Salber, 2001; Cheverst, Mitchell, & Davies, 2002; Abowd, Mynatt, & Rodden, 2002). This not only supports task performance (e.g. less decision errors, faster response time) but also optimizes the user experience of these systems (e.g. higher preference, trust and acceptance). As a design solution, should Jason have a support system with an overview of where his colleagues are and what they are doing, this is expected to maintain his team situation awareness, lower his response time and need to communicate and increase the efficiency and effectiveness of surveillance task performance.

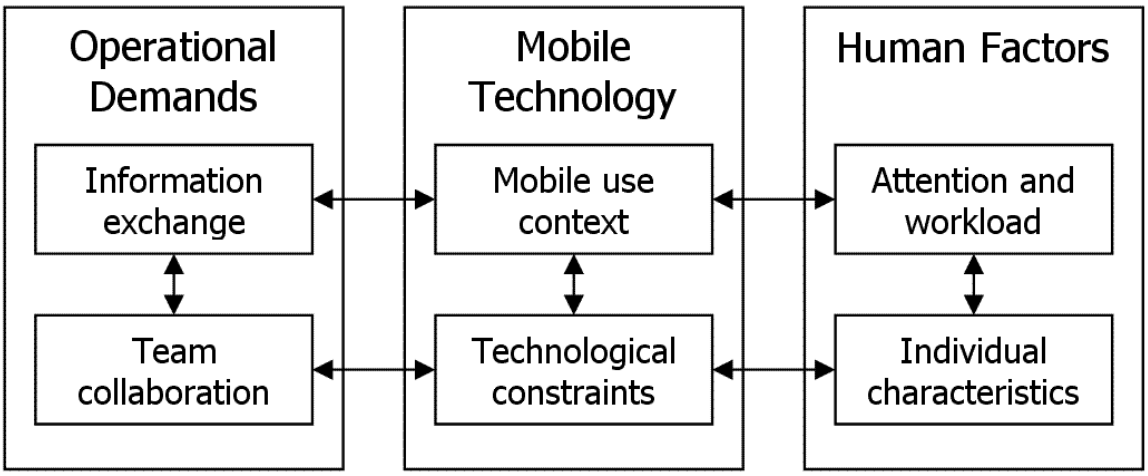


Figure 1-1. Visualizing mutual dependencies between operational demands, mobile technology and Human Factors.

The research in this thesis focuses on designing and evaluating a context-aware support system for (teams of) mobile professionals. The main question is how to attune the interaction and information presentation to the operational demands, to the dynamic mobile use context and to the momentary workload and attentional resources. As police work suffers from the challenges mentioned above, this provides an excellent application domain. To realize such systems, we first need to understand how operational demands, mobile technology and Human Factors issues influence the design of a mobile context-aware support system. Each will be addressed in the next sections.

1.3 Operational demands in mobile professional domains

Efficient work in mobile professional domains becomes more and more dependent on the ability to access and exchange operational information while in the field (Feteke & Kun, 2008). In addition, mobile teamwork requires team members (such as a team of police officers) to collaborate and coordinate actions across distances (Pica & Sørensen, 2004; Carroll et al., 2006). From the systems' design perspective in this thesis, two operational demands are most relevant here: mobile information exchange and team collaboration.

1.3.1 Mobile information exchange

Operational information about incidents needs to reach mobile professionals in an accurate and timely way. As illustrated in the scenario above, the *information need*, defining when and how specific operational information is needed, depends on the situation and changes accordingly (Nulden, 2003; Eijk, Koning, Steen, & Reitsma, 2006). The radio transceiver widely used in mobile domains is used by the emergency room to broadcast incident messages to all people in the field. However, police officers' information needs are not fully met due to three problems. First, incident messages by the emergency room are described too broadly or briefly and lack detail. These details are stored in police databases, to which police officers in the field have no access. Second, the emergency room has little knowledge about the activities of each individual officer. While broadcasting incident messages keeps all team members informed of what is going on, it is sometimes interruptive of police work (Nulden, 2003). For example, police officers have been observed to turn down their radios while in conversation. Third, colleagues exchange information on previous incidents with each other via radio or cell phone. When a specific colleague is unavailable or not on shift, this detailed knowledge is not directly accessible.

1.3.2 Team collaboration

Mobile teamwork requires team members to collaboratively coordinate actions. Due to the mobile context, team members' locations, activities and workload change over time and place, making it difficult to stay aware of distributed team members (Gutwin & Greenberg, 2002). In the scenario above, multiple police officers needed to gather at the incident location. To do this quickly and efficiently requires *team situation awareness*, i.e. being aware of the status, location and activities of colleagues and the operational situation (Ferscha, 2000). Colleagues use their radio or cell phone to keep each other informed. However, this knowledge has to be memorized and updated by each individual team member which is cognitively challenging.

The quality of team situation awareness influences the accuracy of decisions. An important process in policing is *task allocation*, deciding which officer handles which incident (Nulden, 2003; Bouwman et al., 2008). Especially when confronted with high priority incidents, this coordination and allocation process must proceed quickly. Currently, the emergency room assigns incidents to officers, but based on global or outdated knowledge of the whereabouts and activities of these officers. This sometimes results in allocating incidents to officers who are busy or too far away.

1.3.3 Impact of operational demands

The discussion above shows that efficient mobile policing is hampered by unfulfilled, dynamic information needs and limited team situation awareness resulting in inefficient task allocation. These problems affect a large number of police officers carrying out their daily work in the field. In 2008, 32.000 full-time law enforcement officers were employed in the Netherlands and more than 700.000 in the US, averaging about 2 to 2,5 for every 1.000 inhabitants (FBI, 2008). To effectively fight increasing criminality, even more police officers are required.

These problems also have a broader impact on safety in society. Officers who are able to access information in the field can spend more time on the street, an important performance indicator for police organizations. Results from the 2008 Dutch "Safety-monitor" showed that while 44% of the population was very satisfied with the police in general, 52% indicated that police officers often were not noticeably present in the street and only 20% agreed that police fought crime efficiently (CBS, 2008). These figures heavily influence governmental policy of police organizations. In addition, efficient team collaboration and allocation of incidents decreases the risk of escalation. A recent, tragic example was the beach riot at Hoek van Holland in the Netherlands (Police Union, 2009). During a dance party, twenty-one police officers felt threatened by an angry crowd and were not aware if and when backup would arrive, due to limited communication means. Also, the on-scene commander was not fully aware of what was happening, due to unclear organization of command lines. When the crowd did not respond to their warnings, these officers felt compelled to shoot to keep the crowd at a distance, killing a 19-year old boy and wounding several others. Thus, facilitating mobile information exchange and team collaboration will not only impact efficiency and safety of police officers, but also positively impact safety in society. Of course, the organization of command and control (C2) plays a central role in emergency management, but is outside the current scope.

1.4 Mobile police technology

To meet the operational demands outlined above, mobile technology is used in professional domains. Currently, portable radio transceivers are used most commonly in police and other domains, with cellular phones as backup options. Radio technology broadcasts auditory voice communication with the advantage that everybody on one channel hears everybody else. However, radio communication facilitates only one speaker at the same time per channel. As radio technology only has limited bandwidth, cell phones are used for more detailed, person-to-person information exchange. Furthermore, limited bandwidth and limited coverage increases problems in coordination of major incidents, as is demonstrated by the Dutch C2000 communication system for rescue services during the Poldercrash (Safety Region Kennemerland, 2009). Accessing police databases using voice communication always requires a human operator or contact with the district office. Finally, staying aware of the location, status and activities of team members is difficult using only speech interaction via radio.

Introducing mobile computing devices such as Personal Digital Assistants (PDA’s) in the police domain attempts to address the issues of limited information exchange and team collaboration on surveillance (Stijnman, 2004; Van Loon, 2004). Following the recent development towards “Information-Guided Policing”, the Dutch police are developing a number of applications such as MobiPol, Mobile Blue and P-Info (Mente, 2004). Globally, information exchange via mobile devices becomes available for police officers in Spain (K-police), Great Britain (Mobile Information Programme of the National Police Improvement

Agency), the United States (Project54 and the Texas Rangers) and Ireland. These applications and devices use secure internet or satellite communication channels, allowing police officers to access available information while on surveillance (Mente, 2004; Kun et al., 2004; Sorensen et al., 2005). These devices allow multimedia information exchange (video, photo) via multiple modalities, moving beyond "traditional" speech-only interaction. Furthermore, mobile devices increasingly make use of sensor technology (such as Global Positioning System or GPS data) for position estimation. The visual displays on these devices can show geographical maps with locations of incidents and colleagues to support police work (Ioimo et al., 2004). However, while current mobile technology in the police domain increasingly allows mobile information exchange, it does not take into account where the technology is used and who is using it. This results in poorly designed support systems that do not adequately support mobile information exchange and team collaboration (Baber et al., 2001; Stijnman, 2004; Pica et al., 2004; Marcus et al., 2006). Our approach is to design mobile support systems that are *context-aware*, by having them adapt the human-system interaction based on knowledge about the mobile use context and the user.

1.5 Mobile use context

What exactly constitutes "use context" is hard to define, often generically described as everything that influences the mobile human-computer interaction. For our purposes, the use context consists of location (e.g. incident location), task information (e.g. priority), user characteristics (e.g. current availability, expertise) and environmental aspects (e.g. noise, lighting). The mobile use context changes dynamically, influencing the interaction with a support system. For example, the user's information need may change depending on his location. Environmental aspects, such as lighting, noise and movement, influence readability of text on mobile devices or salience of auditory signals.

1.5.1 Adapting to the mobile use context

Mobile context-aware systems make use of this context knowledge to *adapt* the human-system interaction (Dey & Abowd, 2000; Dey et al., 2001; Hong & Landay, 2001; Abowd et al., 2002). The basic assumption is that tuning information presentation to changes in the context will increase its relevance and usefulness (Shafer et al., 2001). Context-aware systems can aid mobile information exchange by matching information presentation to momentary information needs of the user. For example, when Jason had to handle the incident with the juvenile shoplifter, his support system could provide him with specific procedures about handling such an incident. Context-aware systems can also aid team collaboration, for example by using knowledge about user workload or availability to allocate tasks across distributed team members. The main challenges in designing context-aware support systems are which context factors should be used and how the system should subsequently adapt the human-system interaction (Chen & Kotz, 2000; Dey et al., 2001).

1.5.2 Realizing context-aware mobile support

Despite a decade of research on context-aware systems, these systems have not yet been implemented widespread (Chen et al., 2000). A fundamental constraint of these systems is that they are rarely perfect, for example because sensor information can be misinterpreted, outdated or missing (Wickens, 2000). This can result in a mismatch between the actual context and what the system perceives the context to be (i.e. the system's context model). When context-aware systems function reliably, the benefits (e.g. faster response time to incidents) outweigh the costs (e.g. decision errors based on incorrect task allocation). However, due to mismatches, the costs may outweigh the benefits. Decreased task performance, decreased user satisfaction, decision errors and complacency effects have been demonstrated (Rovira, McGarry, & Parasuraman, 2007; Wickens & Dixon, 2007; Gajos, Everitt, Tan, Czerwinski, & Weld, 2008). It is necessary to empirically establish the trade-off of benefits and costs in terms of task performance and user acceptance during actual use. This will help designers to better balance the benefits and costs of context-aware mobile support, which determines in large part whether users will accept and actually use the support (Venkatesh & Davis, 2000; Wiredu, 2007).

1.6 Human Factors

So far, we have focused on how operational demands and the mobile use context influence the design of mobile technology for mobile professional domains. Following theories from mobile Human-Computer Interaction (HCI), we now focus on how this technology can be geared towards the characteristics of the human user, taking into account Human Factors (Dix, Finlay, Abowd, & Beale, 1998). Three mobile HCI challenges stand out when realizing context-aware support in our application domain, related to divided attention, workload and individual characteristics.

1.6.1 Divided attention

When users perform tasks in a mobile work setting, a competition for the user's attention between the mobile device and the environment occurs (Jameson, 2001). In addition, mobile computing is characterized by short, fragmented interactions with a device, such as navigating through the environment, while intermittingly checking the route on your device (Miyata & Norman, 1986; Nagata, 2003; Oulasvirta, Tamminen, Roto, & Kuorelahti, 2005). Due to this competition and fragmentation, distraction and interruption occur frequently, (Adamczyk & Bailey, 2004; Iqbal, Adamczyk, Zheng, & Bailey, 2005; Bailey & Konstan, 2006) causing decision errors and potentially dangerous situations. For instance in the scenario, Jason received a low-priority call on his mobile phone, while navigating to the incident location.

Context-aware systems can help manage such interruptions and focus the user's attention appropriately, for example by presenting notifications. These context-aware notification systems bring valued information to the user's attention, based on reasoning about his current task and context (McFarlane, 2002; Carroll, Neale, Isenhour, Rosson, & McCrickard, 2003; McCrickard et al., 2003a). Context-aware notification systems can also provide team task allocation, for example by notifying team members who will handle which incident. The challenge in designing these systems for the mobile professional domain is *how* to present these notifications and *when* to present them. Notification systems must balance awareness of new information with interruption of an ongoing task (Horvitz, Kadie, Paek, & Hovel, 2003; McCrickard & Chewar, 2003b). The intrusiveness of notifications is influenced by the *notification style*, i.e. the presentation modality, salience and information density of a notification (Sawhney & Schmandt, 2000; Carroll et al., 2003; Kern & Schiele, 2003). In addition, these systems must distinguish appropriate interaction moments, and adapt the timing of a notification to these moments (Fogarty et al., 2005a; Bailey et al., 2006).

1.6.2 Workload

Human cognitive workload in the mobile domain is influenced by the number of tasks that have to be performed, their complexity and their interleaving (Yeh & Wickens, 1988; Parasuraman, Sheridan, & Wickens, 2008). Cognitive workload can change dynamically over time and place and multi-tasking situations occur frequently, especially in mobile computing (Nagata, 2003). This causes specific situations of under- or overload (Young & Stanton, 2002;

Grootjen, Neerincx, & Veltman, 2006). These changes in workload influence the interaction with a mobile device (McFarlane & Latorella, 2002; Neerincx & Streefkerk, 2003; Kjeldskov & Stage, 2004). Mobile computer users have specific strategies to cope with workload changes (i.e. postponing reading an email message). Thus, workload must be mitigated over time for individual users. Individual support to mitigate high workload situations consists of prioritization of tasks and procedures (Neerincx, 2004). Furthermore, workload can be optimized across team members, for example by allocating tasks to team members who are available. The HCI challenge lies in how to design adaptation strategies on mobile devices to support changes in workload for mobile users.

1.6.3 Individual characteristics

Designing context-aware support systems should take into account individual characteristics and capabilities, to meet the needs of individual users (Fischer, 2001). In the scenario, police officer Jason had to navigate to an incident location, while communicating with his colleagues. These tasks address his memory capacity, spatial orientation, task switching ability and information processing capabilities (Carroll, 1993; Neerincx & Lindenberg, 1999; Nagata, 2003). In addition, he might have different police training or experience than his colleagues. These characteristics all influence the human-system interaction with a mobile device and differ across individuals. *Personalization* or adaptation to support individual differences can optimize the interaction, e.g. by individually tailoring and interactively

Table 1-2. Mobile police surveillance scenario illustrating context-aware support.

Context-aware support for mobile police surveillance
Youth officer Jason is on surveillance on a busy Friday evening. He already had to handle a case of domestic violence and a car theft. Via his earpiece, he hears a high-pitched beep: his context-aware support system notifies him of a high priority call. In the display of his mobile device, he reads that an aggressive juvenile shoplifter has been detained by a shopkeeper in the main shopping street. As this call requires two youth officers, the system advises Jason to move to the incident location and indicates that his colleague Paul is on his way. The display shows a geographical map with Paul's current location and estimated time of arrival. Jason notices in the display that youth officer Bob is not available right now, but is in the vicinity as back-up. Because Jason is busy, the support system postpones a low priority call regarding the stolen car he handled earlier. Following the progress of his colleague on the map, Jason quickly reaches the right shop together with Paul. After they finish taking the suspect into custody, the support system notifies Jason that the postponed call is waiting for him.

delivering information (Alpert, Karat, Karat, Brodie, & Vergo, 2003). Interaction with an adaptive system should be understandable for the user, and the user should be able to trust and rely on the system. Therefore, the main challenge in designing such systems is how to design this personalization and validate that the adaptations are appropriate for individual users.

In summary, context-aware support systems adapt their behavior to the dynamic use context to manage divided attention situations, workload challenges and individual user characteristics. How this supports mobile information exchange and team collaboration is illustrated by the scenario in Table 1-2. To address the challenges in realizing this context-aware mobile support, a user-centered design and evaluation approach is employed.

1.7 Design and evaluation methodology

The design of innovative support systems for critical, high-risk environments must meet operational demands and the HCI challenges outlined above. Situated Cognitive Engineering (SCE) is a user-centered design method that iteratively specifies system requirements based on both HCI theory and empirical work (Rosson & Carroll, 2002; Neerincx & Lindenberg, 2008). The design rationales are based on Human Factors theory as well as work domain and support analysis within operational domains (*situatedness*), to incrementally build and test both practical theories and models for adaptive support systems. As not all requirements for these innovative systems are known in advance, the iterative nature of this method allows for adjusting the designs (Preece, Abras, & Maloney-Krichmar, 2004). Furthermore, the method empirically evaluates designs based on domain-relevant criteria with end-users, ensuring that resulting designs meet the needs and characteristics of the intended user group (Rosson et al., 2002). This method is especially suited to develop context-aware mobile support systems in complex environments. It has been previously used to design user-interface support for military, naval and space domains, characterized by the involvement of different actors with different needs in dynamic and critical situations (Grootjen et al., 2006; Neerincx et al., 2006; De Greef, van Dongen, Grootjen, & Lindenberg, 2007; Neerincx, 2010).

The SCE method consists of three processes: analysis, specification and assessment (described in detail in Chapter 0). First, a work domain and support analysis within the police domain identifies how the operational demands of mobile information exchange and team collaboration should be supported. Second, using context modeling and user interface design, the design rationale, concept and design solutions are specified. Context modeling identifies the relevant context factors from the work domain, how they should be sensed and which states these factors can exhibit (Paramythis, Totter, & Stephanidis, 2001). User interface design specifies how the context-aware system should subsequently respond and adapt the human-system interaction, based on changes in the context factors (Dey et al., 2001).

Third, the design solutions are implemented in prototypes to assess whether these support efficiency and effectiveness of (team) task performance and lead to an optimal user experience. As the SCE method is applied iteratively, design solutions move from early semi-functional conceptual prototypes in lab studies to fully functional team support in operational environments. Evaluation setting and methods should be selected and combined based on their benefits and drawbacks (Masthoff, 2002). Furthermore, the changing use context sets high requirements for user evaluation of adaptive systems (Weibelzahl, 2005). There are currently no clear standards for evaluation methods and criteria for mobile context-aware applications. First, because there is a lack in empirical research on these applications. Second, because it is unclear how evaluation methods and criteria optimized for desktop computing translate to mobile professional environments (Kjeldskov & Graham, 2003; Zhang & Adipat, 2005). Thus, in evaluating context-aware mobile support in professional domains, the challenge lies in which methods and criteria to combine and how to tune them to the professional domain.

1.8 Research questions

The main research question stated at the beginning of this chapter can now be specified into separate research questions that will be addressed in the remainder of this thesis. Following the Situated Cognitive Engineering method, prototype systems will be incrementally designed and evaluated in empirical user studies employing laboratory, virtual and real life settings. The effects of context-aware mobile support (CAMS) systems, as specified by the adaptive system behavior in response to operational demands and Human Factors, on mobile surveillance outcome and process measures are studied (see Table 1-3). These outcome and process measures are related to operational demands of mobile information exchange and team collaboration. To address mobile information exchange, we will study the effects of *context-aware notification* (using different styles of appearance and timing) on task performance and the user experience. To address team collaboration, we will study the effects of *task allocation support* as well as context-aware notification on team task performance, team situation awareness and the user experience.

First, we need to analyze the processes in the application domain relevant for mobile information exchange and team collaboration. We need to know what aspects of mobile police surveillance processes need support and how context-aware support can provide meaningful support. This results in a specification of which task, context and user factors are relevant for these processes and requirements for how the support should be designed. Consequently, a work domain and support analysis is performed within the police domain to answer:

1. *Which task, context and user factors in mobile professional domains are relevant to design context-aware support for mobile information exchange and team collaboration?*

As these systems must be used by professionals in demanding environments, thorough evaluation is necessary. To specify appropriate evaluation methods and criteria, we need a framework for selecting, combining and tuning evaluation methods for mobile adaptive systems. This framework explicitly takes into account the mobile professional context. Our evaluation framework will answer:

2. *Which methods and measures should be used to evaluate context-aware mobile support systems in mobile professional domains?*

To address the challenge of **mobile information exchange**, a context-aware mobile support (CAMS) system is designed that provides incidents messages during mobile surveillance. As a first iteration of this support system, we conduct a field study on a location-based notification system, which notifies mobile police officers of operational information in their vicinity. As this system represents state-of-the-art mobile technology in the police domain, we want to know how it influences the work processes in the police domain and whether police officers accept this innovative technology. Consequently, our longitudinal field study in the police domain studies:

3. *What are the effects of a location-based notification system on work process efficiency and effectiveness and the user experience of mobile professionals?*

Table 1-3. Overview of the experiments in this thesis.

Subject (Chapter)	Operational demands and Human Factors	Adaptive system behavior	Evaluation setting	Outcome and process measures
MOBILE INFORMATION EXCHANGE				
Appearance of notification style (Chapter 5)	Message priority (task) Workload (context)	Notification styles (Information density & salience)	Simulated surveillance task in lab study	Task performance Notification intrusiveness Subjective judgments
Timing and appearance of notification style (Chapter 6)	Message priority (task) Activity (user)	Notification styles (Timing, information density & salience)	Mobile surveillance task	Task performance Notification intrusiveness Subjective judgments
TEAM COLLABORATION SUPPORT				
Team awareness support (Chapter 7)	Location (context) Availability (user) Expertise (user)	Presentation modality (auditory vs. visual)	Individual surveillance task in virtual environment	Decision accuracy Team awareness Subjective judgments
Team task allocation support (Chapter 8)	Location (context) Activity (user) Message priority (task)	Task allocation Notification styles	Team surveillance task in virtual environment	Team task performance Subjective judgments
Reliability of task allocation support (Chapter 9)	Location (context) Activity (user) Message priority (task)	Task allocation Notification styles	Team surveillance task in virtual environment with context events	Team task performance Team awareness Subjective judgments

Continuing from this first iteration, we expect that incorporating more context knowledge in the system results in better support. The identified context factors from the work domain analysis and the field study will guide appropriate adaptations in system behavior. In our first lab experiment, a prototype context-aware notification system adapts the visual and auditory notification appearance to message priority and the use context (e.g. a high or low workload environment). To assess the effects of notification style appearance on task performance, we employ this prototype in a simulated surveillance task to study:

4. *What are the effects of notification style appearance, adapted to priority and use context, on task performance, notification intrusiveness and the user experience in a simulated surveillance setting?*

Notification styles should not only focus on how (e.g. visual and auditory appearance), but also on *when* to present an incident message within an ongoing task. In addition, we want to know what the effects on task performance and subjective judgments are of adapting notification styles *dynamically* in an ongoing task. To test this, the second experiment will employ a real life mobile surveillance task using a similar notification prototype. In this case, the prototype adapts both the timing and appearance of notifications to message priority and user activity. This experiment will answer:

5. *What are the effects of a context-aware support system, which adapts notification timing and appearance to message priority and user activity, on task performance and the user experience?*

What we learned from context-aware notification for individuals can now be extended to the support of teams. To address the challenge of **team collaboration**, context-aware notification systems must support team awareness and task allocation. One important aspect of mobile team collaboration is asking colleagues for assistance in handling an incident. In order to choose the appropriate colleague effectively, mobile professionals must monitor ongoing radio communication. We first want to know whether providing distributed team members with a visual overview actually supports team awareness and decision making. Thus, the third experiment will employ a simulated police team task in a virtual environment, where a prototype with visual team information and communication is compared to current auditory support. This experiment will answer:

6. *What are the effects of the visual and auditory presentation modality for team information on team awareness and decision making accuracy?*

In team collaboration, efficient task allocation is essential. During police surveillance, appropriately allocating tasks to team members is challenging in this distributed work environment. Context-aware systems can aid this process by allocating tasks (i.e. incident messages) to team members, based on context factors (officer availability, proximity to the incident and incident priority). Based on decision rules, the redesigned prototype selects and notifies team members of current incidents, by providing advice on who can handle a specific incident best. The fourth experiment will evaluate this prototype system with police end-users in a surveillance task through a virtual environment. This experiment will answer:

7. *What are the effects of providing task allocation advice, adapted to availability, location and incident priority, on team task performance, communication and decision making during mobile team surveillance?*

Finally, designing context-aware systems for critical environments requires that these systems cope with the dynamic nature of these environments. In real life, automated systems will not be 100% reliable. Providing task allocation advice may aid team members when all incidents are "known" to the system, but spontaneous context events (encountering a panicked person or a roadblock) may result in outdated or inappropriate advice. In addition, relying on advice may diminish shared situation awareness (e.g. not knowing what your team members are currently doing). It is necessary to identify how the benefits of such automated support relate to the costs. The fifth experiment studies the effects of partially reliable task allocation support on task performance and subjective judgments. In addition, it studies how (shared) situation awareness and trust in the system help users to maintain task performance. This experiment will answer:

8. *What are the effects of partially reliable task allocation support on task performance and are these mediated by situation awareness and trust in the system?*

1.9 Thesis overview

This thesis is structured as follows (see Figure 1-2). Part I presents the Situated Cognitive Engineering design and evaluation method. Chapter 2 focuses on the work domain and support analysis as well as existing Human Factors literature. Chapter 3 describes the framework for evaluating context-aware support in mobile professional domains. How context-aware notification can support mobile information exchange is addressed in three studies in Part II of this thesis. Chapter 4 presents a field study of a location-based notification system in the police domain. This is followed by two empirical studies on notification appearance (Chapter 5) and adaptive notification styles in a mobile surveillance task (Chapter 6). Results from these studies form the basis for the design of team collaboration support, presented in Part III. First, Chapter 7 focuses on the presentation modality for team awareness support. Then, Chapter 8 focuses on team task allocation support in a study with police end-users and Chapter 9 studies how users deal with situations in which context-aware systems provide partially reliable advice. Finally, Chapter 10 in Part IV concludes this thesis with a discussion of the main findings and directions for future research.

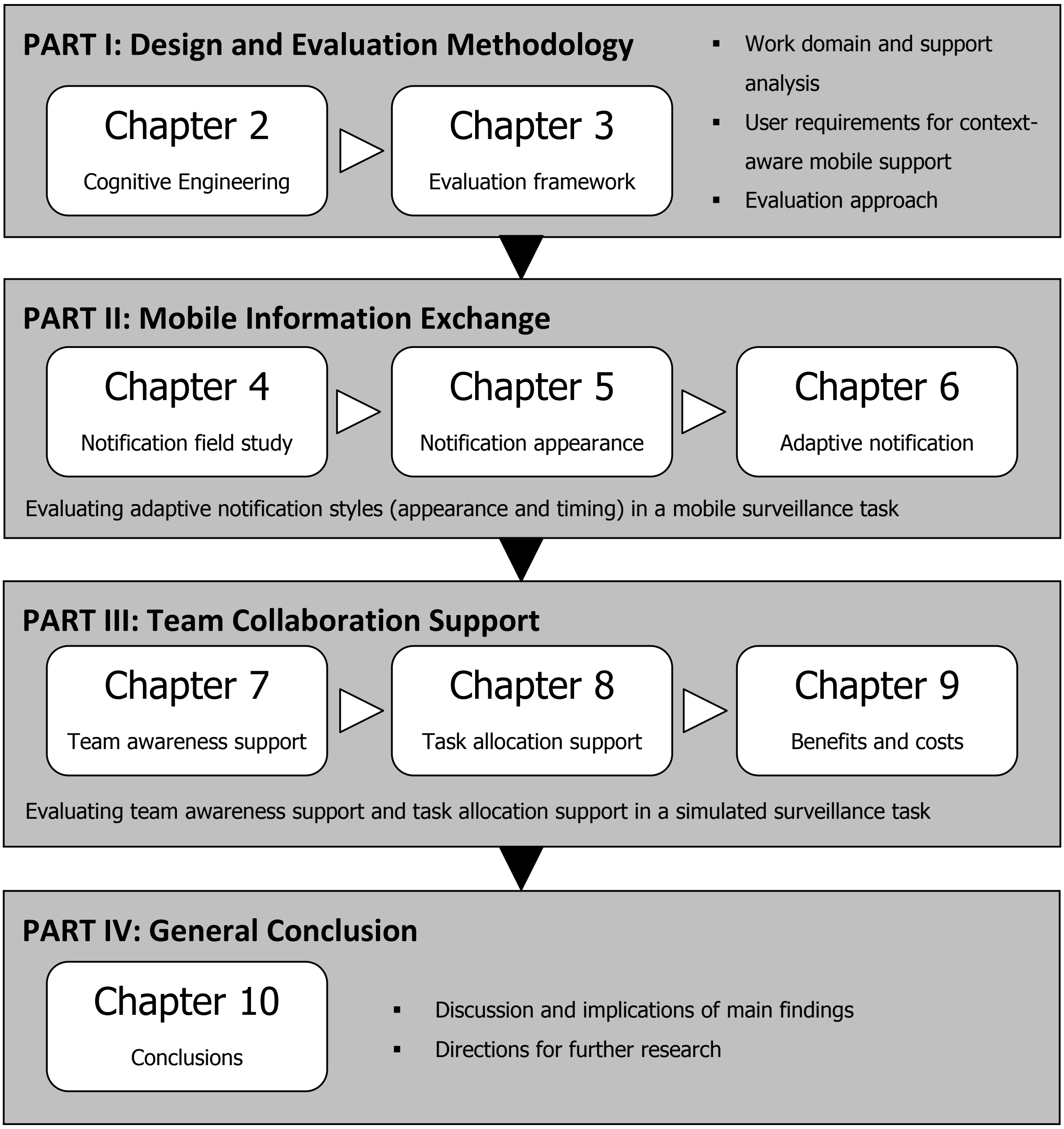
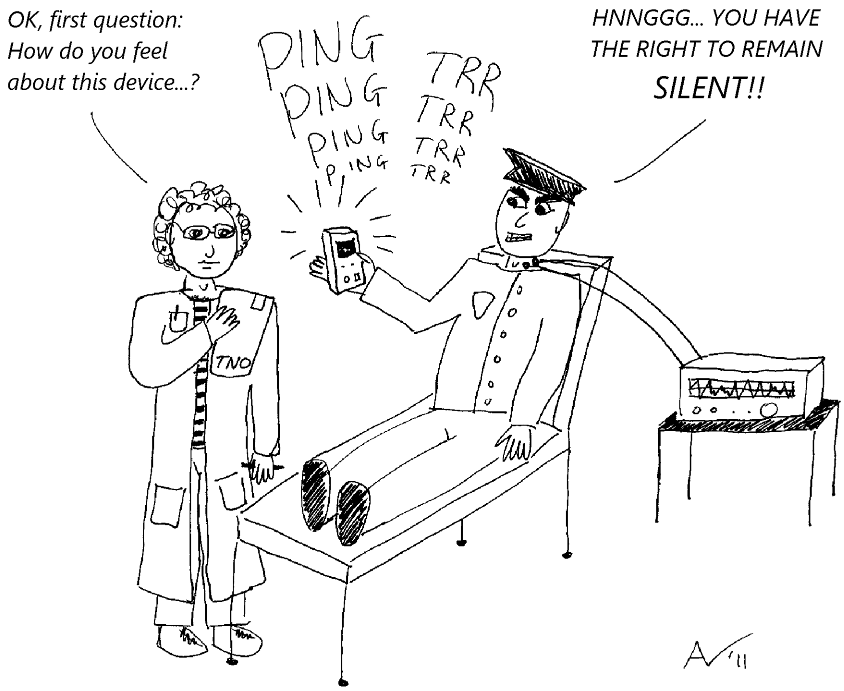


Figure 1-2. Overview of the thesis structure.

PART I

DESIGN AND EVALUATION METHODOLOGY



2 COGNITIVE ENGINEERING TO DESIGN CONTEXT-AWARE MOBILE SUPPORT

Abstract

Designing context-aware mobile support (CAMS) systems for mobile professionals should focus on 1) how processes in the work domain need to be supported and 2) how the use context and Human Factors influence mobile human-system interaction. Situated Cognitive Engineering (SCE) systematically addresses both issues to arrive at a grounded design rationale and user requirements specification. This chapter presents the SCE method, describes the work domain and support analysis and relevant literature. To support mobile police surveillance work, CAMS should provide information access, notification and prioritization of incidents and support team awareness and task allocation. CAMS should adapt the information presentation and interaction based on task, context and user factors (e.g. presenting a notification based on officers' proximity to an incident). The specification of the CAMS concept and core functions and two assessment iterations with police end-users are described.

This chapter is based on the previously published material:

Streefkerk, J.W., van Esch-Bussemakers, M., & Neerincx, M. (2006). Designing personal attentive user interfaces in the mobile public safety domain. *Computers in Human Behavior, 22,* 749-770.

Streefkerk, J.W., van Esch-Bussemakers, M., Neerincx, M., & Ait-Yaiz, R. (2006). *Attentive Services: Relevant factors for design* (Rep. No. TNO DV3 2006 A056). Soesterberg: TNO Defence, Security and Safety.

2.1 Introduction

As stated in Chapter 1, work in mobile professional domains suffers from limited mobile information exchange and inefficient team collaboration. Potentially, networked context-aware support (CAMS) systems can mitigate these problems by adapting information presentation and interaction for (teams of) individual police officers. These systems aim at a more fluid and relevant interaction by adapting their behavior to relevant task, context and user factors (Shafer et al., 2001). Previous mobile Human-Computer Interaction (HCI) research has focused on context-aware and location-based information presentation within a specific domain such as tourism, transportation, construction work, firefighting and mobile public government (Abowd et al., 1997a; Abowd, Dey, Orr, & Brotherton, 1997b; Camp, Hudson, Keldorph, Lewis, & Mynatt, 2000; Cheverst, Davies, Mitchell, Friday, & Efstratiou, 2000; Chen et al., 2000; Cheverst et al., 2002; Krug, Mountain, & Phann, 2003; Wahlster, 2003; Vermeulen, 2004; Jiang, Hong, Takayama, & Landay, 2004b; Aziz, Anumba, Ruikar, Carrillo, & Bouchlaghem, 2006). Location-based services, whereby the mobile device has access to geographical location information, provide tourist information relevant to the users' present location or help users to locate friends in their vicinity (Anhalt et al., 2001). As we will see in Chapter 0, this technology has only recently emerged in mobile professional domains. To our knowledge, CAMS systems have not yet been developed for these domains.

Following a Situated Cognitive Engineering method (described in Section 2.2), this chapter will elaborate the work domain and support analysis relevant for designing CAMS. This analysis aims to answer: *Which task, context and user factors in mobile professional domains are relevant to design context-aware support for mobile information exchange and team collaboration?* The primary application area is the domain of public safety and assistance; e.g. the police work environment. The work domain and support analysis encompasses operational demands (task processes), the mobile use context and Human Factors (user cognition and personalization). First, the analysis focuses on the three general processes in police surveillance (emergency aid, law enforcement and criminal investigation). This results in an overview of police surveillance needs for context-aware mobile support. Then, the analysis outlines how the mobile use context and Human Factors influence the design of a CAMS system and how these factors apply to each of the police processes. For example, during emergency aid police officers' attention should be guided to relevant information or objects (i.e. notification of incidents). Or when an officer has to handle two incidents at the same time, it is important that he is directed to the task with the highest priority. This work domain and support analysis results in specification of a generic concept for CAMS, describing high-level core functions.

2.2 Situated Cognitive Engineering

There are a variety of different approaches to designing for mobile HCI. To design context-aware mobile support for professionals, Situated Cognitive Engineering (SCE) will be employed, which incorporates scenario-based design (Carroll, 2000; Neerincx et al., 2008). Analogous to “classical” CE methods, design solutions are based on knowledge from cognitive psychology and Human Factors and evaluated in an iterative fashion (Hollnagel & Woods, 1983; Norman, 1986). In addition, the SCE method employs work domain and mobile technology analyses, to provide a sound design rationale for the intended work domain. This iterative, empirical design and evaluation approach incrementally refines the requirements baseline and design solutions, based on knowledge about user needs.

The SCE method consists of an analysis, specification and assessment part (see Figure 2-1). Contrasting to classic “waterfall” methods for software engineering, these parts are not strictly separated and may be addressed in parallel. Work domain and support analysis focus on the operational demands within a specific work domain, knowledge about envisioned (mobile) technology and Human Factors research. This analysis forms the input for the design rationale, consisting of the specification of a concept, scenarios, core functions and claims. The concept is a broad description of the proposed system. Scenarios are drafted from the relevant application domain, describing users, their tasks and context in a comprehensive, narrative style. Scenarios are not meant to give a complete and exhaustive picture; instead they focus on the intended use, associated problems and illustrate design solutions. From the scenarios, core functions and claims are specified, describing the expected operational effects of the support system. The claims form the basis for testable hypotheses, which are assessed in the evaluation cycle. Combining the concept, scenarios, core functions and claims, the process of user interface design results in functional specification of system features.

Assessment of the claims and features is done by 1) having HCI experts, end-users or software engineers review and comment on them (lower left cycle in Figure 2-1) and 2) by implementing these features in (semi-functional) prototypes and evaluating them on objective and subjective HCI and user experience metrics (lower right cycle in Figure 2-1). To

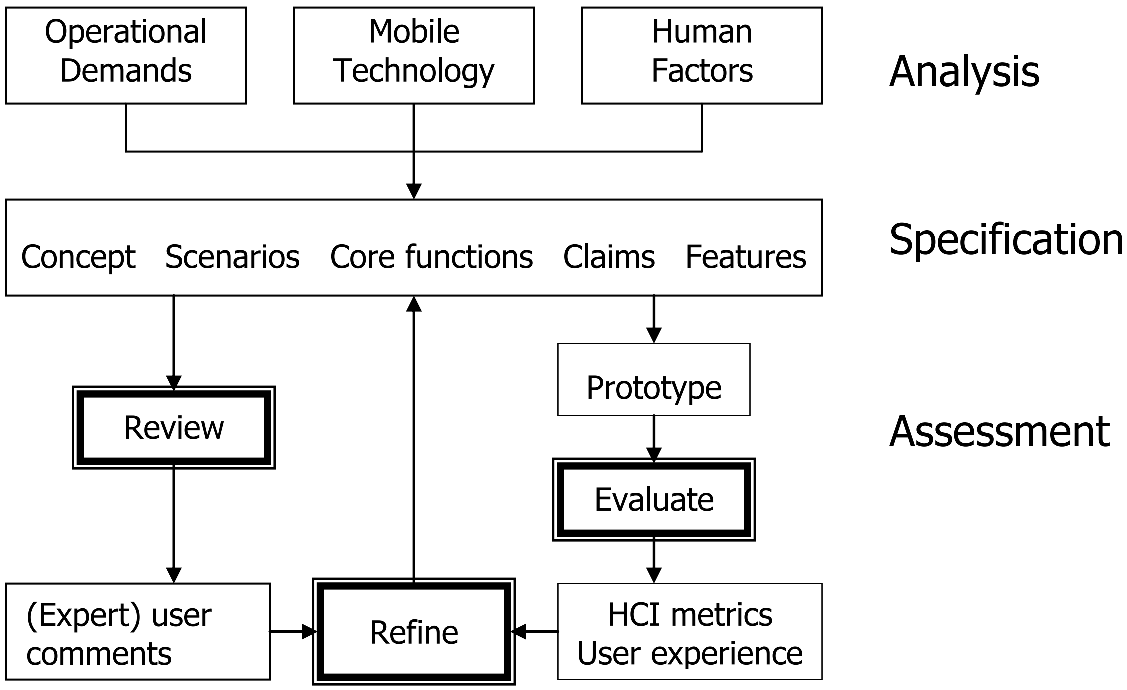


Figure 2-1. Situated Cognitive Engineering (adapted from Neerincx et al., 2008).

select appropriate evaluation methods and techniques, we apply a framework that distinguishes seven evaluation constraints (described in detail in Chapter 3). Both the *stage* in the development process and the *purpose* of the evaluation set specific requirements for the available techniques. In addition, context-aware mobile systems are by nature *complex*. The user interface adapts to the changing context and this emerging, adaptive behavior should be assessed appropriately. This complexity is increased by designing systems for the professional domain. The characteristics of this domain require a different approach than evaluating entertainment systems such as a MP3 player. Thus, the *setting* of the evaluation is important. Access to representative end-users and domain-specific settings in the professional domain can be limited, begging the question *who* to include in the evaluation and for *how long*. Finally, different methods for evaluating mobile devices involve different *costs* in both time and resources.

During evaluation, human participants (end-users or representative users) work with prototypes in a specific task setting. Both established HCI metrics as well as new HCI metrics specifically adapted for context-aware mobile devices are employed. Effectiveness, efficiency and satisfaction are established criteria for evaluating HCI (ISO 9241-11, 1998) but should be matched and tuned to domain-specific performance criteria. In case of the police domain, efficiency and effectiveness of mobile information exchange can be operationalized in measuring decision errors, response time and incident handling time. How well team members collaborate depends on their (shared) situation awareness (SA) (Endsley & Garland, 2000) as well as team coordination (Gorman, Cooke, Pederson, Connor, & De Joode, 2005). User experience metrics include subjective judgments on trust, reliance and user acceptance of the system (Marsh & Meech, 2000; Parasuraman & Miller, 2004; Wang & Emurian, 2005). These metrics are influenced by usability criteria for adaptive systems such as predictability, controllability and transparency of the system (Cheverst et al., 2002; Nagata, 2003; Alty, 2003; Paymans & Lindenberg, 2004).

It is important to note that the SCE method is an iterative process, with a full cycle including the assessment of system features on HCI metrics, and further *refinement* of the scenarios, core functions and claims based on this assessment. It is not necessary to evaluate the whole system at once; specific features or parts of the system can be evaluated separately. The end products of this cycle are practical Human Factors theories, models and methods for an adaptive support system, validated within an application domain.

The remainder of this chapter presents the analysis and specification cycles of the SCE method, applied to designing context-aware mobile support for the police domain. This work domain and support analysis focuses on the operational demands in mobile police surveillance (section 2.3), the mobile use context (section 2.4), and Human Factors (sections 2.5 and 2.6). The results are used to specify the CAMS system in the police domain. Both the specification and a first assessment with police end-users is described in section 2.7.

2.3 Operational demands: Mobile information exchange and team collaboration

The work domain and support analysis identified the operational demands for the three different mobile police surveillance processes. This analysis drew on user studies and a police requirements analysis found in the literature (Baber et al., 2001; Nulden, 2003; Stijnman, 2004; Kun et al., 2004; Marcus et al., 2006; Feteke et al., 2008). In addition, participatory observation in two districts, interviews with police officers and an internal police analysis (Mente, 2004) provided external validity to the literature research. Cognitive task analysis could complement the work domain analysis, for example with detailed knowledge on how police tasks are structured (Ioimo et al., 2004; Neerincx, 2004). For the purpose of this thesis however, a high-level understanding of police surveillance processes suffices.

2.3.1 Mobile police surveillance processes

Police officers' activities include three general processes: emergency aid, criminal investigation and law enforcement. In *emergency aid* police officers respond to incidents or disasters, such as fires, floods or car crashes. This process is characterized by reactivity, high time-pressure and high urgency. Emergencies often require multiple officers to handle, requiring team collaboration. Once officers are informed of the incident, they proceed to the incident scene as soon as possible. Here they execute certain protocols to contain and handle the incident. Contrasting, the process of *criminal investigation* is highly proactive. Before going out on the street, police officers receive a briefing on the focal points for that shift. This information consists of descriptions of suspects, vehicles and relevant developments since the last briefing. Based on the focal points for that shift, police officers pay specific attention to events and objects in the street. Then, during debriefing, officers report acquired information and relevant developments. In *law enforcement*, officers are responsible for maintaining law and order in public spaces. Examples are police surveillance through shopping streets and at events such as football matches. Officers have to be alert of criminal acts and have jurisdiction to intervene and apprehend suspects. Often, surveillance is performed by teams of officers, working distributed over different geographical locations. They have to coordinate actions together, for example to handle a high priority incident as fast as possible with multiple team members.

This analysis focuses on supporting both the individual police officer as well as teams of police officers in executing these three processes on the street. Teams of police officers perform mobile surveillance while paying attention to criminal activity (law enforcement), responding to incidents (emergency aid) and focusing on briefing focal points (criminal investigation). In this mobile surveillance, the operational demands of mobile information exchange (information access, notification and prioritization) and team collaboration (team awareness and task allocation) must be supported (Baber et al., 2001; Ioimo et al., 2004). The operational demands for each process are summarized in Table 2-1 and discussed in the next sections.

Table 2-1. Overview of police surveillance processes and operational demands.

Operational demands	**Emergency aid** *Responding to occurring incidents*	**Criminal investigation** *Proactively focusing on briefing focal points*	**Law enforcement** *Maintaining law and order in public spaces*
Mobile information exchange			
Information access	Mobile access to information on current incidents (location, priority, details)	Mobile access to briefing focal points, assignments and information on historical incidents.	Mobile access to information on suspects, vehicles and delinquents.
Notification	Notification of current incidents.	Notification of previous incidents (e.g. open warrants) and briefing focal points.	Notification of events in public spaces (e.g. aggressive behavior).
Prioritization	Prioritization of current incidents.	Prioritization of briefing focal points based on quality criteria.	Prioritization of procedures and related information.
Team collaboration support			
Team awareness	Keeping track of location and activity of team members as well as location, status and details of incidents.		
Task allocation	Optimal division of work by allocating incidents to distributed team members.		

2.3.2 Mobile information exchange

As was stated in Chapter 1, timely access to accurate, dynamic information is very important in the processes of emergency aid, criminal investigation and law enforcement. The mobile officer needs task information such as his current assignment, resource information such as database access and immediate information about incidents such as type and location (Baber et al., 2001). Currently, this kind of information is only accessible upon request via the emergency room using a radio transceiver or during briefing prior to work. Recently, the police organizations improved information access for mobile officers (Ioimo et al., 2004; Stijnman, 2004; Van Loon, 2004). Police officers could query databases and access digital briefing information using a Personal Digital Assistant (PDA). User studies found that officers felt supported in their task execution but were dissatisfied with the implementation and interaction style. They had to create an opportune usage moment so they could devote their attention to interacting with the device. As a result, they could not always maintain eye-contact with their environment.

Notification helps to guide police officers' attention to important events, information or objects in their environment. Currently, the emergency room plays an important role in notifying officers of incidents via their radio transceivers. When an incident occurs, the emergency room broadcasts the incident call, without knowledge of the location or current activity of individual officers. For officers who are busy handling an incident, not all communication from the emergency room is directly relevant and might cause unwanted interruptions. During participatory observation, busy police officers were observed to turn off their radios e.g. when talking to suspects or victims. Still, they need to be aware of high priority emergency situations (e.g. when colleagues in danger are requesting assistance).

Notification needs are determined by what the officer is doing (activity) and his mode of transport. Especially for emergency aid, early warning and notification on incidents is essential, because of the urgent nature of these incidents. Police officers in cars can reach an incident location faster than on foot, consequently they need more timely notification. In criminal investigation, notification can aid the police officer to focus attention to criminal "hotspots", certain problem areas or locations where criminal behavior is encountered often. For law enforcement, mobile police officers need to be notified of dangerous situations that are developing in public spaces.

Prioritization helps officers to ensure that they are working on the incident with the highest priority and that they are not missing any information (Baber et al., 2001). Currently, categorizing incidents as low, normal or high priority is already in use in emergency aid. Low priority incidents or tasks are without time-pressure or risk of personal injury. Contrasting, high priority incidents are time-critical situations involving either personal injury or a high chance of apprehending a suspect and require multiple police officers. In criminal investigation, prioritization of briefing focal points can help police forces to make choices in which type of criminality is addressed (Mente, 2004). In law enforcement, police work is dependent on protocols and procedures that need to be followed. Here, a mobile support system could (pro-actively) present prioritized steps in procedures on a PDA.

2.3.3 Team collaboration support

Police officers on surveillance operate as geographically distributed, ad-hoc teams. For all three police processes, they have to keep track of the activities and location of colleagues to work effectively as a team. Staying aware of team members' status and activities is known as team awareness. This ability improves team collaboration, for example in asking an available colleague to assist in handling an incident. Currently, mobile police communication systems (e.g. radio transceivers) transmit all communication between team members via the auditory modality. The problem is that it is cognitively demanding to keep track of activity and status of other team members from auditory communication alone.

Efficient team collaboration depends on appropriately allocating incidents to team members. Especially in the process of emergency aid, incidents usually require two or more police officers to handle. The emergency room does not know which officers are currently available, but relies on officers to indicate their availability themselves via radio. Police officers communicate with colleagues to determine who should handle which incident. They currently have no overview of availability and location of team members to support this task allocation process. This makes it sometimes unclear who is available to handle an incident, potentially resulting in decision errors and ineffective communication.

2.3.4 Conclusion

Concluding from the work domain analysis so far, mobile police officers need support in all three processes of emergency aid, criminal investigation and law enforcement. Providing mobile information access, notification of incidents and prioritization of tasks will keep officers informed of occurring incidents and help improve incident handling. Team awareness and task allocation support will ensure an optimal division of work across distributed team members. To support these demands, a context-aware mobile support system must have knowledge of police officers' current activity, incident priority and mode of transport.

2.4 Mobile technology: Mobile use context and technological constraints

How context-aware mobile technology can be implemented for mobile police officers, depends on the mobile use context and technological constraints. This section specifies how context-aware mobile support systems can use knowledge about the mobile use context (location, time and social interaction) to improve information presentation. In addition, these systems have technological constraints in network architectures and infrastructures that influence their implementation.

2.4.1 Mobile use context

Mobile professionals do not perform their tasks in isolation, but in rich interaction with people and the environment around them (i.e. the mobile use context). Two of the most widely used context factors in context-aware systems are location and time (Abowd et al., 1997b; Dey et al., 2000; Chen et al., 2000; Cheverst et al., 2002). Location can be used for location-based notification (as illustrated in the introduction), while time can be used for automatic reminders at certain points in time. For our CAMS system, location information can aid information access and notification for police officers. For example, by automatically notifying police officers of incidents or important locations (e.g. criminal “hotspots”) in their direct vicinity (Anhalt et al., 2001). Also, prioritization of incidents can be done based on location information, for example to advise which team member can reach the incident location first. Team collaboration can also be supported with location information. For example, when a police officer searches for the nearest colleague, the system provides him with sorted results based on proximity to his own location (Baber et al., 2001; Anhalt et al., 2001). Time information can be used by the system for appropriate timing of notifications or reminders for events (e.g. a scheduled demonstration). As location and time change more quickly depending on the mode of transport, an officer in a car will need more timely notification than his colleague on bike. In addition, recording when actions took place can facilitate subsequent reporting and information access.

All three police processes involve social interaction with colleagues, suspects and victims. Police officers indicate that social interaction is central to their work, especially in criminal investigation and law enforcement. The interaction with a support system should not hamper maintaining eye contact with their environment (Mente, 2004; Stijnman, 2004). Furthermore, a context-aware system can postpone incident calls when officers are interacting, or on the other hand, interrupt them when this is required. However, as social interaction often occurs spontaneous, context-aware support systems may not always be aware of this interaction occurring. Empirical research (specifically addressed in Chapter 9) will need to establish the benefits and costs of context-aware support vis-à-vis the unexpected nature of police work.

Of course, more context factors could be relevant for designing for the police domain, such as environmental noise, lighting conditions, etc. For example, if the system is aware the user is in a noisy area with insufficient lighting, appropriate adaptations are to increase display luminance or provide tactile notification instead of auditory notification. However, we believe the context factors of location, time and social interaction to be sufficient to support mobile information exchange and team collaboration.

2.4.2 Technological constraints

Realizing CAMS systems requires taking into account the constraints of mobile computing software and hardware platforms, system architectures and network infrastructures. To enable adaptive information presentation, software has to ensure the separation of content and layout (for example using XML; Kerer & Kirda, 2001). This separation is a key element to present information in different formats and modalities on different devices. Current mobile hardware platforms (such as smartphones, PDA's and handheld computers) have different display sizes and input-output options, influencing their appropriateness for implementation of context-aware mobile support (Cremers, Engels, & Vlaskamp, 2003; Cremers & van der Pijl, 2005; Streefkerk, van Esch-Bussemakers, Neerincx, & Ait Yaiz, 2006). Smartphones are interesting options because of their integrated network connectivity and computing power. However, they often have smaller display sizes and less interaction possibilities (e.g. buttons instead of pen-based or touch-screen input) than PDA's or handheld computers (York & Pendharkar, 2004). The physical design of a support system should allow police officers to maintain eye contact with other people and the environment and keep their hands free as much as possible. Wearable computing devices, integrated around the body or into clothes might be appropriate options, as they contain displays positioned on the user's forearm or mounted on glasses in the user's visual field (Abowd et al., 1997b; Rypkema, van der Lee, van der Meiler, & Weitenberg, 2005).

To design a system architecture for context-aware mobile devices, a common approach is to distinguish between a sensing, modeling and output layer (Chen et al., 2000; Dey et al., 2001; Hong et al., 2001; Hofer, Schwinger, Pichler, Leonhartsberger, & Altmann, 2003; Bardram & Hansen, 2004; Sorensen et al., 2004). Data from sensor technology (e.g. GPS location trackers) provide input for the modeling layer. In this layer, the data is fed into rules or reasoning algorithms and cut-off values determine adaptive behavior. The computations in this layer can be performed on a remote server or locally on the device. Based on the outcome, the output layer takes care of the information presentation in the appropriate format on the device. These devices require a robust network infrastructure to connect to other devices and remote servers or databases (via Wireless Local Area Network or Universal Mobile Telecommunications System). Based on device type and proximity of other devices, “ad-hoc networks” enable automatic connections and communication links between distributed mobile workers. For aid workers in crisis situations, this resulted in a very robust communication network that efficiently used available computing power (Abowd et al., 2002; Jiang et al., 2004a; Jiang et al., 2004b).

Architectures and network infrastructures for CAMS face a number of constraints. First, network connectivity is limited in bandwidth and dependent on the environment. Location is often sensed outdoors with Global Positioning System (GPS) technology. The latency of this signal determines how fast a context-aware system can respond. When connection with sensors cannot be established, errors may occur in the adaptive system behavior (Paymans et al., 2004; Van der Kleij, de Jong, te Brake, & de Greef, 2009). Second, the interpretation of sensor data inherently has uncertainty in precision and accuracy (Hong et al., 2001). It may be that the wrong piece of context is sensed, or that the context is interpreted inaccurately. Third, users must have insight in the adaptive behavior of the system, to understand why certain adaptations are made (Paymans et al., 2004). This requires a transparent or modular approach to designing the architecture for adaptive systems (Höök, 2000; Paramythis et al., 2001; Weibelzahl, 2005). Fourth, privacy, network security and reliability must be ensured as much as possible, especially in mobile professional domains.

2.4.3 Conclusion

Focusing on context-aware mobile technology, this part of the analysis showed that CAMS for mobile professional domains needs to take into account the mobile use context factors of location, time and momentary social interaction. Furthermore, designing CAMS needs to take into account constraints in mobile computing technology, specifically the separation of content and layout, the characteristics of mobile or wearable computing devices and the robustness of system architectures and network infrastructures.

2.5 Human Factors: Attention and workload

As stated in the introduction, mobile human-computer interaction is influenced by Human Factors such as momentary attention and workload. Specifically, efficient interaction with CAMS systems in the mobile domain is constrained by (limitations in) human information processing and cognitive resources. This section discusses how cognitive abilities such as attention, task switching, workload and situation awareness are relevant for designing context-aware support.

2.5.1 Divided attention and distraction

Attention can be regarded as a selection process whereby conscious access to working memory is made possible through direction of attentional focus (Wickens, 1987). Attention is inherently limited in focus and can be either voluntarily directed or involuntary drawn by objects, sounds or movement in the environment. At the perceptual level, (changes in) object features such as salience, color, motion and sudden appearance easily grab attention (Bartram, Ware, & Calvert, 2003). At the cognitive level, attention is directed to those elements that are consistent with the user's task, expectations or goals (Wood, Cox, & Cheng, 2004). These characteristics of attention entail two problems in directing attention. First, when people have to attend to two or more tasks simultaneously (dual task situations), divided attention results in performance decrements (Horvitz et al., 2003). Second, grabbing attention at the wrong moment (e.g. receiving a telephone call while navigating to an address) can lead to inappropriate focus of attention through distraction (Klapp, 1987). These attention problems occur especially in mobile computing environments: interacting with a device while navigating through the environment can be regarded as a dual task situation, while distraction occurs through spontaneous context events (e.g. encountering a panicked person).

To accurately support attentional processes, CAMS needs to ensure the users' attentional focus is on the right task and place (e.g. on the device or on the environment). Attention management systems should gather knowledge about users' task and context to decide when and how to draw attention (McCrickard et al., 2003a; Adamczyk et al., 2004; Fogarty et al., 2005a; Bailey et al., 2006). Based on this knowledge, the system can dynamically adapt the output modality, information presentation and interaction style to attract and maintain attention. Examples in desktop settings include "Attentive User Interfaces" that prioritize information presentation based on reasoning about information processing resources (i.e. attention or workload) (Jameson, Schäfer, Weis, Berthold, & Weyrath, 1999; Vertegaal, 2002; Horvitz et al., 2003; Zhai, 2003). As they rely on estimating users attentional focus from eye-tracking data, this is impractical to implement in mobile settings (Jameson, 2002; Abowd et al., 2002; Horvitz et al., 2003). Ideally, information from multiple "sensors" (user actions, tasks, location, etc.) should be used to make inferences about attention.

The goal of context-aware notification systems is improving the human-system interaction by adapting to user attention. Previous research has focused on how notifications are presented (e.g. presentation modality, salience, information content). In mobile situations, visual attention is often away from the display, making it more effective to attract user attention via auditory or tactile modalities, for example using sounds or vibrations. Context-aware notification systems can adapt the presentation modality of notifications (e.g. visual, auditory and tactile signals), for example to limit notification intrusiveness. Kern and Schiele (2003) adapted the salience (e.g. beeping or ringing) and amount of information in 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 increasingly salient auditory signals and more elaborate information content as message importance increased. While tactile cues are used to limit distraction, especially to relieve visual attention (Hopp, Smith, Clegg, & Heggestad, 2005), these require the device to be in close contact with the body.

For mobile police officers, context-aware notification can play an important role in police surveillance. Information access may be helped by presenting information auditorily, so officers can keep their attention on the environment. Notification and prioritization can be supported by adapting notification presentation to incident priority and location, by providing officers with salient notifications for high priority incidents in their vicinity. Or notifications for low priority incidents may be presented in a condensed form (e.g. only an outline of the message). While attentional processes may be supported by *how* a notification is presented, *when* notifications are presented is important for task switching and interruption.

2.5.2 Task switching and interruption

In a dynamic mobile environment, multi-tasking situations occur frequently, where a user has to perform multiple tasks simultaneously (e.g. Nagata, 2003). Switching between these tasks influences attention in two ways. First, the interruption of a task causes attention to move away from that task to the source of the interruption. Depending on user goals, this is either an unwanted distraction from the primary task or an attraction to valued and necessary information (McCrickard, Chewar, Somervell, & Ndiwalana, 2003). Unwanted interruption has negative consequences for task performance, causing disorientation on part of the user (Nagata, 2003). Longer task completion time, higher error rates, and increased frustration and anxiety have been demonstrated (Cutrell, Czerwinski, & Horvitz, 2000; Nagata, 2003; Adamczyk et al., 2004; Bailey et al., 2006). Second, research indicates that the process itself of switching between tasks places high demand on attention, thereby increasing cognitive workload. In addition, a switch between tasks is more disruptive than sequential finishing of tasks and thereby negatively affects performance (task switching costs) (Neerincx, van Doorne, & Ruijsendaal, 2000). How high these costs are depends on the relatedness of the tasks, the duration of the work on the tasks and the cognitive workload experienced by the person working on the tasks (Wickens, 1987). Consider for example, a police officer engaged in checking a license plate on a car. Suddenly, he is notified of a theft of another car and receives a description of the stolen vehicle. He now has to distinguish between the two sets of car information, which is more cognitively demanding than when the two sets were less alike.

The goal of CAMS is to support switching between tasks by avoiding unwanted interruption of the ongoing task, but to interrupt when necessary. To make this decision in a meaningful way, notification systems must have knowledge about the user’s task and context (such as activity, priority or location) to subsequently time interruptions appropriately (McFarlane, 2002; Horvitz et al., 2003; McCrickard et al., 2003b; Gievska & Sibert, 2004; Iqbal & Bailey, 2008; Bailey et al., 2008). A context-aware notification system can predict appropriate moments for interruption, such as at activity breakpoints or changes in location and task priority (Adamczyk et al., 2004; Kostov, Tajima, Naito, & Ozawa, 2006). The negative effects of unwanted interruption (frustration, increased mental effort) are mitigated by timing interruptions at these points in task execution (Gievska et al., 2004; Fogarty et al., 2005b; Bailey et al., 2006). In addition, interruptions showed less negative impact when users expected an interruption to occur. Thus, creating anticipation of when an interruption will occur facilitates task switching (McFarlane, 2002). Overcoming task interruptions in mobile computing can be done by deploying certain interface elements, such as the Point of Return Indicator, which directs users’ attention back to a specific point of a suspended task. A user study showed that these aids supported users effectively in continuing their work after a task switch (Nagata, 2003).

It should be clear that task switching and interruption influence attention and task performance. For the mobile police officer, notification and prioritization are supported by timing notifications at appropriate moments and locations. This supports either staying with the current task or a valuable switch to a new task. Examples are postponing a low priority message while engaged in a high priority incident, or interrupting a low priority incident with a message about a new high priority incident. This requires a context-aware notification system to be aware of task information (officer activity, incident priority) as well as the context (location).

2.5.3 Situation awareness

Situation awareness (SA) is defined as the continuous perception (Level 1) and comprehension (Level 2) of the state of the world, and projecting (Level 3) this state into the near future (Endsley, 1988; Endsley et al., 2000; Wickens, 2008). For example, police officers on surveillance have to be aware of their environment to recognize and understand a high-risk situation and anticipate actions to minimize the risk. When underway to an incident, police officers strive to obtain a complete overview of the incident situation beforehand and attempt to integrate all the relevant incident, location and resource information. In collaborating with colleagues, *shared* situation awareness (Salas, Prince, Baker, & Shrestha, 1995) or team awareness (Ferscha, 2000) is necessary for efficient cooperation and communication. Shared situation awareness can be regarded as an understanding of the activities of others in dynamic work environments (Ferscha, 2000). Shared situation awareness is most often defined and measured as the knowledge held in common by team members. While knowledge is an important prerequisite for SA, as a measure it is dependent on working memory capacity (Salas, Cooke, & Rosen, 2008). How teams coordinate their perceptions and actions gives a valid measure of their situation awareness (Gorman et al., 2005). The problem is that it is cognitively demanding to maintain shared situation awareness in such a dynamic and distributed environment (Gutwin et al., 2002). For this reason, police officers indicate they like to have situation awareness support (Baber et al., 2001).

Our CAMS system is focused on supporting team awareness and team task allocation. A number of team awareness prototypes, some of them on mobile devices, have been described in the literature, e.g. Portholes (Dourish & Bly, 1993), Teamspace (Ferscha, 2000), Awarenex (Begole & Tang, 2007). Most prototypes gave information on people's location and availability, availability of resources and notified users of changes. Team awareness systems in emergency management support decision making by providing strategic information to distributed team members. This kind of information (such as availability and location) is considered most critical among team members to perform coordinated action (Convertino, Ganoe, Schafer, Yost, & Carroll, 2005). Other awareness support systems explicitly visualize team information such as the locations of colleagues and resources in geographical overviews (Rogers, Brignull, & Scaife, 2002; Cai, 2005; Carroll et al., 2006) or using mobile awareness cues (Oulasvirta, Petit, Raento, & Tiitta, 2007).

Staying aware of team members' tasks and activities is a necessary requirement for team task allocation. For example, a graphical overview of current tasks and activities can help deciding who should handle which incident or who to approach for back-up (Nulden, 2003; Kaber & Endsley, 2004). Such designs for activity awareness in mobile computer-supported cooperative work (e.g. Carroll et al., 2006) lead to increased team performance and awareness as well as reductions in mental workload (Kaber et al., 2004). Alternatively, a task allocation support system can directly assign tasks to team members based on expertise or momentary workload. This helps to avoid overload situations, where multiple incidents are allocated to one team member while other team members are available. This was shown in two studies of a task allocation system for incident handling in the Navy domain. This system ensured that tasks were performed by the right person at the right time (Neerincx, Grootjen, & Veltman, 2004) or by an automated system (Arciszewski, de Greef, & van Delft, 2009). On the downside, some authors argue that such high levels of system automation are not advisable in time-critical environments (Cummings, 2004). Still, few empirical users studies address automated task allocation in mobile, operational domains (Weibelzahl, 2005; Arciszewski et al., 2009). By supporting task allocation, it is expected that CAMS can distribute incidents optimally over available team members.

2.5.4 Conclusion

In this section, we have outlined the Human Factors of attention and workload and how these influence designing context-aware mobile support for police surveillance. In this domain, divided attention and distraction can be mitigated by adapting information presentation (modality, salience, content). CAMS systems can support task switching and appropriate interruption by appropriate timing of these notifications. Context-aware mobile support should help users maintain situation awareness, for example by providing overviews of team information, tasks and current status. In addition, team collaboration should be supported by allocating tasks to the appropriate team members. However, context-aware system adaptations need to be tailored to individual users. For example, presenting condensed information will suffice for one user, but may not be enough for another. We will address personalization and individual differences in the next section.

2.6 Individual characteristics: Personalization

Up to now, we have discussed police officers as prototypical users with no individual characteristics, experience or preferences. However, individual differences in attention, working memory and spatial ability do exist (Vicente, Hayes, & Willeges, 1987; Derryberry & Reed, 2001) and are relevant for task performance. Younger people generally have better spatial ability than elderly people, making it easier for them to navigate and orient themselves. Individuals with higher working memory capacity exhibit greater attentional control (Conway & Kane, 2001). These findings show individual users have different needs for support. Personalization is aimed at providing this support.

2.6.1 User modeling

In the introduction, we presented personalization as the adaptation of system behavior to individual users, based on accurate user models (Jameson, 2001; Fischer, 2001). For our CAMS system, modeling relevant user information includes user role or expertise, preferences and history. Police officers have different roles or expertise, that influence which incidents they want to be notified of. Roles or expertise can also be used to find experts in your vicinity. A user model containing expertise information can make this match effectively and thus support team collaboration. For example, a mobile service on a PDA enabling police officers to find contact details of district officers based on the location and subject of the incident (Eijk et al., 2006). User preferences can include notification preferences, explicitly stated in the user model. The system subsequently adapts the information presentation or output modality to these preferences, such as delivering a low priority call as an email message instead of a telephone call.

User history can be inferred by monitoring the user interacting with the support system. Police officer history must include which incidents were handled and what actions were taken. In supporting team awareness and collaboration, it is very important that an officer knows when a colleague has visited a location previously, what he encountered and which actions he subsequently performed. This information can be stored in a user model and used for subsequent updates or notifications to other colleagues. For example, by providing notifications with addresses of repeated delinquents.

The success of personalization relies in part on accurate predictions about user goals and tasks. This has proven difficult yet feasible, due to recent developments in statistical modeling techniques (Horvitz et al., 2003; Isbell, Omojokun, & Pierce, 2004). For example, goal-oriented adaptation uses formal models to predict user behavior based on reasoning about users' goals. In addition, generalizations based on earlier user behavior (when in situation A, users' next action will be action X) can serve as an 'educated guess' (Weld et al., 2004). Because the user model itself is adapted, personalization is a dynamic process and results in different support for different individuals. This points to the main challenge for personalization: supporting team collaboration. Here, a balance must be struck between shared situation awareness and personalization based on user models of individual team members. For example, can two police officers collaborate effectively, when one officer receives only concise information on an incident, and another receives more elaborate information (as dictated by preferences in their user model)? This question has to be addressed by empirical research.

2.6.2 Conclusion

Concluding from the work domain and support analysis, support opportunities have been identified in the three processes of mobile police surveillance. There is a clear need for a system that provides information access, notification and prioritization, as well as team awareness and task allocation. Table 2-2 summarizes relevant factors to guide the context-awareness of CAMS, including task (activity, priority, mode of transport), context (location, time, social interaction) and user factors (expertise, preferences and history). Based on this knowledge, CAMS can adapt the modality or salience of a notification, or provide task allocation advice on which team member can best handle which incident. Personalization can further optimize these processes by attuning the interaction to user expertise or preferences. This is expected to support the operational demands of mobile information exchange and team collaboration.

2.7 Specification of context-aware mobile support

In the remainder of this chapter, a first specification and assessment cycle is performed for context-aware mobile support for police officers. This section presents the design rationale, based on the work domain and support analysis. It specifies the concept, scenario, and core functions for CAMS. Following the Situated Cognitive Engineering method in Figure 2-1, two assessment iterations are performed, by reviewing the concept, scenario and core functions in two focus groups with police end-users.

Table 2-2. Relevant factors for context-aware mobile support.

Operational demands	
Activity	Police officer activity determines whether officers are available for a new incident, or should not be disturbed in their current activity.
Priority	Information on incident priority can help prioritization of incident messages. This ensures the officer is working on the incident with the highest priority.
Mode of transport	The mode of transport influences when the system needs to respond: officers in a car need more timely notification than those on foot.
Mobile use context	
Location, time, social interaction	The geographical location of police officers, the time of day, relative time between incidents and momentary social interaction should be incorporated in a context model.
Attention and workload	
Distraction	Preventing distraction and maintaining attention should proceed via adaptations in notification presentation, modality and interaction style.
Interruption	Task switching should be supported by appropriate timing of interruptions, prioritization and interface elements.
Workload	To avoid overload situations, incidents should be prioritized and allocated to available team members.
(Shared) situation awareness	Situation awareness should be supported by a geographical overview of team information and incidents.
Personalization	
Preferences, role, expertise, history	A user model should contain the role, expertise, preferences, and user history of individual officers. This model guides subsequent personalization such as adaptations in information presentation, modality and interaction style.

2.7.1 Concept specification and assessment

The definition of context-aware mobile support for police officers (introduced in Chapter 1) will now be further specified based on the work domain and support analysis in this chapter. We envision an adaptive support system for mobile information exchange and team collaboration, which has specific knowledge about the task, the use context and the user. The following high-level system concept is defined (see Figure 2-2). Police officers operate within a network of different information sources, such as the emergency room and (multimedia) databases. Information on criminal activities and their location must be presented to a team of police officers, each having their own task, function, expertise and

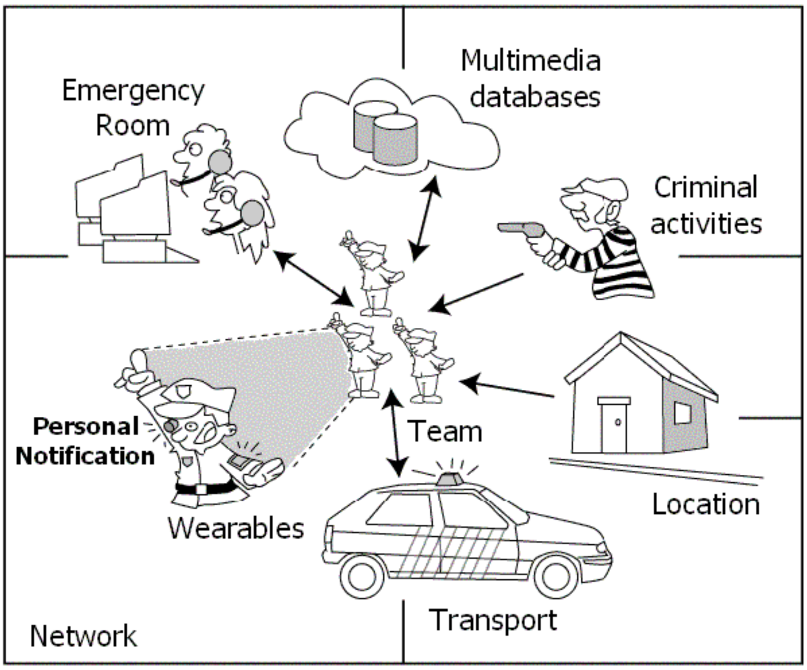


Figure 2-2. Concept of context-aware mobile support for police officers (illustration by Kim Kranenborg).

mode of transport. Based on these factors, individual officers receive personal notification at the right time and place, possibly via a wearable or handheld device. This system

1. attracts the attention of the user to high priority events and objects ("notification")
2. adapts to the attention of the user ("attentive")
3. adapts to the individual user and context ("personalized and context-aware")
4. supports task allocation within a team of police officers
5. uses multimodal input and output means.

To validate this high-level concept to police work processes and domain knowledge, a focus group was conducted with police end-users. In this one-day focus group, eighteen executive and management personnel of the Dutch Police organization participated voluntarily. A brainstorm discussion was held on innovative usage scenarios and use contexts from different perspectives, because the purpose was to gather as much and diverse information as possible.

During the focus group, the proposed concept was explained to the participants and they brainstormed about possible work situations where the context-aware system could have added value (see Figure 2-3, left). Participants used examples from the processes of criminal investigation, emergency aid and law enforcement to illustrate expected benefits of

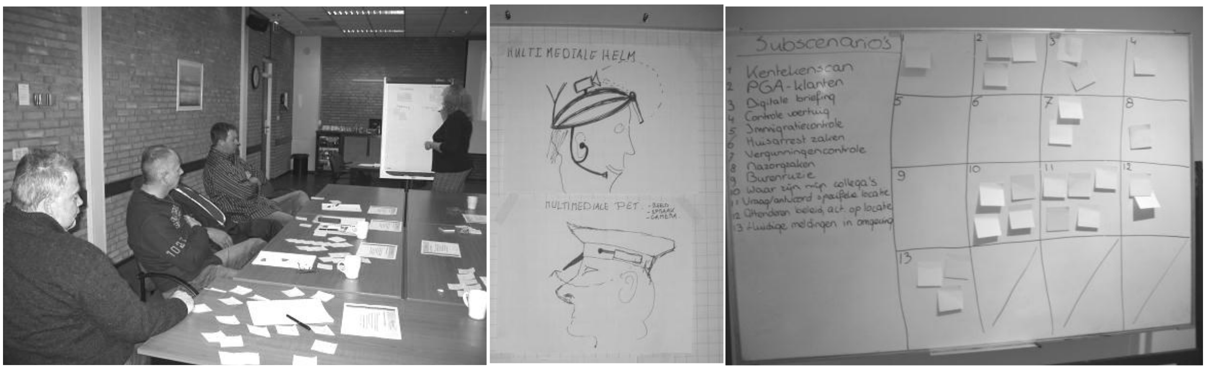


Figure 2-3. The police officers' focus group (left), illustrations of context-aware support (center) and the voting mechanism (right).

the context-aware mobile support system. In four small groups (separate for police officers and management personnel), ideas were written down in the form of short stories with illustrations. Participants were stimulated to "think out of the box" to give room to new and refreshing ideas and not to focus directly on technical feasibility. The groups were moderated by a researcher who guided the discussion and recorded comments. At the end of the session, a plenary vote was taken on which stories illustrated the most added value for police work practice and would be used further. Every participant could give one positive and one negative judgment and the totals were added up for every story. The scenario with the highest positive score and demonstrating the most benefit of context-aware mobile support was selected for further use.

2.7.2 Scenario specification and assessment

The resulting scenario (partly described in Table 2-3) clearly illustrates problems police officers experience when on surveillance. Examples include quickly assessing when a call is relevant for them or what priority a call has. Police officers considered these problems to result in unwanted interruption from their surveillance task and to hamper optimal task execution. They believe the concept solved these problems, with benefits including early warning and notification of incidents and dynamic task allocation between team members. From the scenario in Table 2-3 (simplified for illustrative purposes), crucial constraints were identified. These constraints, combined with the factors from Table 2-2, were translated into an initial set of core functions for context-aware support for mobile police officers.

A second focus group with twelve police end-users was held to evaluate the scenario and the list of core functions. This focus group took half a day and was similar in setting to the first. Its purpose was more restricted however, because participants were instructed to think of *realistic work situations* where a context-aware system would provide added value. Based on discussion of the scenario with participants, the list of core functions was extended. Again, a plenary vote decided which core functions were either indispensable, necessary, or merely worth considering (see Figure 2-3, right). This process resulted in a prioritized list of domain-specific core functions for CAMS, validated by end-users.

2.7.3 Core functions specification

The second focus group resulted in a set of five core functions for context-aware mobile support. These core functions are illustrated in the scenario and are presented in Table 2-4. In short, the system will need to adapt the system behavior to the attentional state of the user, notify the user of relevant information, support task switching and information presentation and personalize the interaction based on a user model. These core functions require the system to have access to information on location, attentive state and

Table 2-3. Scenario from the focus group, illustrating context-aware mobile support (core function codes between parentheses refer to Table 2-4).

Attention-guided police work
Youth officer Jason receives a briefing before work. His team gets an assignment (U) to check out a disturbance report and is instructed to focus on pick pocketing and gang-activity. Jason straps his wrist display on and enters in his profile that he wants to be informed of an important court case that is serving today (U). While cycling on patrol in the centre of Utrecht (U), his location aware system notices that Jason is near a hotspot of gang-activity. A colleague entered this information in the database yesterday. The system notifies Jason that he is close to the hotspot (N) and Jason decides to check six adolescents sitting nearby.   *While Jason talks to the adolescents, a call comes in about a burglary. The system holds back the notification, because it senses that Jason cannot be disturbed (A) and the priority is low (T). The notification is kept to a minimum by only showing an icon on the display. When Jason is finished with the youths, the system gives a vibration to his wrist. Jason notices the icon, presented on a map (T). He cycles to the indicated location while the system gives him auditory directions.*   *When Jason arrives, he asks the system to provide him with the relevant procedure for handling a burglary case (I). The correct procedure is accessed by the system and the wrist display shows the actions Jason has to take (T). He is filling out the form on the burglary, the last action in the procedure, when suddenly he receives an urgent notification (N) about a team member in need of backup. He rushes to his colleague, abandoning the task. When he has verified all is well with the officer, his system notifies him that there is still a form waiting to be filled out (T). When Jason is finished, he sends his information to the database (I) and returns to his patrol. His wrist display beeps and vibrates, drawing his attention to the screen (N, A). The court case was successful and the perpetrator was convicted.*

Table 2-4. Overview of the core functions for context-aware mobile support.

Core function	Description
Adaptation to attention (A)	The context-aware system knows when the officer is available, when he is "interruptible". The interaction is adapted to fit the attentive state of the user and relevant context. E.g. in the scenario, the system postpones non-necessary interruptions.
Notification (N)	Notification of high-priority objects and events through multiple modalities based on relevant task and context factors. E.g. Jason is notified of the hotspot based on his location and current activity.
Task switching support (T)	Prioritization of tasks ensures officers are working on the most relevant task. When switching is necessary, the system guides the user through a switch with interface adaptations. E.g. Jason can resume working on the procedure just where he left off.
Information support (I)	The system supports the teams of officers with relevant task information such as incident details, procedures and status of team members. This will help their shared situation awareness and task allocation. E.g. the system presents Jason with the relevant procedure on the burglary case.
User model (U)	The user and system can both change the user model which guides the adaptivity of the system. The user model contains preferences, user history, team duties etc. E.g. Jason can tell the system to notify him when necessary.

preferences of the user. In addition, information on team members' activities, duties and their task priorities is necessary.

2.7.4 Claims

The core functions in Table 2-4 are the result of a first assessment iteration and necessarily not exhaustive. For each function, an initial set of claims for CAMS systems are derived from this baseline. These claims and associated system features are specified and elaborated in each separate study in this thesis (Chapters 0 – 9). In each study, the claims form the basis for testable hypotheses that are evaluated in controlled empirical studies. The system features are designed and implemented in (semi-functional) prototypes suitable for evaluation purposes. Based on the results of the studies, the design rationale for CAMS is further refined. We will discuss the evaluation approach in more detail in the next chapter.

2.8 Opportunities for context-aware mobile support

Implementing mobile technology in professional domains provides opportunities to support mobile information exchange and team collaboration. From a work domain and support analysis in the police domain, operational demands were identified for the primary processes of emergency aid, criminal investigation and law enforcement. Police officers on surveillance need accurate and timely information access, early warning and notification of incidents, prioritization of incidents, and support for team awareness and task allocation. Context-aware mobile support systems have the potential to support these needs. To this end, specific task (activity, priority, mode of transport), context (location, time, social interaction) and user factors (preferences, expertise, history) must guide the adaptive system behavior. CAMS systems should adapt information presentation and modality (e.g. appropriate timing and presentation of interruptions) and task allocation in order to support mobile team collaboration. The concept, scenario and core functions for CAMS were specified and assessed with police end-users.

In designing CAMS systems, care should be given to create relevant and meaningful adaptive system behavior through accurate predictions on user goals and behavior. The challenge in designing these systems lies in realizing system behavior that is still meaningful for the user, given a current or future task (Shafer et al., 2001). Users will need and want to understand why certain adaptations were made, especially when these adaptations change over time, place and user. As such, a trade-off exists between adaptivity and predictability or consistency (Paymans et al., 2004). Empirical evaluation through user tests and field studies is needed to find how to balance this trade-off. In the next chapter, we describe how to select, combine and tune appropriate evaluation methods and techniques for context-aware mobile support in professional domains.

3 EVALUATING CONTEXT-AWARE MOBILE SUPPORT

Abstract

Evaluation is an indispensable part of cognitive engineering for mobile HCI, as it helps to refine and validate design solutions in order to establish adequate user experiences in the mobile domain. For mobile user interfaces in dynamic and critical environments, user experiences can vary enormously, setting high requirements for evaluation. Furthermore, evaluating desktop computing experiences does not transfer one-on-one to mobile computing. This chapter presents a framework for the selection, combination, and tuning of evaluation methods. It identifies seven evaluation constraints that influence the appropriateness of the method, specifically the development stage, the complexity of the design, the purpose, participants, setting, duration, and cost of evaluation. We applied this framework to our case study on context-aware mobile support for the police. The framework helped to select a combination of appropriate methods in different settings (such as Wizard-of-Oz, game-based, and field evaluations) and gather a concise, complete, and coherent set of user experience data (such as performance, situation awareness, trust, and acceptance). Using the framework we succeeded to capture the mobile context and its relation to the user experience.

This chapter is based on the previously published material:

Streefkerk, J.W., van Esch-Bussemakers, M., Neerincx, M., & Looije, R. (2008). 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.

3.1 Introduction

In designing mobile support systems following the Situated Cognitive Engineering (SCE) method, evaluating designs at various stages in the development process is used to refine and adjust the design when needed. Furthermore, evaluation validates that the user needs and requirements are met for the intended user group. Thorough evaluations are required when the risks and costs of errors are high, when innovative interactive support systems, such as context-aware systems, are developed, or when the system is designed for use in a dynamic and critical environment. These needs for evaluation are even higher for mobile user interfaces, because of 1) the dynamic use context, 2) specific constraints of devices and 3) risk of negative transfer from desktop experiences to mobile experiences (Nagata, 2003; Tamminen, Oulasvirta, Toiskallio, & Kankainen, 2004).

Due to these three issues, the user experience of mobile user interfaces is still an important bottleneck for services in the professional domain (Marcus et al., 2006). User experience encompasses the cognitive, affective and social responses that are induced by the use of a product or service. Realizing adequate user experiences is done by selecting the right evaluation method, based on specific constraints for evaluation of mobile, context-aware applications. Combining methods should capture the dynamic context factors and their relations to the user experience in a complete, concise and coherent way (cf. Neerincx et al., 2008). Finally, tuning of techniques and measures should ensure that the obtained results are relevant to the application domain.

This chapter presents a systematic approach to the assessment of context-aware mobile support (CAMS) systems for professionals, guided by the research question: *Which methods and measures should be used to evaluate context-aware mobile support systems in mobile professional domains?* We will focus on how evaluation in the mobile professional domain differs from evaluation in other settings. This results in a framework of seven evaluation constraints, which helps to specify when to use which evaluation methods, techniques and metrics. First, the constraints in the framework are described in depth in section 3.2. Then, in section 3.3, our framework is applied to designing and evaluating CAMS for mobile police officers, outlining the evaluation steps we will take in the rest of this thesis. Throughout this chapter the police officers' work domain will be used as application domain.

3.1.1 Application Domain

The professional domain can be characterized as an environment where mobile workers are dependent on correct and relevant information to make critical decisions, where individuals are trained for their tasks and where tasks are goal-directed. Evaluation for the professional domain is distinguished from other domains by the following three aspects. First of all, evaluation methods and measures should be tuned to specific user experience criteria within the application domain. For example, it seems less relevant (although interesting: see p. 31) to ask police officers about their emotional response towards the interaction with a mobile device. It seems more relevant to measure the increase of criminal cases that get solved in a specific period of time with the new device. Second, not all situations for which the device is intended can be assessed in the field. Situations may not happen frequently enough or the risks are too high. For these situations, other research settings such as simulations or synthetic task environments may prove useful. Third, access to professional end-users for evaluation purposes may be limited due to busy schedules and limited resources.

3.1.2 Evaluating within the mobile use context

Mobile systems have been assessed insufficiently in their use context (Goodman, Brewster, & Gray, 2004). Traditionally, evaluation is limited to laboratory settings and lacks the use of methods such as survey research, case study research and field evaluation in real use contexts that give validity to the research results (Kjeldskov et al., 2003). Recreating central aspects of the mobile use context in the lab has been shown to be sufficient for identifying usability problems (Kjeldskov, Skov, Als, & Hoegh, 2004). But field evaluation has an added benefit: deeper insight into the effects of the mobile use context, such as social interaction, lighting conditions, body movement and unreliable wireless networks, on the user experience (Zhang et al., 2005; Duh, Tan, & Chen, 2006). For CAMS, use context is especially important as it changes more than the use context of desktop applications, thereby influencing user performance, acceptance and the user experience.

The lack of field evaluation characterizes the professional domain as well. Only a few field evaluations of context-aware systems for professionals have been documented in the literature. In one effort to design context-aware support for firefighters, the application was evaluated with end-users in a focus group setting, although a field study was used to guide the initial design (Jiang et al., 2004a). Results showed that the application was accepted by the firefighters, and it supported their work practices. However, the researchers state that field testing of the application is necessary to get deeper insight into how the application influences work organization and processes. A related project Freeband FRUX aims to design mobile applications for police and rescue workers (Eijk et al., 2006) by incorporating end-users in the analysis stage and a final field test in the use environment. However, these projects are exceptions, stressing the need for a comprehensive approach to evaluation.

One of the problems with using real use contexts may lie in the fact that traditional evaluation methods are insufficient and inappropriate for evaluating context-aware applications in dynamic environments (Vetere, Howard, Pedell, & Balbo, 2003; Kellar et al., 2004; Zhang et al., 2005). A shift can be seen towards employing new techniques to sample the user experience *within the use context*. Examples are a heuristic walkthrough especially developed for mobile use (Vetere et al., 2003), and a context-aware questionnaire, which is presented to the user after a specific event. This results in more specific user reactions than using a general questionnaire (Kort & De Poot, 2005). These solutions are still in the development stage. Concluding, evaluation of mobile context-aware systems is lacking a coherent and concise set of methods, techniques and tools to “chart” the user experience in context. A more elaborate framework is necessary which takes into account the specific constraints of context-aware computing in the mobile, professional domain. This framework should provide guidance for the selection, combination and tuning of evaluation methods. Furthermore, it should be flexible enough for evaluators who have different expertise and preferences. Finally, it should apply to other professional domains where mobile context-aware applications are designed and evaluated.

This chapter proposes such a framework and applies it to evaluating context-aware mobile support in professional domains, specifically the police domain. The framework is not intended to fully capture all existing evaluation methods, but to provide a systematic and practical approach for evaluation of professional mobile systems and present a “core” set of methods. It should be noted that this framework is general and can be applied to other evaluation methods than the ones mentioned in this chapter.

3.2 Framework of evaluation constraints

We aim at an effective and efficient use of evaluation methods at different moments to improve the quality of design solutions. However, selection of techniques is not straightforward as researchers are confronted with a diversity and multitude of evaluation methods and techniques. Kjeldskov and Graham (2003) propose a categorization of current mobile HCI research methods on the constraints of setting and purpose. They signal a lack of basic research and promote the development of theoretical frameworks to better describe, compare and understand evaluation methods. Another framework for usability research methods for mobile devices is presented by Zhang and Adipat (2005). It emphasizes the setting of the evaluation (field vs. lab) based on the need to evaluate the application in context. While the frameworks help to select a particular research method, both lack specific guidance for deciding between and combining different evaluation techniques and measures in the evaluation of context-aware systems for professionals. Combination of methods should result in a more complete and sound knowledge base for design decisions, e.g. by complementing and cross-validating results between methods. Further tuning of methods should ensure that results are relevant to the application domain. As introduced in section 2.2, our framework distinguishes the following seven constraints that influence which methods, techniques and measures can be employed:

- *Stage*: Which stage of the development process is the design currently in?
- *Purpose*: What is the purpose of the evaluation?
- *Complexity*: How complex is the design?
- *Participants*: Who are your participants?
- *Setting*: In which setting will the evaluation take place?
- *Duration*: What is the duration of the evaluation?
- *Cost*: What is the cost of the evaluation?

3.2.1 Stage in development process

The development process for mobile context-aware applications can be separated into an analysis, design and implementation stage. Mobile design solutions can be evaluated at every stage in the development process both within and outside of the actual use context.

The stage of the development process determines which techniques can be employed and what can be presented to participants during the evaluation. In early analysis stages, only high-level concepts and usage or problem scenarios are subject of an evaluation. In addition, the mobile work environment and tasks of professionals are analyzed, identifying tasks in need of support, problems in task execution and appropriate characteristics to guide the context-awareness of the application. The focus is on gathering as much and diverse information as possible. In intermediate design stages, early versions of the adaptation model, mobile design solutions and support for professionals' tasks can be evaluated on usability, appropriateness and suitability for current work practices. Near the end of the process a functional demonstrator or prototype can be implemented. A benefit of early evaluation is that design flaws or errors are uncovered relatively early. Sometimes it suffices to evaluate only parts of a system, such as support for a specific task. Early prototyping and field testing is even more important for mobile applications than desktop applications as the usability of the mobile application is very dependent on the device used and the dynamic context (Zhang et al., 2005). Here, evaluation provides an important proof of concept that the adaptation model and application result in meaningful support.

3.2.2 Purpose of evaluation

A second constraint is the purpose of the study. For mobile, context-aware applications, purpose can be gathering factors on which to base adaptive system behavior, evaluate influence of environmental factors and mobility or evaluate suitability for a specific task. We distinguish between formative methods, used to generate design solutions, and summative methods, used to measure acceptance of designs. Within our framework, formative evaluation can be used to identify the factors on which to base the adaptive behavior. Contrastingly, summative evaluation focuses on how the system impacts the work processes of professionals and the correctness of the adaptivity model.

On a more fundamental level, the innovativeness of context-aware mobile systems also determines the purpose of evaluation. These evaluations must often take place without established benchmarks or design guidelines. In our case study, the evaluation of innovative concepts is adapted to specific police contexts and tasks. This purpose is in contrast with redesigning or improving existing applications.

3.2.3 Complexity of design

How complex the design is constitutes the third constraint. Complexity in adaptive systems can be defined as "the directness of transformation from user input to system output" (Zipf & Jost, 2006) i.e. the adaptive system behavior. Design solutions with different degrees of complexity need different evaluation approaches. This point is closely related to the innovativeness of the system. The evaluation of a calendar application on a mobile phone requires a different set of techniques and measures than the evaluation of a context-aware adaptive system. However, for mobile devices evaluation should always be fitted to the dynamic use context. A factor that further increases the complexity is the fact that users themselves also show adaptive behavior towards the system. A system that dynamically adapts to changing user behavior can cause unpredictable effects.

From a user perspective, evaluating adaptive systems means evaluating the appropriateness of the adaptive behavior, given the context and user task. Optimally, the system should be tested in the use context, because the adaptations are based on this use context. Depending on the goal of the evaluation, the question is whether or not to make the underlying rules or model explicit for users. Often the goal of an adaptive system is to seamlessly support the user's flow of work, making comparison to non-adaptive systems difficult or irrelevant (Weibelzahl, 2005). In other situations, the adaptation rules or models need to be made explicit in order to be evaluated. Here, a "modular" approach could be adopted by evaluating the appropriateness of the input, the model and the resulting behavior separately. This approach provides adequate feedback into the design process (Paramythis et al., 2001).

3.2.4 Participants

The fourth constraint is choosing the right participants and the right number of participants for testing. Evaluation shows to which extent the design meets the requirements of the end-user group. For professionals such as police officers, their diverse roles, skills, training and experience impose specific requirements on the design (e.g. Pica et al., 2004). Determining these requirements, characteristics and needs of users is the first step in evaluation. Next, during evaluation, an assessment is made how well the adaptive system supports specific roles or tasks. Often in professional settings, access to end-users is limited and deciding which method to use must take into account the availability of participants. When no actual end-users can be involved, a careful selection of participants has to ensure they are representative of end-users (e.g. in age). Depending on the research question, student participants can be employed, provided they are trained on the evaluation tasks. End-users are particularly necessary during the analysis and implementation stages because of their knowledge of the mobile and dynamic use context and work processes. In addition, prior training on or experience with certain tasks has to be taken into account, as well as prior experience with mobile devices. Negative transfer from desktop experience to mobile experience can cause longer task execution times and more switching between tasks (Nagata, 2003). End-users' experience may sometimes hamper evaluation, as their standard work routines may limit their openness to innovations.

3.2.5 Setting of evaluation

Furthermore, the setting of the evaluation is of importance. The setting of mobile systems evaluation can be defined as environment independent, natural or artificial (Kjeldskov et al., 2003). Environment independent methods are not situated in the use context. Their focus is on creating a general overview of system use instead of describing specific tasks. For context-aware mobile systems, gathering information about the use context is particularly important during the analysis stage. Hence, contrasting to evaluation of desktop applications, environment independent methods must be combined with methods that provide a rich description of the dynamic use context. The results can be captured in for example scenarios, storyboards and use cases.

This contrasts with the natural or artificial setting of task-based evaluations. In essence, choosing between a natural or artificial setting is balancing a trade-off between the degree of reality of the evaluation setting and control over extraneous variables. The purpose of evaluation in a natural setting is proving that the system works as intended in a realistic use environment. For example for context-aware systems, the correctness of the adaptive behavior with respect to the use context is evaluated. However, when a high degree of control over extraneous variables is needed, an artificial laboratory setting can be used. Recreating or simulating core task features from the use context in the lab has specific benefits for evaluating professional systems. In this domain, field evaluation may interfere with ongoing work and imposes on the time of participants. In addition, situations for which the design is intended may not happen frequently enough to evaluate them properly, for example large-scale disasters. In this case, a good alternative is to simulate the use context and test the context-aware system in the lab (Te Brake, de Greef, Lindenberg, Rypkema, & Smets, 2006). Finally, if actual mobile use is subject of evaluation, simulation of an application on a real mobile device has advantages over simulation on a desktop computer. Specific constraints for the device (e.g. screen size) and environmental factors (e.g. low bandwidth) are taken into account during the evaluation, providing more realistic results (Zhang et al., 2005).

3.2.6 Duration of evaluation

The duration of the evaluation is constrained by the type of data that is collected during evaluation. Some data can be collected relatively fast and easy by interacting with a prototype for a couple of hours. Examples include usability questionnaires or task performance data on a specific task. This data is focused and specific, i.e. only valid for the task and can not be generalized to other tasks and settings. In contrast, evaluation in a longitudinal study gives deeper insight into how learning effects, the dynamics of trust and user experience develop over time. These measures are particularly important in evaluating mobile, adaptive systems. Interpretation of this general, broad data makes it necessary to take into account the whole use context (Kort et al., 2005). Tuning measures to the application domain can be done by relating them to performance criteria for professionals (Neerincx et al., 2008).

3.2.7 Evaluation cost

As the final constraint, the cost of an evaluation can be expressed in time and resources. Thus, the cost-effectiveness of the evaluation method can be viewed as the amount and severity of uncovered design flaws versus the cost of investing time and resources. For evaluating mobile applications, video logging with behavior analysis is a widely used but time consuming and expensive method of which the added value remains debated (Kjeldskov et al., 2005). Recent comparisons between methods show that rapid review by experts is a very cost effective procedure, uncovering the majority of critical usability problems in a short time. However, for evaluating mobile adaptive systems, issues like ecological validity of the design can only be tested in field situations. These studies entail higher costs due to the mobility of the setup and the participation of professional end-users. Furthermore, there is less room in the professional domain for flawed designs leading to usage errors, calling for a more extensive evaluation. Possible cost-efficient solutions for evaluating these innovative systems are using Wizard-of-Oz prototypes or simulations.

3.2.8 Conclusions

From the discussion so far, it is clear that the stage, purpose, complexity, participants, setting, duration and cost each impose constraints on which evaluation technique to use. When these constraints are specifically applied to evaluating context-aware professional user interfaces, the following can be concluded.

- Evaluation within and outside the use context with participation of end-users can take place at every stage in the development process, each stage having its own focus.
- The purpose of evaluation is influenced by the innovativeness of the system and determines whether formative or summative techniques are used.

- Evaluating complex, adaptive systems in the use context increases appropriateness of final designs.
- Actual end-users should be involved as participants as much and as early as possible because of their intimate knowledge of mobile use context and domain-specific tasks.
- Information about the dynamic use context must be gathered as early as possible. Furthermore, when access to the actual use context is restricted, simulation yields a realistic yet controlled evaluation setting.
- Evaluation over longer periods of time in the mobile application domain is particularly important to gather rich, broad user experience data.
- Using simulation tools can reduce the cost of evaluation, but the user experience and ecological validity are preferably evaluated in (relatively expensive) field testing.

It is important to note that all seven constraints are interdependent. For example, the setting of an evaluation depends on the complexity as it makes little sense to evaluate a non-functional, paper prototype for a specific police task in a real-life setting. However, each constraint has its own unique contribution to the selection of techniques.

3.3 Applying the framework to evaluate context-aware mobile support

The SCE method introduced in section 2.2 is a general methodology to design cognitive systems, but does not specify when to use which evaluation method. In this section, the assessment cycle of the SCE method (see Figure 2-1) will be filled in, by applying the framework to the central case study in this thesis: the police domain. Based on the evaluation framework described above, for each study the design stage, purpose, complexity, participants, setting, duration and cost are specified. This process results in the incremental design and evaluation approach in Figure 3-1, specifying one field study and five experiments to realize context-aware mobile support. As the evaluation steps progress, the interactivity of both the prototypes and the evaluation environment increase. For each step, the rationale behind the evaluation methods, techniques and metrics is described. We will address the evaluation methods of ethnographic field study, “Wizard-of-Oz” evaluation in lab settings, mobile experimentation and evaluation using synthetic task environments. Alternative methods and techniques to the ones employed here will be discussed in separate boxes. The actual studies will be described in the subsequent chapters of this thesis.

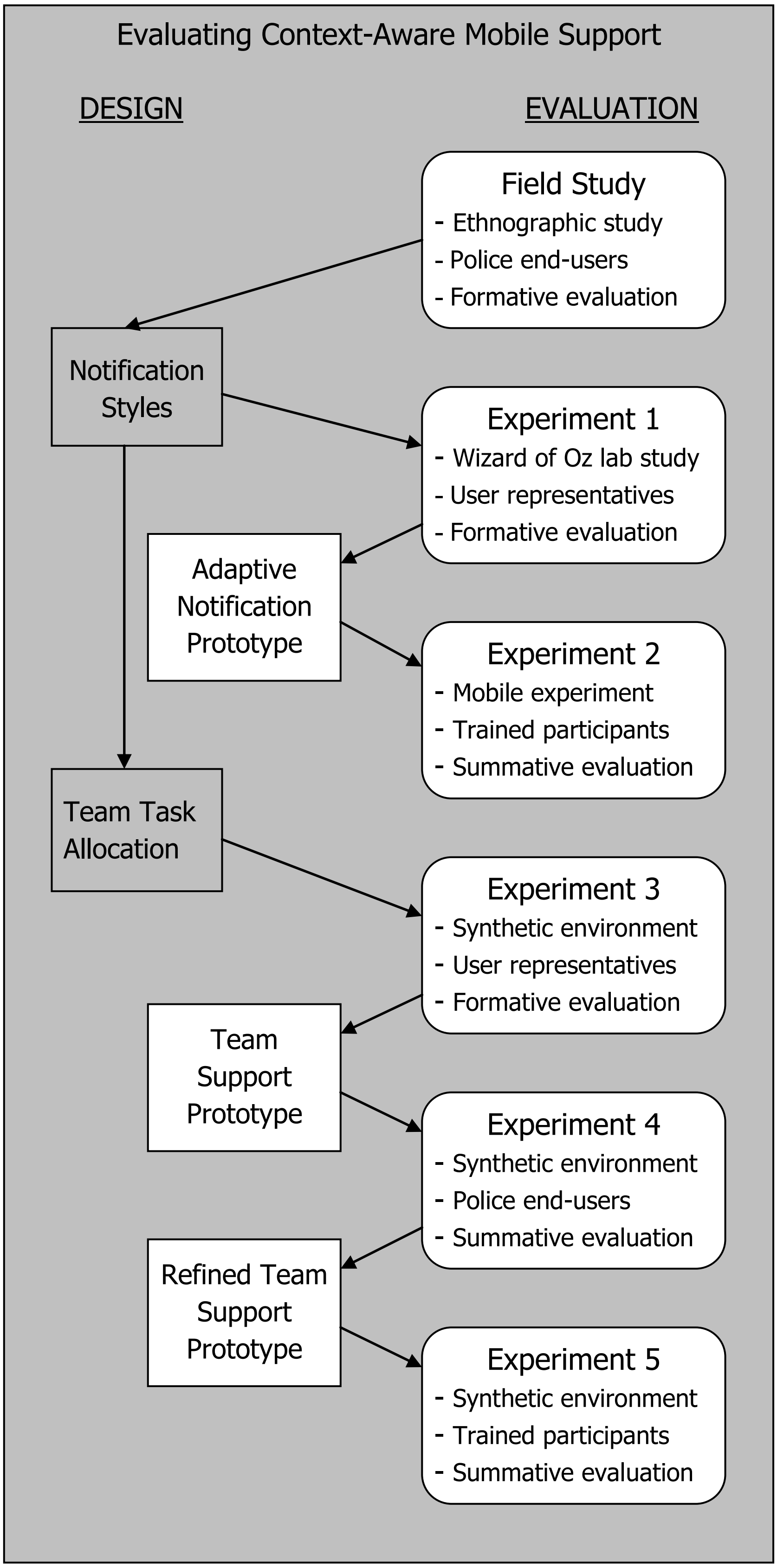


Figure 3-1. Incremental design and evaluation approach to realize context-aware mobile support.

3.3.1 Ethnographic field study

Field evaluation is conducted in natural work environments, and can be conducted during any stage in the development cycle and over longer periods of time. It requires a stable and reliable functioning system, participation of end-users and mobile data gathering tools. As the functioning of the mobile system is dependent on the dynamic context and often unreliable wireless networks, evaluating context-aware support systems in the field provides validation that the design works as intended. The added benefit of field evaluation over other methods has been criticized and disadvantages are possible interference with ongoing work, difficulties to encompass the richness of mobile contexts, and the difficult data collection and control due to the dynamic context and physically moving users (Kjeldskov et al., 2004; Goodman et al., 2004; Zhang et al., 2005). Field studies, combined with environment independent methods such as focus groups, interviews and surveys (see Box 1), can be used in early stages for user requirements analysis.

Box 1: Environment independent evaluation

Focus group
Focus groups provide a powerful method for envisioning future system use. For designing mobile, context-aware systems, this evaluation method is best used in an early stage of the process, when user requirements need to be defined for the system. During a focus group session a small, selected group of people is brought together for an interactive and spontaneous discussion on a specific topic. The purpose of a focus group session is to gather broad information and to get insight into user needs and opinions through interaction between group members. Focus group research can be used for evaluating concepts, scenarios and high-level user requirements (see for example the focus group with police officers in section 2.7.1).

Interview
An alternative technique is interviewing domain-experts or expert users such as police end-users. Questions are asked to get expert opinions and deeper understanding of the problems in the domain. For designing in the professional domain, interviews with police end-users can help establish domain-specific evaluation criteria to which a context-aware system can be judged.

Survey
Distributing questionnaires to the end-user population is a way of getting a large quantity of opinions from a diverse group. Specific advantages of survey research to evaluation in the professional domain are that surveys provide an overview of the police organization and allow user needs to be related to specific roles. For example, a police officer on surveillance may have different needs than a police officer visiting a crime scene.

References

- Jiang et al. (2004)
- Kjeldskov and Graham (2003)

Figure 3-2. Police officer participating in the field study.

As a starting point for the design of CAMS, we studied an innovative mobile notification system within the police domain (described in detail in Chapter 0). This location-based notification system (LBNS) for mobile police officers was implemented in 2007 by the Dutch police organization. The service notifies police officers of incidents and criminal activities in their direct vicinity. It shows police officers the current location of colleagues and supports police procedures. The LBNS can be regarded as a first iteration of a context-aware support system, in that it uses context (location) to trigger notification messages. This system constitutes the current state-of-the-art in mobile computing for law enforcement.

The evaluation of this mobile notification system was done using an ethnographic field study (see Box 2). An ethnographic study describes the functioning of the mobile system in the work domain and requires presence of the researcher during work activities. The purpose of this study was to validate the full functioning of the location-based system with end-users and to provide a “proof of concept” in the application domain. Through the ethnographic field study we gained insight into the functioning of the system with respect to police processes and organization. The system was evaluated in a longitudinal study to measure impact on work processes, trust, acceptance and learning effects. The cost in both time and resources were relatively high compared to other methods, as the police organization and personnel had to participate. In addition, collecting and analysing data in the field was necessary, further increasing the cost. Thirty officers with different roles, such as emergency aid, district surveillance and prevention participated in this evaluation. After an initial training phase, the system was used during daily work for a period of four months. The constraints of this study, based on the framework, are:

- *Stage*: final implementation stage of functioning system.
- *Purpose*: summative; provide a “proof-of-concept” within real-life context.
- *Complexity*: high; full system functionality.
- *Participants*: thirty experienced police end-users.
- *Setting*: natural use environment; police surveillance.
- *Duration*: longitudinal study during four months.
- *Cost*: high in time, low in software resources (as the system was already developed).

This evaluation focused on the user experience in context, integration of the system in work practice and acceptance within the organization. Techniques included participatory observation, interviews and questionnaires. To evaluate the user experience, critical incidents in task execution with the system were reported weekly by the officers. These reports were then related to the specific context variables logged by the system. Researchers conducted monthly participatory observation sessions on surveillance with officers. This technique aimed at getting deeper insight into the system’s impact on work processes. The system was evaluated in a pre- and post-test setup, thereby giving insight into learning effects and changes caused by the system. Prior to evaluation, the expected effects of the system were captured in specific performance criteria. These criteria were tuned to the police application domain: the amount of fines collected, response time to calls and amount of time spent on surveillance are important measures. This data was collected by recording events from police databases and analyzing system events on the PDA.

The benefits of this method were that observing professional end-users interacting with the system in their work environment gives insight into usability, user experience and impact on work processes. In addition, only in field studies could the system be assessed in the actual and diverse work situations that occurred naturally. This was a necessary and valuable step before actual implementation of the finished system as it allowed final changes and tuning of the system. However, limitations were that this method was also costly as an advanced prototype, a mobile evaluation setup and the participation of end-users were required. In addition, judgments by participants (i.e. reporting critical incidents) should be given as fast as possible to avoid recall problems, but this may interfere with ongoing work.

Box 2: Evaluation in natural settings

Field evaluation is conducted in natural settings, often during the final phase in the development cycle and over longer periods of time. It requires a stable, reliably functioning system, participation of end-users and mobile data gathering tools. As the functioning of the mobile system is dependent on the dynamic context and unreliable wireless networks, evaluating context-aware support systems in the field provides validation that the design works as intended. The added benefit of field evaluation over other methods has been criticized and disadvantages are possible interference with ongoing work, difficulties to encompass the richness of mobile contexts, and the difficult data collection and control due to the dynamic context and physically moving users.

Ethnography
To study mobile applications use through (rapid) ethnography, researchers immerse themselves in the work practice. They meticulously describe the context and common practices of the domain. This technique leads to deeper insight into end-user practices in their natural work setting. This insight is of extra importance to understand the dynamic context of mobile end-users. For example, studying the police work environment provides a detailed description of common and uncommon tasks and critical incidents that a context-aware system can support.

Ethnographic field studies and field experiments
Evaluation in natural settings can be distinguished in ethnographic field studies and field experiments. An ethnographic study describes the functioning of the mobile context-aware system in the work context and requires participation of the researcher in the work activities. Contrastingly, field experiments test two versions of a context-aware system under different conditions to evaluate the influence on task performance. Mobile experimentation in the field allows for more control but can only be used for restricted evaluation purposes, such as usability evaluation.

References
- Goodman et al. (2004)
- Kjeldskov et al. (2004)
- Zhang and Adipat (2005)

3.3.2 *Lab-based Wizard-of-Oz evaluation*

Both from the field study as well as the list of core functions in Chapter 2, the requirements of notification and adaptation to attention were the first and most important focus of designing CAMS. The CAMS system should notify police officers of relevant information in their environment, without distracting them unnecessarily from their primary surveillance task. What was needed was a system that presented incident messages to the user in different notification styles, based on user workload and the message priority. Notification styles were designed by adapting the salience (e.g. auditory volume) and information density (e.g. summarizing text) of the notification. This adaptive notification concept was evaluated by implementing it on a handheld computer (PDA).

To evaluate the support concept of context-aware notification styles, we selected a lab experiment with a Wizard-of-Oz setup (Experiment 1 described in Chapter 5). The Wizard-of-Oz (WOz) method is widely used in evaluation of mobile context-aware applications (see Box 3). It involves letting participants interact with a seemingly functional system that is actually operated by the researcher (Dahlback, Jonsson, & Ahrenberg, 1993). We selected this method based on the following considerations. The purpose of the evaluation was to guide the further design effort. We needed a simulated setting that allowed us to recreate basic aspects of the police officers' surveillance task. Furthermore, a flexible environment was necessary because we were testing the influence of changes in the context (e.g. workload and message priority) on interaction with a mobile device. Finally, the WOz setup allowed us to empirically test our concept by systematically comparing two conditions. As our concept dealt with general instead of task-specific abilities, a representative participant group was used. The constraints of this study, based on the framework, are:

- *Stage*: early design stage; first iteration of requirements evaluation.
- *Purpose*: formative; to validate context-aware notification styles.
- *Complexity*: moderate; (simulated) context-aware system behavior.
- *Participants*: twenty; representative to police officers.
- *Setting*: artificial; lab experiment employing Wizard of Oz setup.
- *Duration*: short; two hours per participant, two weeks in total.
- *Cost*: moderate in time (data-analysis) and software resources (prototyping).

This study recreated basic aspects of police surveillance (divided attention, notification) in the lab. Twenty participants were involved in this study, representative to end-users in age and education. They performed a simulated police surveillance by watching videos, reporting targets and answering questions on these videos. Simultaneously, the test-leader sent low, normal or high priority messages at predefined moments to the PDA (cf. the WOz setup). Participants needed to recognize and report the messages. Adaptive notification (different notification styles) was directly compared with non-adaptive notification (uniform notification styles) in a within-subjects design. Each evaluation took approximately two hours, including training, two scenarios and debriefing.

Both qualitative and quantitative techniques were combined in this evaluation. Performance data (time on task, number of errors) were collected using event-logging on the PDA. Subjective judgments (notification intrusiveness, preference for condition) were

Figure 3-3. Wizard-of-Oz setup and videos to recreate a mobile evaluation setting in the lab.

measured with rating scales and questionnaires. The specific performance measures were tuned to realistic aspects of the police officers' surveillance task. For example, the messages were representative of police reports. In addition, participants had to report and describe different "targets" from the videos, which is an important surveillance skill.

The WOz setup created the illusion of a working, adaptive support concept (see Figure 3-3). This allowed participants to compare multiple conditions, improving the accuracy and validity of their subjective judgments. However, extra training was necessary to facilitate the distinction between the different support concepts (e.g. notification styles). In further evaluations, to accurately gather performance data and see how the mobile use context influences the interaction, a more realistic task setting is required.

3.3.3 Mobile experimentation

Based on the results from Experiment 1, the adaptive notification support prototype was refined by allowing different timing of notifications. The next step was to use this prototype in a more realistic, mobile task setting, which allowed measurements of task performance within the context and observe how the mobile context influences the interaction with the support prototype. This is termed "quasi-experimentation" because of the decreased control over extraneous variables (Tamminen et al., 2004; Roto et al., 2004). In addition, in the previous lab study, participants could only make a limited set of decisions in a task they did not directly control (i.e. by watching videos). Evaluation of the refined prototype took place in a mobile task environment, representative for police surveillance.

Thus, Experiment 2 (described in Chapter 6) employed a refined CAMS prototype in a real-life mobile surveillance task. Because the purpose was to validate different design solutions, multiple conditions needed to be compared, again requiring a WOz setup.

Figure 3-4. Participant (left) in mobile experimentation, accompanied by the test leader.

Employing the prototype in a real-life mobile setting further increased the complexity and the cost of the evaluation. Because the system was still in an early phase (e.g. selecting between different design solutions), trained student participants instead of end-users participated in this study. Thus, the constraints of this study, based on the framework, are:

- *Stage*: intermediate design stage; second iteration of requirements evaluation.
- *Purpose*: formative; further validation of context-aware notification.
- *Complexity*: moderately high; simulated context-aware system behavior while mobile.
- *Participants*: thirty trained student participants.
- *Setting*: real-life building walkthrough employing a Wizard-of-Oz setup.
- *Duration*: short; three hours per participant, three weeks total.
- *Cost*: high in time (data-analysis) and moderate in software resources (increased prototyping).

This mobile experiment employed a combination of both (mobile) data logging as well as participatory observation and questionnaires. The experiment was tuned to the police domain by employing an experimental scenario containing police incidents. Performance metrics were representative of police surveillance (response time, time on task, number of incidents handled). The downside of this mobile experiment was decreased control over extraneous variables, such as people interrupting the experiment or problems with navigation. In addition, the test leader needed to accompany the participant for observation and to operate the WOz-prototype (see Figure 3-4).

Box 3: Evaluation in artificial settings

Wizard-of-Oz evaluation
The Wizard-of-Oz (WOz) evaluation method is widely used in evaluation of mobile context-aware applications. It involves letting participants interact with a seemingly functional system (possibly in the mobile context) that is actually operated by the researcher. This avoids programming a functional context-aware system and allows for early and relatively low-cost evaluation of design solutions. However, the weakness of the WOz technique is that it requires human intervention. This technique is appropriate when no time-critical system performance is required.

Game-based evaluation
Game-based evaluation provides the best of both worlds for evaluation of mobile applications: a realistic task environment with control over extraneous variables. It provides an ideal simulation environment for task-based evaluation for professionals. Control over context factors means that the application can be evaluated under a wide variety of situations. Measurement of performance data can be done accurately due to integrated logging procedures. In addition, data gathering tools do not have to be taken into the field to evaluate mobile technology. Game-based techniques have been used frequently in learning and training environments and as simulation for crisis management situations.

References
- Dahlback, Jonsson and Ahrenberg (1993)
- Lewis and Jacobson (2002)
- Cooke and Shope (2004)

3.3.4 Evaluation using synthetic task environments

Evaluation in synthetic task environments or virtual environments provides the best of both worlds for evaluation of mobile applications: a realistic task setting with control over extraneous variables (see Box 3). It allows flexibility in recreating task-specific aspects of the use context, such as navigation, incident handling and team tasks with multiple actors (Cooke & Shope, 2004). Because they allow multiple participants to interact within the environment, team coordination and situation awareness can be reliably studied. Control over the use context means that the application can be evaluated under a wide variety of situations that can be reliably replicated each time. Measurement of performance data can be done accurately due to integrated logging procedures. In addition, data gathering tools do not have to be taken into the field to evaluate mobile technology. Game-based techniques have been used frequently in training environments (Lewis & Jacobson, 2002).

Up to now, the CAMS prototypes were evaluated by individual participants, focusing on mobile information exchange and notification. Evaluating team collaboration however required a more extensive evaluation setting, where multiple actors can collaborate. The CAMS system prototype was extended to 1) support team task allocation through advice to team members and 2) support team awareness through a map overview with location of team members and incidents. The next evaluations assessed the redesigned prototype in a rich yet controlled synthetic task environment. The simulated reality of the surveillance task through this environment required using end-users as participants, although their availability limited their participation to Experiment 4 only. Based on these considerations, we selected a lab experiment in a game-based simulation environment for Experiments 3, 4 and 5. The specific experimental designs will be addressed in each chapter (7, 8 and 9 respectively). The constraints for these studies are:

- *Stage*: final iterations in design stage.
- *Purpose*: formative and summative; test and validate context-aware team collaboration support.
- *Complexity*: high; (simulated) interactive team notification prototype.
- *Participants*: both police end-users, user representatives and trained student participants in teams of three.
- *Setting*: artificial; lab experiment employing interactive synthetic task environment.
- *Duration*: moderate; four hours, two months in total.
- *Cost*: moderate in time and in software resources (creation of synthetic task environment).

A mobile surveillance environment was created within the PC game Unreal Tournament (see Figure 3-5; for a description, see Te Brake et al., 2006). We included multiple teams of three participants, depending on availability. Teams navigated through this environment in a surveillance task, consisting of reconnaissance, gathering information and communicating with team members. In addition, participants received notifications about incidents (unconscious persons, thefts, vandalism) via the context-aware system, simulated on a PDA. Both task performance and appropriateness of the adaptive behavior were subject of evaluation. Therefore, an experimental condition with the adaptive system was compared to a non-adaptive system. The duration of the evaluation was approximately four hours to allow thorough training on using the environment and an extensive debriefing at the end.

During these evaluations a combination of qualitative and quantitative measures was collected. Performance data include time on task, number of decision errors and distance

Figure 3-5. Team of police officers participating in the experiment (left) and the game-based environment showing a victim (right).

traveled. In addition, a measure for effectiveness of the system was the number of incidents handled. Trust, acceptance and preference were measured using questionnaires and rating scales. Situation awareness (SA) is measured with a technique called “freezing” (Endsley et al., 2000), where the task environment was paused at irregular intervals to answer a question about the environment, such as “indicate on the map the location of the car accident”. In addition, the “critical incidents” technique used a think-aloud protocol to collect both positive and negative incidents in using the context-aware system. The evaluation setup was tuned to the police environment by using an experimental scenario that is established by police end users. Furthermore, the critical incidents reported by the police officers participating in this evaluation were analyzed carefully. These incidents provide suggestions for the appropriateness of the context-aware system in the field.

The synthetic task environment technique allowed for accurate quantitative measures of performance data and SA, because the behavior and navigation path of the participants were recorded. In addition, the appropriateness of the adaptive behavior could be accurately assessed by the participants. Furthermore, multiple participants worked collaboratively on one task in the same environment, allowing evaluation with teams. Finally, critical or unexpected events were pre-programmed into the scenario running in the game simulation. Three factors negatively influenced the game-based evaluation. First, asking participants to fill out SA questions and rating scales interfered with the task flow at certain moments. Second, prior gaming experience should be well documented, as this influences participants’ performance. Third, some participants are susceptible to simulator sickness, which can occur in game-based simulation (Kolasinski, 1996). This did not occur in any of the three experiments in the synthetic task environment.

3.4 Discussion and conclusion

Earlier research identified the lack of structuring in methods and techniques for evaluation of mobile and adaptive technology (e.g. Paramythis et al., 2001). The effects of this lack include inability to interpret and generalize results across applications and user groups. Traditionally the focus in designing mobile systems has been on producing engineering solutions, rather than conducting ecologically valid evaluations, leading to a prevalence of lab evaluations (Kjeldskov et al., 2003). In this chapter, we argued that evaluation of context-aware mobile support systems for professionals benefits from a systematic, incremental approach. As there is not one evaluation technique that delivers answers to all design questions, combinations of techniques have to be sought. By considering the development stage, the design complexity, the purpose, participants, setting, duration and cost of evaluation, a specific set of methods, techniques and measures can be determined. This framework of evaluation constraints was applied to designing context-aware mobile support (CAMS) for police officers. This resulted in specification for four methods for evaluation of mobile, adaptive systems (ethnographical field study, Wizard-of-Oz lab evaluation, mobile experimentation and evaluation in synthetic task environments).

The main argument in this chapter is that researchers should explicitly establish an "evaluation rationale" to match their design rationale. Our framework helped in this respect, by selecting a concise and coherent set of appropriate evaluation methods and techniques in a flexible way. By tuning these to the mobile professional domain, the framework takes into account the specific constraints for context-aware computing in this domain. Moreover, the framework can be applied to evaluation of context-aware computing in other domains. In conclusion, evaluation of adaptive mobile systems is expected to specifically benefit from this approach. It stresses the need to incorporate end-users in the evaluation, emphasizes the critical and dynamic professional environment and interprets evaluation results within the use context. In this way, both short-term usable services as well as long-term innovative support concepts for police officers are realized.

PART II

MOBILE INFORMATION EXCHANGE

4 LOCATION-BASED NOTIFICATION FOR MOBILE POLICE OFFICERS

Abstract

To increase police officer awareness of incident locations, the Dutch police developed and implemented a location-based notification system (LBNS). This mobile service notifies police officers proactively of warrants, agreements and police focal points in their current vicinity. To assess the efficiency, effectiveness and user experience of this service, a longitudinal field evaluation was conducted with thirty police officers over four months. The results indicate that using the LBNS, police officers were better informed of relevant information in their environment. Users considered the interface clear and easy to use. However, users indicated that the system presented too many or non-relevant notifications and that the system is overly complex. Recommendations for further development are to mitigate unwanted interruption by intelligent filtering of notifications and integration of system components.

This chapter is based on the previously published material:

Streefkerk, J.W., van Esch-Bussemakers, M., & Neerincx, M. (2008). Field Evaluation of a Mobile Location-Based Notification System for Police Officers. In *Proceedings of Mobile HCI 2008* (pp. 101-108). Amsterdam, the Netherlands: ACM.

4.1 Introduction

Context-aware mobile support for police officers should facilitate mobile information exchange. For police officers, increasingly more information becomes available on surveillance. They have to be aware of locations of incidents and colleagues, crime hotspots and police focal points. Most of this information is only necessary and relevant at specific locations. For example, police officers only have to be aware that a particular location is a criminal hotspot so they can act on this information when they are in the vicinity. In addition, this information is dynamic in nature, changing over time. New incidents may come up during the day, or colleagues may finish their shift and are no longer available to provide assistance. The challenge police officers face is to be aware of relevant locations at the right time or place. Furthermore, they rely heavily on information from other sources (emergency room, colleagues) which may not always be available. This entails a need for information support for police officers on the street.

Current developments in geographical information systems (GIS) and mobile internet enable location-based information presentation on mobile devices (e.g. PDA's). Location-based systems in critical domains, such as the Land Warrior System for the military, are designed to support soldiers in navigation and situation awareness (Murray, 2000). This particular system provides information about enemy presence, terrain and landmarks using GPS coordinates. However, it relies on user "information pull" (e.g. by looking at the map at the appropriate moment). In the police domain, an in-car location-based system is proposed to support communication between dispatchers and patrolling officers (Nulden, 2003). Based on GPS coordinates, the display indicates the status of all patrols and their distance to the incident location. This is expected to support more efficient task allocation and awareness of other patrols but the system was not evaluated. These efforts only partly address the mobile information exchange issue described above. Location-based notification systems (LBNS) present relevant information *proactively* (i.e. via notifications) based on the current geographical location of the user (Munson & Gupta, 2002).

The Dutch police implemented a LBNS to support three police procedures that require awareness of incident locations (warrants, focal points and agreements). These three types of locations were indicated on a geographical map on a PDA using GPS location tracking. The system notifies police officers proactively of warrants and police focal points when they are in the vicinity of such a location. When an incident occurs, police officers can check for agreements at that location (e.g. an agreement regarding noise disturbance). The PDA application employs auditory signals and pop-up screens to notify police officers and provides access to detailed incident information in police databases. Results of police actions can be inputted into the application, informing colleagues who visit the location next.

This system is expected to help mobile information exchange in three ways. First, police officers' awareness of incident locations will increase when they are notified of relevant information on location. Second, because they have information available within the use context, officers will be more self-reliant and rely less on communication with colleagues and the emergency room. Third, by allowing officers to input information based on actions they performed, information sharing between colleagues will be more optimally supported. These three effects are expected to have a positive influence on operational results, such as handling incidents faster and more effectively, a higher chance of apprehending criminals, or reducing nuisance in public spaces. In addition, these effects also positively influence the user experience of this service, because officers notice it supports their work processes. However, the actual effects of this innovative system on police work efficiency and effectiveness and the user experience must be carefully evaluated. One risk of a LBNS is officers getting distracted due to too many or non-relevant notifications. For example, when a police officer is en-route to a high-priority incident, a notification about a warrant that needs to be collected might be distracting. Thus, in evaluating this system, the trade-off between receiving valued information while minimizing non-relevant interruptions needs to be addressed.

This chapter presents a longitudinal field study of this location-based notification system in the police domain. Twenty-six police officers used this system in their daily work for a period of four months. Using information from interviews, observation, questionnaires and log-file analysis, the following research question was addressed: *What are the effects of a location-based notification system on work process efficiency and effectiveness and the user experience of mobile professionals?* The field study shows the positive and negative effects of this LBNS on efficiency and effectiveness of police surveillance. In addition, because of the long duration, it provides insight into the user experience over time as well as the impact on the police organization. In the following sections, first the system and the police procedures it was designed to support are described. Then, the method and results of the field study are presented. The chapter concludes with how the results of this field study influence the design of CAMS for police officers. This will guide addressing the issues of mobile information exchange and team collaboration in the rest of this thesis.

4.2 Location-based notification system for police officers

In the current situation, police officers have to contact the emergency room via their radio transceiver with requests about a specific location. They will ask whether agreements, open warrants or police focal points are in effect at a particular location. Alternatively, they can approach colleagues informally with questions via their mobile phone. The final way to get information while on the street is the use of paper notebooks. During the morning briefing, police officers write down incidents, addresses, names and other operational information and check their notes when necessary during the day. As police officers work in shifts, the briefing constitutes an important moment of information sharing between most (but not all) active officers. Different types of shift can be distinguished, such as emergency response or a “free” shift. During emergency response, officers are only concerned with time-critical incidents. Warrants, agreements and police focal points are generally of a lower priority than emergency response incidents. During a free shift, police officers do not have a specific task to perform. Similarly, officers with specific roles (e.g. a district officer or an officer with juvenile delinquent expertise) often have to have more detailed knowledge about incident addresses in a specific region.

Current practice shows that officer awareness of incident locations is dependent on their own proactivity (e.g. checking their notes at the right time and place). The Dutch police was interested in the potential benefits from location-based notification, such as improving officer awareness of their environment and increasing chances of apprehending criminals. Their vision was a mobile system that proactively notifies police officers on surveillance when they are in the vicinity of a location that requires their attention. The LBNS system was designed to support awareness of three types of location-specific information, particularly open warrants (such as outstanding fines), agreements on location (such as environmental regulations) and police focal points (such as criminal hotspots). Note that these three types of locations are not concerned with high-priority or time-critical incidents. Officers are still notified of such incidents by the emergency room via a radio transceiver.

- *Open warrants*. These warrants are issued when perpetrators have to pay a sentenced fine or spend time in jail. Police officers need to be aware of these open warrants, so they can visit the address when they are in the vicinity. This increases the chance of apprehending the perpetrator or collecting the fine.

- *Agreements on location*. These are locations where regulations are in effect that specify what is and is not permitted at that location. Examples are regulations about noise disturbance or bar opening times. Police officers only need to be aware of these agreements when there is an incident reported at the location. Thus, the LBNS does not proactively notify officers of agreements, but shows them in the geographical display.
- *Police focal points*. These focal points are locations with increased criminal activity. For example, when police officers are aware of burglary hotspots, they can be extra vigilant when they are in the vicinity. At the time of the evaluation, this functionality was not yet implemented in the LBNS, so no notifications to police focal points were given.

With an implemented LBNS, the scenario in Table 4-1 would become reality for police officers on surveillance. In short, using a mobile location-based notification system, officers will be better informed and aware of relevant locations in their vicinity as well as less reliant on having to remember information at the right time and place. This will positively influence the efficiency and effectiveness of police work as well as the user experience of the service.

Table 4-1. Police surveillance scenario illustrating location-based notification.

Location-based notification during police surveillance
Youth officer Jason walks on surveillance, equipped with a PDA with the LBNS. When he is some distance away from the playground area in his district, his PDA beeps. He checks the display and is informed that he is near a hotspot for gang activity. Yesterday complaints were received about six adolescents causing nuisance around the playground area. The notification further informs him that a colleague checked this location yesterday and sent the perpetrators away with a warning. Prompted by this notification, Jason checks the playground again. He does not encounter any problems, but has a chat with the resident that filed the complaint. The resident is satisfied that the police checks up on complaints so quickly and Jason continues on his way. *On the next block, he receives a notification that the resident of number 47 has a outstanding fine for speeding. Based on this information, Jason confronts the resident, who is at home. He succeeds in making an appointment that the resident will come to the district bureau tomorrow to pay the fine. Jason inputs this appointment into his PDA and continues on his surveillance round.*

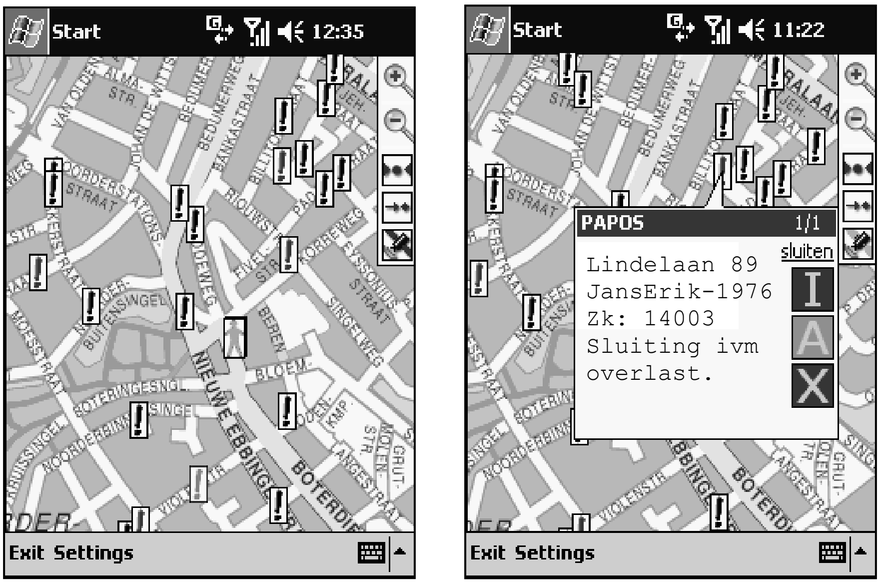


Figure 4-1. LBNS screenshots showing the POI overview (left) and a notification pop-up with a (fictive) warrant (right). Buttons indicate more information (I), take action (A) or close (X).

4.2.1 LBNS system description

Together with two commercial IT-companies, the Dutch police designed and implemented the LBNS on a PDA. The system was called Attentive Services (“Attendering Service” in Dutch). Five police officers with mobile computing experience were included in the design and implementation process as representative end-users.

The interface of the LBNS was designed as a geographical map application, showing relevant locations as icons (“points-of-interest” or POI’s) on this map (see Figure 4-1, left). Exclamation marks were used as icons, showing different colors for different information sources (e.g. blue for agreements on location). In addition, the user’s own location as well as the locations of colleagues were indicated on the map. Possible map manipulations included zooming, panning and showing or hiding the POI’s.

The LBNS was implemented on a Compaq IPAQ PDA. For an overview diagram of the system components, see Figure 4-2). The PDA was connected to the police database (P-info) which contained the agreements, warrants and their addresses and details. The connection was established by means of a secured (Virtual Private Network) tunnel over the UMTS data-network. In addition, the user carried a separate GPS device, connected to the PDA via Bluetooth. The LBNS coded the addresses to GPS coordinates and compared these coordinates to the current user location. When the user was within 40 meters of a POI location, the LBNS generated a notification.

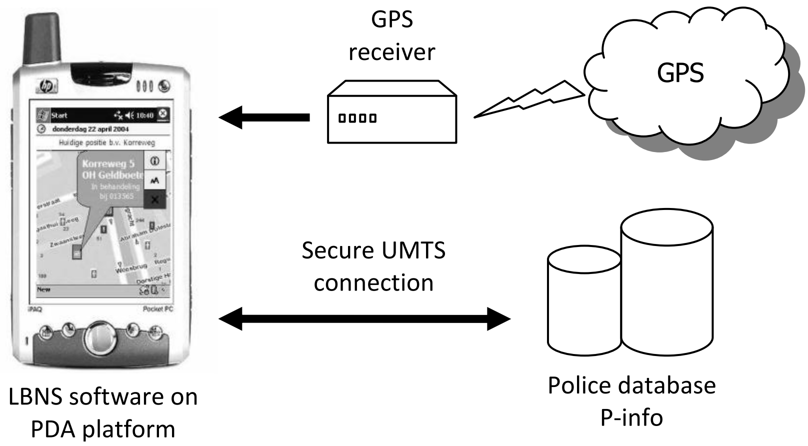


Figure 4-2. LBNS architecture and implementation.

A Kevlar protective casing was specially designed for this PDA. This casing allowed officers to attach the PDA directly to their belts, or let the PDA rest against their upper leg from a slightly longer attachment cable. This allowed them to view the display without having to detach the PDA from the belt.

4.2.2 Interaction with the system

The standard view of the LBNS was the geographical map shown in Figure 4-1 (left). Users could click on icons on the map, to see what the warrant or agreement was about. In addition, they could access the police database on the PDA using a web browser. The system presented notifications with a salient auditory signal and a small pop-up screen, pointing to the relevant location (see Figure 4-1, right). The pop-up screen showed the address, perpetrator and details of the location. By clicking on one of three buttons in the pop-up, the user could decide to close the notification (X), get more information (I) or directly take action on the notification (A). When clicking "more information", details in the police database about the warrant or agreement where shown in the browser of the device. When "taking action" (i.e. going to this location), officers could indicate the results of their actions (nobody home, suspect in custody, etc.) in the "action-screen". These notes could be read by colleagues. The officer could indicate whether the notification could be removed, because the incident was finished (e.g. fine collected). When the officer ignored the notification and did not click any button, the pop-up disappeared after five minutes or when a new notification came in. When multiple notifications appeared at the same address, these were presented in the same pop-up screen, indicating the total number of notifications. Users could switch between them using "previous" and "next" buttons.

4.3 Evaluation method

To test the efficiency, effectiveness and user experience of the LBNS for police officers, a longitudinal field evaluation was conducted over a period of four months.

4.3.1 Participants

In total, thirty police officers participated in this study. Of this group, twenty-four completed their participation till the end of the study and six participants dropped out. Twenty-two (17 male, 5 female) participants filled out the starting questionnaire. Their mean age was 36.6 years (SD = 7.6), with on average 12.2 years (SD = 7.9) of police experience. Within this group, four participants were designated "district officers", whereas the rest had no specific role. All participants had elaborate experience in using desktop and mobile computers. Only three participants had never before used a mobile computer.

4.3.2 Setting

The evaluation took place in the district "Korrewegwijk" in the city of Groningen, the Netherlands. This district has a surface area of approximately 1.8 sq. kilometers (1.12 sq. miles) and 16.700 inhabitants[1]. The police department in this district employs 50 uniformed police officers.

The PDA with the LBNS was handed out to the thirty participating police officers. All users received one full day of training with the device. They were instructed on how to use it in their daily work practice (see Figure 4-3). After the training, they were asked to use the PDA for the whole evaluation period (four months).

4.3.3 Measures

This field evaluation employed a set of objective and subjective measures to establish users' expectations of the system, the experienced system accuracy, efficiency and effectiveness and the user experience.

Expectations. Users' expectations regarding the LBNS were established prior to the evaluation in interviews. Questions considered expectations about information delivery, work process support and success factors of the system.

[1] Keyfigures 2008 available online at groningen.buurtmonitor.nl

Figure 4-3. Police officers on training (left) and using the LBNS on surveillance (right).

Experienced system accuracy. The LBNS should provide relevant information correctly and at the right time and place. For example, receiving a notification about an address two blocks away, or that has already passed does not constitute high system accuracy. Thus, both the accuracy of the moment of notification presentation as well as the accuracy of the information itself is assessed using questionnaires, interviews and participatory observation.

Efficiency and effectiveness. Efficiency was measured with objective methods (usage data) and subjective methods (interviews and questionnaires). It is expressed as the subjective user judgments about how well users were informed by the notifications, the relevance of the notifications and the influence of the system on the work processes. In addition, efficiency was regarded as the number of repetitions of notifications (how many identical notifications were given to each user). Effectiveness is measured as actual operational results based on the notifications (e.g. police officers undertook action or apprehended a person). Also, effectiveness was expressed as users' comments in the interviews and questionnaires.

User experience. The user experience of the LBNS consists of the user's appraisal of the ease with which they can work with the system and the usability of the interface. It was assessed whether police officers felt they could reach their goals with the system in a fast and efficient way. The number of interface operations, system errors or inaccuracies may have a negative impact on the user experience. In addition, system complexity, system speed and physical comfort were investigated.

4.3.4 Materials

This study employed three different online questionnaires, two semi-structured interviews and log files that captured system events. The starting questionnaire asked about participants' age and police experience, use of (mobile) computers, email and internet. The usage questionnaire consisted of fourteen statements concerning the daily use of the LBNS. The answering options ranged from -3 "strongly disagree" to +3 "strongly agree". The statements dealt with the experienced support of the LBNS in the two work processes (agreements and warrants), the system performance and distraction or interruption by the notifications. The acceptance questionnaire consisted of 25 statements about support, trust and usability of the system. The same answering options were used as in the usage questionnaire. In addition, four open questions regarding positive and negative aspects as well as possible improvements to the system. The semi-structured interviews lasted approximately thirty minutes and contained questions regarding current work processes and users' expectations of the system. The log files recorded all notifications, user actions and system events on the PDA with a timestamp. From these files, usage data on notifications and user actions were calculated.

4.3.5 Procedure

The evaluation took place from April until August 2007 (see Table 4-2 for a timeline of research activities). The evaluation period started with a training on the system and filling out the starting questionnaire. After the training, the first round of interviews was held. Police officers used the LBNS in their daily work activities for the rest of the evaluation period. At two moments in the evaluation, a researcher accompanied the police officers on surveillance for participatory observation. Every week, users were asked to fill out the usage questionnaire. In addition, at two moments during evaluation, the acceptance questionnaire was filled out. All users received an email, prompting them to fill out the questionnaires online. Finally, at the end of the evaluation period, another round of interviews was held. The response rates to the questionnaires and the number of participants in each research activity is specified in the final column of Table 4-2. Note that because of varying availability and response rate, the number of participants is not equal in each activity.

Table 4-2. Overview of the research activities and number of participants in the different phases of the field evaluation.

Phase	Research activities	Number of participants
Training (1 month)	Starting questionnaire	N = 22
Pre-test	1st Acceptance questionnaire Interviews	N = 9 N = 8
1st Evaluation (2 months)	Usage questionnaires Gathering log-files Observation	N = 26 N = 26 N = 2
Mid-test	2nd Acceptance questionnaire	N = 11
2nd Evaluation (2 months)	Usage questionnaires Observation	N = 14 N = 2
Post-test	Interviews	N = 5

4.4 Results

All data were analyzed and frequencies or percentages were calculated. For selected questionnaire items, statistical significance was calculated using t-tests for a single mean to see whether the answer deviated significantly from neutral.

4.4.1 Expectations

In interviews with eight participants, prior to the evaluation, their expectations about the system were established. All participants expect to be better informed due to the LBNS, which would result in higher efficiency and effectiveness of their work. Especially apprehending persons and collecting open warrants should improve. In addition, the awareness of agreements on location and changes made by colleagues is expected to increase by using the system. They expect the system to specifically aid the district police officers and the officers on "free shifts", but not the officers concerned with the emergency response shift. These officers generally do not have the time to respond to notifications. Having information about colleagues' location seemed helpful to them as well.

Half of the participants indicate they think system usability and stability to be critical success factors for them to accept the service. In addition, the speed of the system is important as well as the accuracy; notifications have to be given at the right moment and at the right location. Two officers are concerned about distraction due to too many interruptions from the system.

4.4.2 Usage data

From the logfiles, usage data were analyzed from 26 users over a 55-day period. Only usage data were included from users who had received ten or more notifications during this time. In total, users received 3647 notifications, regarding 239 unique incidents. Further analysis showed that only four locations were responsible for 2566 notifications. These locations were exceptional institutions (such as a forensic psychiatric institution) that housed persons with multiple open warrants, resulting in 10 to 25 notifications at the same time. To calculate the usage data more accurately, multiple notifications at these locations were counted as a single notification whenever they occurred. This filtering resulted in 1658 notifications over the whole period. 94% of the notifications concerned open warrants, while the remaining 6% dealt with agreements on location.

The log files specified system events as well as user reactions (e.g. ignore, close, request more information, took action on the notification, remove notification at the location). Usage data (in frequencies and percentages) is presented in Table 4-3. The usage data showed that to 66% of the notifications, no response was given. Because the notification pop-ups were presented directly in the display, it could not be determined whether users still noticed them, but did not respond, or that the notification went unnoticed altogether. 19% of the total amount of notifications was followed up by requesting additional information (clicking button *I*). Action was undertaken on only 6% of the notifications (clicking button *A*). Police officers often did not indicate the results of their

Table 4-3. Usage data over the first evaluation period.

			Total	%
Receive notification			1658	100.0
User reaction	Ignore notification (no interaction)		1101	66.4
	Close notification pop-up (*X*)		140	8.4
	Requesting more information (*I*)		313	18.9
	Take action (*A*)		105	6.3
	Action result	No interaction	32	1.9
		Nothing to report	51	3.1
		Finished, do not remove	16	1.0
		Finished, remove	6	0.3
Requesting agreement on location			138	8.3

action in the action screen (no interaction). In 3% of the notifications, there was nothing to report (e.g. the perpetrator was not at home). Twenty-two (16+6) specific instances were counted where the notification was handled with positive operational results; this represents 1.3% of the total number of notifications.

The average number of notifications was calculated per notification, per user and per 24 hour time period. On average, each unique notification was presented 5.7 times during the whole period. Each user received 69 notifications on average, with two users receiving more than 200 notifications over the whole period. Almost half of the participants received less than 50 notifications.

The log files showed approximately four active users per 24 hour period. During this period, on average 30.4 notifications were sent, although this number fluctuated heavily (SD = 21.3). Most of the notifications were sent during the day shift (63% between 9:00 and 15:00). During the afternoon, evening and night shifts, respectively 20%, 8% and 9% of all notifications were generated. Although less officers are on shift during evening and nights, these numbers still show a preference for using the system during the daytime.

4.4.3 Experienced system accuracy

In general, users commented that the icons are represented in the interface accurately at the right location. It could not be determined if the notifications were always delivered at the right location. The questionnaires contained two statements concerning this issue; "Notifications are presented on the right moment" and "Notifications are presented too far from the location". Considering the average neutral response to these statements and the high SD (see Table 4-4), users apparently sometimes agreed and sometimes disagreed with these statements. During observation, most of the notifications were presented right before or at the address. However, some notable exceptions were observed as well, where notifications were received blocks from the actual address, back at the station or were not presented at all. Presumably, these exceptions were due to technical errors.

The accuracy of the information in the notification is also important. Users disagree with the statement "Notifications often are not relevant", this effect was significantly different from neutral. However, in the questionnaires and interviews, users indicated that in general, too many notifications were delivered, they contained obsolete information, or they were repeated too often.

Table 4-4. Mean scores and standard deviation (SD) on selected questionnaire items.
Scale ranges from –3 (disagree) to +3 (agree). (* significantly different from neutral at *p* < 0.05).

Questionnaire item	Score	SD
Notifications were presented at the right moment	0.11	1.00
Notifications were presented too far away from the location	0.38	1.25
Notifications often contained obsolete information	0.54*	0.64
Notifications often are not relevant	- 0.63*	0.95
The system provided necessary information	0.19	0.73
Using the system, I can work faster	0.28	1.32
The system has a negative impact on my work	0.66*	1.21
Using the system, I can work more effectively	0.69*	1.18
The system improved the handling of warrants	0.62*	0.94
The system improved my awareness of agreements	0.44	0.94
The notifications were interruptive in my work	0.55*	1.21
The notifications often interrupted handling incidents	0.57*	1.20
The system presented too many notifications	0.49	0.83

4.4.4 Efficiency and effectiveness

In the final interviews, users indicated to feel positive about the concept of being notified of certain locations. They regard their work to be a little more effective due to the system, but not more efficient. They are better informed of open warrants and agreements on location. However, all users indicate that the LBNS is an addition to their work, not a guiding principle. For example, they ignore notifications that are not directly relevant.

The notifications provide information clearly and quickly. In the questionnaires, users related 14 specific instances where the LBNS provided a positive contribution to operational results (e.g. apprehension of persons). Also the opportunity to request additional information from police databases was considered very valuable. Finally, the LBNS showed the location of colleagues in the display. Users indicated that they did not use this information. This could be because of the relatively low number of users per day (4 on average) but also because police officers kept track of each others location via mobile phone or radio transceiver.

The downside of using the LBNS is that notifications can be interruptive or distracting during work. In the questionnaires, users indicated that notifications were often interruptive, especially during another incident or talking to people. As mentioned before, the system provided too many notifications (see Table 4-4). Users themselves indicated filtering of notifications (e.g. based on importance) as a solution to this problem.

4.4.5 User experience

Users mention that the LBNS is usable, understandable and clear, and that learning to work with the system is easy. However, there are three aspects that negatively influence the user experience of the system. First, the LBNS in its current form is a complex and vulnerable system, which relies heavily on the correct functioning of Bluetooth, WiFi and Virtual Private Network (VPN) connections. This lack of robustness results in occasional system malfunctions, while the complexity makes it hard for users to determine what went wrong and troubleshoot. Often, the only way to get the system up and running again is to reset the whole PDA. Second, the log-in procedure is very cumbersome, requiring three different login name and password combinations, including a code that changes every 30 seconds (provided by a physical token). Third, the system is relatively slow. Reloading or refreshing the interface can take up to a minute, while system reactions to user actions can be very slow. This results in frustration on the users' part, because the system is not available when they need it.

The interface of the LBNS is positively commented on. The icons are clear, the three buttons I (information), A (action) and X (close) are clear and concise. Some users indicate that the buttons and text are too small. There is certainly more space available in the notification pop-up screens to enlarge the text and buttons. The PDA itself is considered an extra burden to carry on the police belt. This belt is often completely filled with gun, pepperspray, gloves, handcuffs, etc. In the questionnaires, twelve users mention the PDA as physically "burdensome", and the separate GPS module as "unnecessary".

4.5 Discussion and recommendations

To see whether mobile information exchange can be supported by location-based notification, the Dutch police implemented a LBNS for mobile police officers. In this longitudinal field evaluation, thirty police officers used the LBNS during four months in their daily work practice. Using questionnaires, interviews, usage data and observation, we evaluated the system's accuracy, its effects on efficiency and effectiveness of work processes and the user experience.

In presenting the notifications, the accuracy of the LBNS was variable. Notifications were not always presented at the right location and the information is sometimes obsolete. The system accuracy might be improved by consequently presenting notifications at the right location. The system supports the work processes of open warrants and agreements on location; police officers are better informed and have necessary information about (prior) incidents available within the use context. This awareness led to a number of positive operational results such as the apprehension of perpetrators.

Users consider the interface of the LBNS easy and usable to work with and the screen design clear and concise, despite their perceived complexity of the system. This complexity might be mitigated by integration of different system components (such as the separate GPS device).

The downsides are that the current system can be slow, the log-in procedure is cumbersome and the system is lacking robustness. These downsides led six participants to cease further participation before the end of the evaluation period. In addition, users consider the notifications often distracting or interruptive. This can be concluded from the usage data and the answers to the questionnaires. Currently, only 19% of the notifications is followed up and 1% is acted upon and "solved". This might seem as a low number, but actual operational results are dependent on a number of factors. When a police officer wants to apprehend a person for an open warrant, as suggested by the notification, but no prison cells or police backup is available, it is considered not safe to proceed with action.

In its current implementation, the LBNS sends notifications to every police officer within the notification perimeter. Consequently, police officers often are already aware of the information in the notification, thus making it less relevant. In addition, not all notifications are relevant for every police officer. For example, district officers are very familiar with the agreements on location in their district. In contrast, police officers with the duty of emergency handling do not have the time to respond to open warrants. Finally, thought must be given on how to handle locations with 10 to 25 different notifications, such as special criminal institutions. These locations can be considered known to police officers, making it not necessary to notify them. These issues seriously affect the use of the system and must be addressed.

The current LBNS only notifies police officers of low-priority warrants, some of which are already known. The results mentioned above (e.g. the low "action rate") are influenced by this implementation choice. Briefing focal points were not yet implemented in the current LBNS. As this information is more dynamic then either warrants or agreements, it is expected to be very relevant to include in the system. In addition, these focal points are more limited in number then warrants and agreements, so this is not expected to lead to increased notification intrusiveness. In addition, should the notification system be extended to also notify of high-priority incidents in stressful situations, such as a car crash with several victims, the strengths and limitations of the system would become apparent.

This field evaluation provided some valuable lessons on methodology. The quality of the results was dependent on the varying availability and response rate from the police officers. It proved unrealistic to conduct both the first and last round of interviews with exactly the same participants. The response rate on the weekly questionnaires was especially low in the second evaluation period, around 50%. Participatory observation provided valuable insight into how the users used the system, but the officers were conscious that they were being observed. Finally, it proved difficult to quantify improvements in efficiency and effectiveness of the two work processes (agreements and warrants) that could be attributed to the system. Before the introduction of the notification system, efficiency and effectiveness were not recorded, making it hard to establish a baseline. One logical solution would be to measure on how many notifications action was taken and could subsequently be removed. However, agreements are necessary knowledge and do not need to be removed, making this not a valid performance criterion for agreements. Number of removed notifications is not a valid performance criterion for warrants as well, because it depends very much on the situation whether a warrant can be handled. The perpetrator can be not at home or there can be no room in the prison cells. In these cases, the LBNS and the user can have performed accurately, but still the action did not have the desired effect.

In its current implementation, the functional design of the LBNS is not appropriate for actual use in the police context. The downsides of the system (interruption, slowness, complexity and non-relevant notifications) currently outweigh the positive operational results. Possible solutions might be intelligent filtering or personalization of notifications and integration of different system components.

- *Intelligent filtering.* Filtering of notifications can lead to less interruption or distraction from other incidents. Notifications can be filtered based on incident priority, time of day or specific characteristics such as the height of the fine in the warrant. Furthermore, employing a personal user model will ensure that officers only receive notifications that are relevant for their task, role, availability or specific shift. Finally, the LBNS should keep track of who has received a specific notification, and not repeat this within a particular timeframe.

- *Integration of different system components.* Integration will lead to reduced system complexity, making it easier for the users to comprehend the system and troubleshoot when errors occur. For example, the separate GPS module can be integrated with the PDA, or the three different log-in procedures can be integrated into one.

Relevant for the design of a context-aware mobile support (CAMS) system, this field evaluation showed that a notification system used on surveillance increases officer awareness of operational information. However, the current system clearly has a number of downsides: presenting too many or non-relevant notifications and not presenting the notifications at the right time and place. These findings point out that the design of CAMS should address how and when notifications should be presented; i.e. the *notification appearance* and *timing*. From a human-computer interaction point of view, appearance and timing of notifications are expected to influence task performance, user distraction and notification intrusiveness. The next two chapters will present two controlled experiments that investigate the effects of different notification styles on surveillance task performance.

5 CONTEXT-AWARE NOTIFICATION STYLES: APPEARANCE

Abstract

To minimize unwanted interruption and information overload during surveillance, police officers need to be supported by a mobile, context-aware notification system. This system adapts notification style appearance to message priority and the use context (attending to the environment or to the device). A prototype is designed and evaluated in a simulated surveillance task, requiring users to attend to (high and low workload) videos while handling messages on a mobile device. Adaptive notification was compared to non-adaptive, uniform notification by measuring task performance (number of targets reported, messages recalled and message handling time), message intrusiveness, mental effort and subjective judgments. In high workload situations, adapting the notification appearance led to better performance and less intrusive messages than non-adaptive notification. Subjective judgments showed a positive user experience with the adaptive notification system. These empirical findings are used to improve the design of mobile notification support systems for police officers.

This chapter is based on the previously published material:

Streefkerk, J.W., van Esch-Bussemakers, M., & Neerincx, M. (2007). Context-aware notification for mobile police officers. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics, LNAI 4562* (pp. 436–445). Berlin: Springer-Verlag.

5.1 Introduction

The previous chapters demonstrated that police surveillance suffers from limited mobile information exchange and that officers' awareness of operational information should be improved. As a solution, context-aware mobile support (CAMS) systems proactively provide incident information to police officers in the relevant use context (such as notification at a specific location). The field study in Chapter 0 showed that such location-based notification improves officers' awareness of operational information. However, the study also emphasized that *how* notification systems present notifications determines how well users can work with them (performance) and whether these systems are preferred and used. Thus, the *notification style* is a major constraint for their acceptance and use (McCrickard et al., 2003a). Following our iterative design method for CAMS systems, this chapter presents the first experiment, investigating the effects of notification style appearance on surveillance task performance and user preference.

Designing CAMS should take into account divided attention during surveillance. An intrinsic part of police officers' task is dividing their attention between detecting incidents in their environment and attending to incoming messages on a mobile device. Officers on surveillance need to focus their attention on their surroundings 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. 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. Two potential risks to effective surveillance are unwanted distraction by low priority messages and information overload in high workload situations. These situations are caused by monitoring and integrating complex information from the environment, while simultaneously performing tasks on a mobile device (Nagata, 2003). Imagine a police officer on surveillance handling an argument between two neighbors. During this intervention, his mobile device interrupts him with a message. He has to attend to and read the message to perceive its urgency and content. Focusing his attention on his device distracts him from the argument, which can have serious consequences. When he can perceive a highly urgent message by its appearance, e.g. by a highly salient sound, unnecessary distraction is minimized. Similarly, when the message content is presented concisely instead of elaborately, information presentation is appropriately matched to the high workload use context. This way, he can still receive valued information, while focusing his attention appropriately on the environment or on the device.

Context-aware notification systems can provide this match by adapting information appearance (notification style) based on message priority and the use context (e.g. attending to the environment or to a mobile device). This is expected to help task switching in divided attention situations. The current chapter presents a controlled lab study of an adaptive notification prototype that presents messages in the appropriate notification style. The notification style consists of the message appearance, i.e. the salience of auditory and visual signals and the amount of information in the user interface. We state claims about the appropriateness of the notification styles (e.g. in a high workload situation, a normal priority message will be presented with a less salient auditory signal than a high priority message), and test these claims empirically. This study aims to answer: *What are the effects of notification style appearance, adapted to priority and use context, on task performance, notification intrusiveness and the user experience in a simulated surveillance setting?*

A prototype of this adaptive notification system is implemented on a mobile device and evaluated in a simulated surveillance task using a Wizard-of-Oz setup (see also section 3.3.2). Participants are required to perceive, read and remember messages on a mobile device, while attending to (high and low workload) videos or attending to the mobile device. We compare adaptive notification to a non-adaptive prototype (i.e. using a uniform notification style) so participants experience and can compare the dynamics of both systems. Because adaptive notification is expected to address general cognitive abilities instead of police task-specific abilities, we employed a non-professional participant group in this experiment, representative in age to police officers.

We expect that adaptive notification leads to better task performance during high workload situations, less intrusive messages and higher user preference compared to non-adaptive notification. Furthermore, we examine possible downsides of adaptive notification, such as increased mental effort, low recognition of notifications and difficulty in understanding the adaptive system behavior. The next sections describe the design and implementation of the context-aware notification prototype. Then, the evaluation method and results are presented and implications for the design of CAMS are discussed.

5.2 Design of notification styles

This section describes the notification styles, matches these to the context factors of message priority and use context and describes the implementation of the prototype.

5.2.1 Notification styles and rules

The design space of notification styles is determined by two important aspects of notification appearance: salience and information density (Kern et al., 2003). Because highly salient notifications are perceived more easily, message priority was matched to salience. Thus, high priority messages were presented highly salient (with visual flashing effects and loud auditory signals) and normal and low priority messages less salient. This was expected to limit interruption by lower priority notifications and at the same time facilitate recognition of message priority.

Information density is the amount of information conveyed by the notification, for example the amount of text in a display or the richness of an auditory signal. Use context was matched to information density, because the amount of information that has to be processed is an important factor in determining the speed of the response. For the present study, use context was categorized as attending to the environment or attending to the device. We further distinguished between attending to a low or high workload environment. In low workload environments, elaborate information was presented. In high workload environments, more condensed information was presented (e.g. by using keywords). When attending to the device, even more condensed information was presented (e.g. by using an icon). This was expected to benefit the speed of reaction to notifications. Based on these considerations, the adaptive behavior of our notification system was dictated by three rules:

1. *The higher the message priority, the more salient notification signals should be used.*
2. *The higher the workload from the environment, the more condensed information should be presented.*
3. *When attending to the device, even more condensed information should be presented.*

5.2.2 Implementation of the prototype

A prototype of this adaptive notification system was implemented on a mobile device (PDA) for evaluation. We used a Wizard-of-Oz setup by having the test leader send the messages at the proper time and in the appropriate notification style. High priority messages were presented with red visual flashing bars and icons and a sharp sound () (see Table 5-1).

Table 5-1. Matching notification styles to message priority and use context.

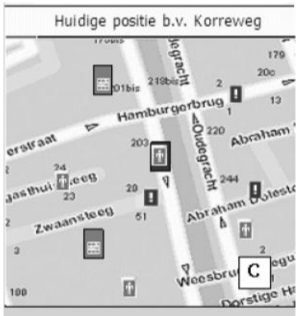
<table>
<tr><th rowspan="2"></th><th colspan="2">Attending to the environment</th><th rowspan="2">Attending to the device</th></tr>
<tr><th>Low Workload</th><th>High Workload</th></tr>
<tr><td>High priority</td>
<td>

Full message with red flashing bars and sharp sound</td>
<td>

Full message with red flashing bars and sharp sound</td>
<td>

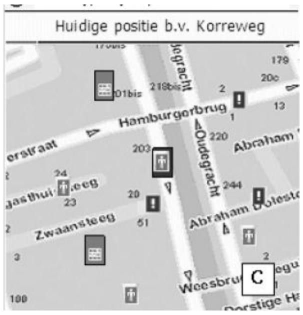
Red flashing icon and sharp sound</td></tr>
<tr><td>Normal priority</td>
<td>

☼
Full message and soft sound</td>
<td>

☼
Summary with soft sound</td>
<td>

☼
Icon with soft sound</td></tr>
<tr><td>Low priority</td>
<td>

–
Full message without sound</td>
<td>

–
Summary without sound</td>
<td>

–
Icon without sound</td></tr>
</table>

Normal priority messages were presented with a soft sound (☼) and low priority messages without sound (–). In low workload environments, the full message text was presented at once. In high workload situations, first a summary of the message was presented, except for the high priority messages. As we assumed that in this case a task switch was necessary, the full message was presented instead of a summary. To avoid interruption while users were attending to the device (e.g. reading a message), an icon in the interface signaled a new message (right column in Table 5-1). Summaries and icons could be clicked to open the full message. The non-adaptive prototype employed a uniform notification style (full message text with a sharp sound) for all messages, regardless of priority and use context.

5.3 Evaluation method

5.3.1 Participants

Twenty participants (10 male and 10 female), aged between 22 and 45 years (M = 29.1; SD = 6.8) completed the experiment. All used a personal computers and internet on a daily basis, 17 participants used a cellular phone daily and most of them (N = 19) had never before used a mobile device. None had working experience in the police domain.

5.3.2 Experimental design and manipulation

A within subjects, repeated measures design was used with an adaptive notification condition (notification style adapted to message priority and use context) and a non-adaptive condition (uniform notification style). In both conditions, similar surveillance scenarios were used with similar messages and videos. The order of conditions and scenarios was counterbalanced. All messages were relevant for the participants, however their required speed of response differed between high, normal or low priority messages. High priority messages, such as a colleague in need of backup, were considered urgent and needed immediate attention. Normal priority messages (e.g. about an outstanding warrant) were urgent, but did not require immediate attention. Low priority messages, such as questions from colleagues on past events, were not considered urgent and could wait.

Use context was operationalized as attending to the environment (watching a video) or attending to the device (taking notes or reading a message on the PDA). In two different “districts”, represented by two adjacent rooms, low or high workload was induced by showing videos. The low workload district showed videos of bicyclists riding through a shopping street, whereas high workload district showed videos of fights, arguments and attacks on police officers (see Figure 5-1). These videos were used in common police training and required focused attention from the participant to accurately remember perpetrator and situation characteristics. When participants physically moved from one district to the other, as prompted by the messages, this constituted a switch in the use context.

Figure 5-1. Video stills from the low (left) and high workload district (middle and right).

5.3.3 Surveillance task

The mobile surveillance task consisted of moving between both districts, watching the videos and reporting targets from the videos. In the low workload district, the number of bicyclists that rode into a shopping street had to be counted and noted on the PDA. In the high workload district, participants took notes on the PDA about the perpetrator and situation characteristics in the videos. In total 18 targets had to be identified from the videos. One scenario consisted of alternating three low workload videos and three high workload videos and took about 25 minutes to complete.

During the videos and note taking on the PDA, the system presented incident messages to the participants. These messages contained subject, priority, location and a description of the incident. A total of nine messages were sent to the participants; two messages were

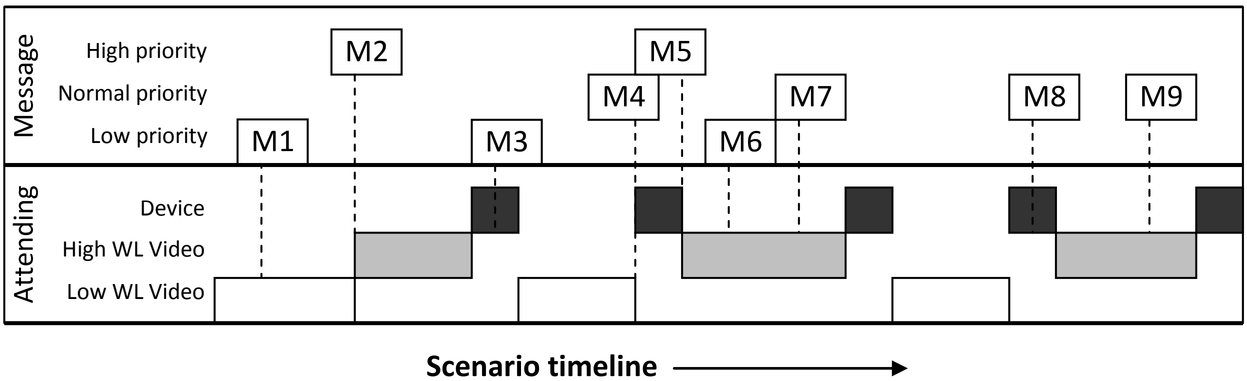


Figure 5-2. Scenario timeline indicating the presentation moment of the messages.

high priority, four normal and three low priority (see Figure 5-2). High priority messages had to be read as fast as possible and prompted participants to move from the low workload to the high workload district to watch the next video there. Normal and low priority messages could be read when possible. To read the whole message, scrolling was required.

5.3.4 Measures

In this experiment, we measured effectiveness and efficiency of task performance, mental effort, message intrusiveness and preferences. During each scenario, surveillance task effectiveness was measured as the number of *targets* reported from both the low and high workload videos. In addition, the number of *messages* recalled correctly was measured with six multiple choice questions after the scenario. Efficiency was measured as the *message handling time* in seconds. After each scenario, *mental effort* was measured with the Rating Scale Mental Effort (Zijlstra, 1993).

After each message, participants rated the *message intrusiveness* using a small rating scale from 1 (not interruptive) to 7 (highly interruptive) on the PDA. After each scenario, *subjective judgments* are measured with a 5-point rating scale on the ease of the surveillance task, difficulty in directing attention, interruption and irritation experienced by the messages. In addition, recognition of adaptation was measured with open questions after the adaptive condition asking participants which differences they noticed between the messages and districts. Finally, after the experiment participants stated *preferences* by comparing both conditions and offered improvements to the prototype.

5.3.5 Apparatus

The adaptive notification system was simulated using a HP 5550 PDA. A Wireless LAN connected the PDA to a message server, used by the test leader to send the messages. The server kept an event log of all events on the PDA (messages received, messages opened, etc.). Videos were projected full-screen on the wall of the two rooms using two beamers controlled by a desktop PC. The rooms, PDA screen and the video images were videotaped as a back-up procedure.

5.3.6 Procedure

At the start of the experiment, participants were instructed they had to perform two surveillance scenarios by watching videos and handling messages on a mobile device. In one condition, the notification style of the messages on the PDA would be adapted to the message priority and their use context, explained as either “relaxed” (district 1, low workload), “busy” (district 2, high workload) or “working with the device”. In the other condition, all messages would be presented as full messages. Participants could decide for themselves when to read a message. After training with the PDA, messages and notification styles, participants performed both conditions with a short break in between. The timeline of the scenario is presented in Figure 5-2. After each condition, they filled out the multiple choice questionnaire, recognition questionnaire and subjective rating scales. After both conditions, they filled out the preference questionnaire, were debriefed and paid for their participation.

5.3.7 Statistical analyses

Message handling time and message intrusiveness were analyzed with repeated measures ANOVA with condition (adaptive or non-adaptive), use context (low workload, high workload or device) and message priority (high, normal or low) as within-subjects factors. Bonferroni post-hoc comparisons were performed on these two variables. Number of targets, number of messages and mental effort data were analyzed with dependent samples *t*-tests. Finally, the questionnaires and rating scales were analyzed with non-parametric tests (Wilcoxon Matched Pairs test). In the adaptive condition, message M6 (low priority, high workload) was only perceived by four participants and therefore excluded from the analysis.

5.4 Results

5.4.1 Surveillance task effectiveness

Effectiveness was measured as the number of targets reported from the videos and number of messages recalled correctly. The number of targets reported in the low workload use context did not differ significantly between adaptive (A) and non-adaptive (NA) conditions (M_A = 12.4; M_{NA} = 11.4; $t(19) = 1.04$; $p = 0.31$). However, in the high workload use context, significantly more targets were reported with the adaptive than with the non-adaptive system (M_A = 13.1; M_{NA} = 11.6; $t(19) = 2.15$; $p < 0.05$). There was no significant difference in number of messages recalled correctly (M_A = 3.3; M_{NA} = 3.2; $t(19) = 0.21$; $p = 0.84$). Thus, effectiveness of the surveillance task, in terms of targets reported, profited from the adaptive system in high workload use contexts.

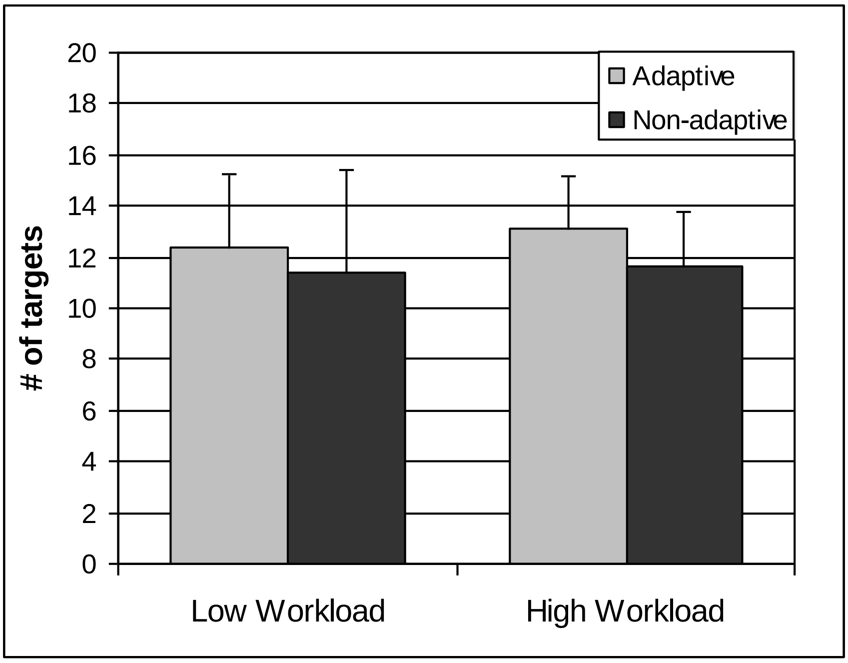


Figure 5-2. Mean number of targets reported in low and high workload videos.

5.4.2 Surveillance task efficiency

Efficiency was measured as the message handling time (MHT). MHT was significantly longer in the adaptive than in the non-adaptive condition ($M_A = 42.6$; $M_{NA} = 26.0$; $F(1,19) = 35.1$; $p = 0.000$). MHT was also significantly different between use contexts ($F(2,38) = 18.4$; $p = 0.000$; see Figure 5-3, top). Post-hoc comparison showed that MHT was longer for messages in the high workload use context (51 sec.) than in the low workload use context or when attending to the device (both 25 sec.). A significant interaction effect ($F(2,38) = 5.37$; $p < 0.01$) showed that MHT varied across both conditions and contexts. Post-hoc comparison showed significantly longer MHT for the adaptive condition compared to the non-adaptive condition in the high workload use context (p =.000), but not in the low workload use context ($p > 0.05$). Thus, message handling time showed that with the adaptive system, participants postponed handling (lower priority) messages in the high workload use context.

We did a follow-up analysis of MHT per priority level. Message handling time was also significantly different for different message priorities ($M_{low} = 32.5$; $M_{norm} = 41.8$; $M_{high} = 17.3$; $F(2,38) = 29.3$; $p = 0.000$; see Figure 5-3, bottom). Post-hoc comparisons showed significant longer MHT for low and normal priority messages in the adaptive condition compared to the non-adaptive condition (both $p = 0.000$). This confirms that participants recognized the priority level and postponed handling lower priority messages.

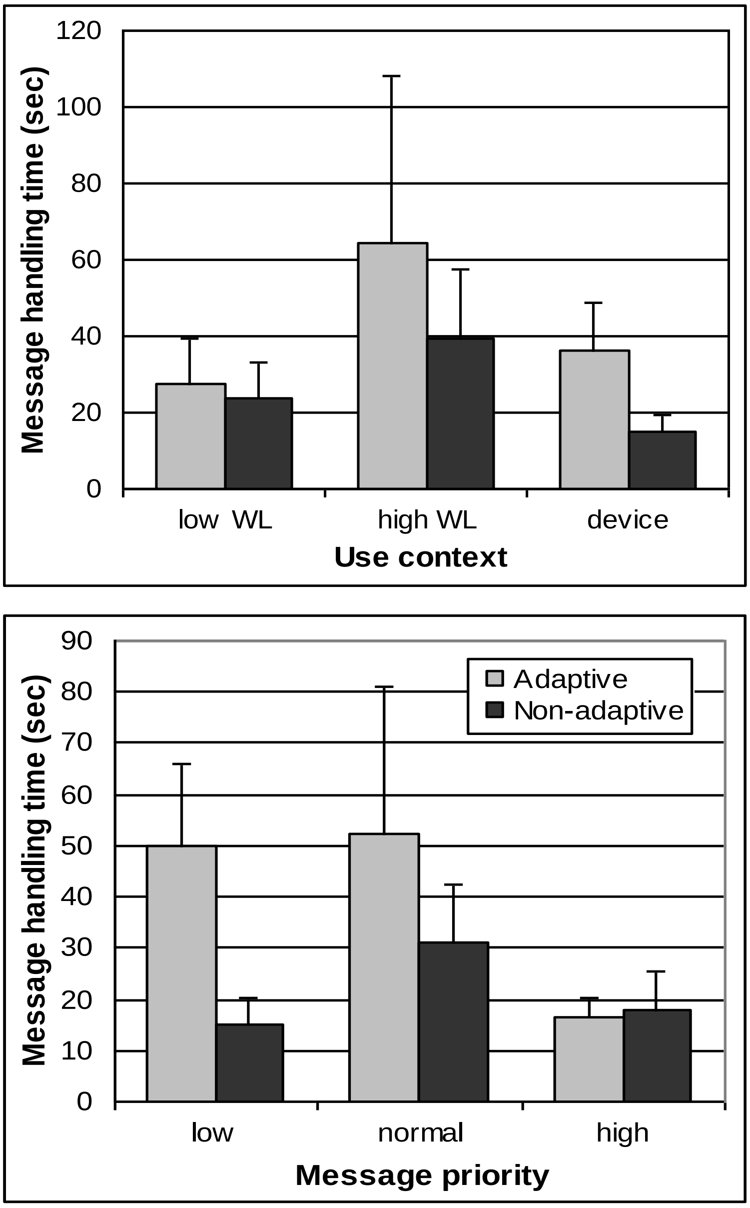


Figure 5-3. Message handling time for use context (top) and message priority (bottom).

5.4.3 Message intrusiveness and mental effort

From data on the message intrusiveness scale, it shows that messages in the adaptive condition were experienced as less intrusive compared to the non-adaptive condition ($F(1,19) = 17.6$; $p = 0.000$; see Figure 5-4). In addition, post-hoc comparison showed that messages in the adaptive condition were experienced as less intrusive in the high workload use context ($p = 0.000$) and when attending to the device ($p = 0.000$). Mental effort was not significantly different between the adaptive and non-adaptive conditions but showed a trend towards lower mental effort in the adaptive condition ($M_A = 46.7$; $M_{NA} = 50.9$; $t(19) = 1.93$; $p = 0.06$). Thus, with the adaptive system, the intrusiveness of messages is lower.

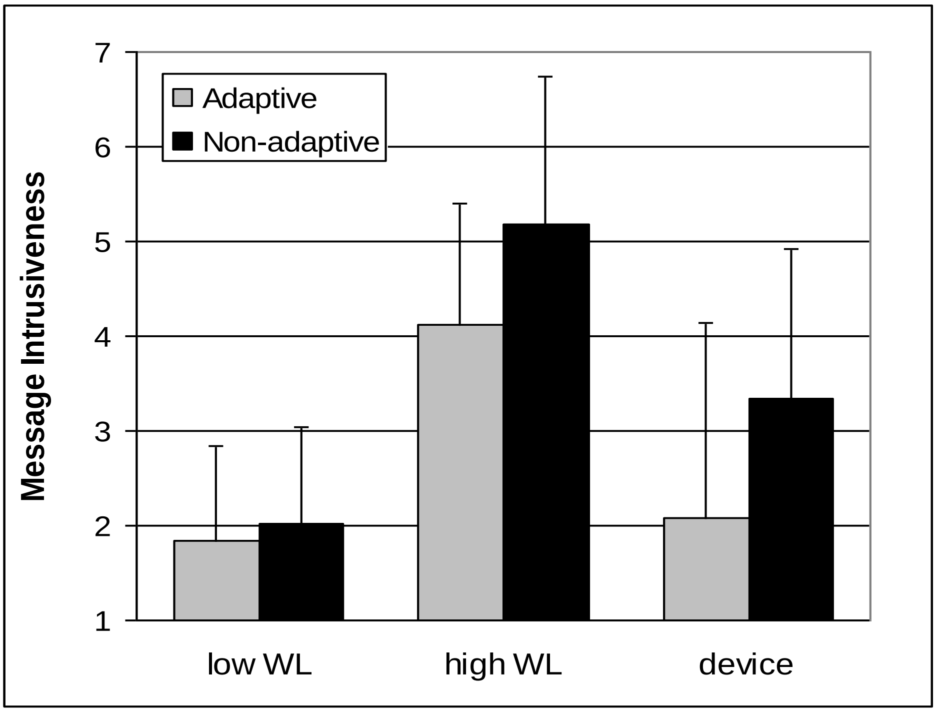


Figure 5-4. Message intrusiveness scores per use context.

5.4.4 Recognition of adaptation

In the open questions, in total 25 correct remarks were made about the differences in message presentation between the adaptive and non-adaptive condition. Of these, 20% stated all differences, 24% were only about differences in priority, 24% only about visual aspects and 20% only about the auditory signals. When asked about the differences in message presentation between the districts, 22 remarks were made. Of these, only 23% identified the correct differences (low workload district: whole message, high workload district: summary). 36% of the answers were wrong or did not state any differences. The rest (41%) of the remarks did not concern message appearance.

5.4.5 Subjective judgments and preferences

Participants did not find the surveillance easier in either the adaptive or non-adaptive condition ($Z(20) = 1.48$; $p > 0.05$). However, in the adaptive condition they found it significantly easier to direct their attention between the messages and the videos ($Z(20) = 2.56$; $p < 0.05$), were less interrupted and less irritated by the messages ($Z(20) = 2.28$; $p < 0.05$ and $Z(20) = 2.82$; $p < 0.01$ respectively). Their preference was almost unanimously for the adaptive condition (90%) because they could distinguish message priority better (50%)

and the messages were less interruptive (25%). These results show high preference and positive judgments for the adaptive system.

5.5 Discussion

This study evaluated adaptive notification compared to non-adaptive notification in a simulated surveillance task. Notification style appearance was adapted to message priority and use context (attending to a low or high workload environment or to the device). Results showed that with adaptive notification, slightly more targets were reported from high workload environments than with non-adaptive notification. Message handling time was overall longer with adaptive than with non-adaptive notification, especially in the high workload environment and for low and normal priority messages. So, while task performance was slower, this shows that participants appropriately postponed handling low and normal priority messages when in a high workload environment. Handling high priority messages was equally fast with both adaptive and non-adaptive notification. Furthermore, with adaptive notification the message intrusiveness was rated lower than with non-adaptive notification. We did not find effects of adaptive notification on recall of messages or on mental effort. User preference was almost unanimously in favor of the adaptive notification system. When using the adaptive system, users could direct their attention better between the environment (videos) and the device (messages). These empirical findings support the anticipated usefulness of context-aware notification systems and positive subjective judgments found in earlier research in other domains (Jiang et al., 2004a).

Overall, the adaptive notification system allowed users to perceive, comprehend and act on messages, suggesting the notification system was appropriately designed for this task. The finding that more targets were reported with adaptive notification implies that adaptive notification distracted participants less from the environment than non-adaptive notification. The variance in message handling time (short for high priority, longer for low and normal priority) showed participants could recognize message priority correctly from the notification style and respond appropriately based on this priority. The correct answers on the recognition questionnaire support this conclusion. Participants did remark that low priority messages should also be announced by an auditory signal. In addition, the current prototype uses the visual modality heavily, keeping visual attention away from the environment. Instead, the auditory modality could be used to provide more information-rich signals, for example with synthesized speech.

Many participants did not notice and recognize the differences in message adaptation between use contexts. It was not clear to participants that the appearance of a summary or icon depended on their location or their attending to the device. However, participants found the summary before a message useful and task performance was better than with full-text messages only. This shows that the matching between use context and information density was intuitive and appropriate. The adaptive system behavior supported the users' task flow, even though users may not actively notice it. Thus, the adaptation of information density is useful for notification systems.

The Wizard-of-Oz setup allowed varying the message priority and use context between conditions and controlled measurements without the need to implement a fully functional system. This setup was successful in creating the simulation of a functioning adaptive system. In addition, using videos and messages successfully created divided attention situations, characteristic of police surveillance. One of the limitations of the current setup is that the workload induced from the videos was not validated. Moreover, determining whether an environment actually induces high or low workload is hard to accomplish in real life situations. Fluctuations in workload occur often, especially in the police domain. Another limitation is that participants in this experiment were not required to actively engage in the surveillance situations. They could only make a limited set of decisions in a task they did not directly control. The mobility of participants within the setup was limited and did not require navigation. A more elaborate and immersive mobile task environment is needed, such as a real life task setting or game-based environment, to also take into account the task flow, navigation, situation awareness and the dynamics of the user experience.

In conclusion, this study provided empirical data on the positive and negative impact of a context-aware notification system on task performance and message intrusiveness in a task-relevant setting. When using this system, task performance in terms of targets reported was slightly better in high workload situations and users felt less interrupted than with a non-adaptive system. While recognition of adaptive behavior was low, user preference was high for the adaptive system. We demonstrated that using this adaptive system incurred a cost of postponing normal and low priority messages, resulting in longer message handling time. This shows that *when* notifications are presented within the task is important to consider in the design of CAMS. As argued in Chapter 0, appropriate timing of notifications is expected to support task switching and to limit interruption and distraction. Timing notifications within the task flow should be based on what the user is doing (user activity) and how important the message is, relative to the activity (relative priority). To test these assumptions, the notification styles from the present experiment need to be redesigned to include both notification timing and appearance. In addition, the experimental paradigm requires an ongoing task flow in a mobile task environment, to systematically manipulate user activity and priority. The next chapter presents the second experiment on context-aware notification styles that employs a redesigned notification prototype. In a mobile surveillance task, we evaluate how both timing and appearance of notification styles influence surveillance task performance.

6 CONTEXT-AWARE NOTIFICATION STYLES: TIMING

Abstract

To prevent unwanted interruption in police surveillance, the context-aware notification system from the previous study is redesigned to adapt the timing and appearance (notification style) of incident messages. These notification styles are adapted to user activity (available, in transit or busy) and relative message priority (higher, equal or lower than current activity). This chapter presents the design and user evaluation of a prototype mobile notification system. In a mobile surveillance task with thirty-two trained student participants, adaptive notification was compared to three uniform notification styles (presenting full messages, postponing messages or presenting indicators). The effects of notification styles on surveillance task effectiveness (decision errors and targets), efficiency (response time and handling time) and message intrusiveness were measured. Full messages caused short response time to messages, but also high message intrusiveness. Postponing all messages resulted in decision errors on incident handling. Indicators decreased awareness of the environment (low number of targets) and were forgotten or overlooked. Adaptive notification resulted in a low decision error rate at the cost of increased response time. We conclude that adaptive notification maintains awareness of messages without decreasing awareness of the environment. Implementation of adaptive notification systems for mobile police officers can help prevent unwanted interruption in police surveillance.

This chapter is based on the following material:

Streefkerk, J.W., McCrickard, D.S., van Esch-Bussemakers, M.P. & Neerincx, M.A. (submitted). Do (not) disturb?: Using context-aware notification to improve effectiveness of a mobile patrol task. *Submitted to Int. J. of Mobile Human-Computer Interaction.*

6.1 Introduction

To support mobile information exchange, notification appearance on a mobile device was adapted to message priority and the use context. The previous chapter showed that this context-aware notification can (slightly) improve task performance in terms of targets reported from the environment and a reduction in notification intrusiveness. However, unwanted interruptions still occurred due to inappropriate timing of notifications within the task flow. In addition, the task did not require active engagement or navigation from the participants. This chapter addresses these issues by focusing on how to time notifications appropriately within an on-going mobile surveillance task. The notification system prototype is redesigned in two ways; first by taking into account user activity instead of workload from the environment and second by taking into account the priority of the message *relative* to the current incident. In this chapter, our second experiment investigates what the effects are of context-aware notification styles (timing and appearance) on surveillance task performance. This experiment aims to validate and refine the results from the previous study.

As stated in the previous chapter, police officers on surveillance must divide their attention to ensure awareness of their direct surroundings and of incidents elsewhere. Notification systems maintain officers' awareness of incident messages, but can diminish awareness of the environment due to unwanted interruption. This illustrates the cost-benefit trade-off that exists between awareness and interruption. Awareness of incident messages 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 (McCrickard et al., 2003b; Horvitz, Apacible, & Subramani, 2005). As argued in section 2.5, 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) (McCrickard et al., 2003b; Bailey et al., 2006; Streefkerk, van Esch-Bussemakers, & Neerincx, 2006). Previous research has shown that postponing, scheduling or deferring interruptions until appropriate moments mitigates the negative effects of these interruptions (Adamczyk et al., 2004; Iqbal et al., 2008). Also, the presentation modality (e.g. visually, auditorially) and salience of the message influences its intrusiveness (Nagata, 2003; Kern et al., 2003; Streefkerk, van Esch-Bussemakers, & Neerincx, 2007).

Our redesigned context-aware notification system takes into account officers' 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. Based on these two context factors and a set of rules, this system adapts the notification style; i.e. the timing (e.g. postponing a message) and appearance (e.g. using an indicator icon) of an incident message. The current study focuses on 1) designing appropriate notification styles and 2) evaluating the effects of adaptive notification (compared to uniform notification styles) on task performance and the user experience. This study aims to answer: *What are the effects of a context-aware support system, which adapts notification timing and appearance to message priority and user activity, on task performance and the user experience?*

We operationalized core task features of police surveillance relevant to the awareness trade-off (observation, navigation, notification, incident handling) in a controlled experiment with trained student participants. A mobile surveillance task required participants to notice targets and handle incidents while their notification system presented incident messages in different notification styles. We expect that adaptive notification will improve the effectiveness (e.g. decreasing decision errors) and efficiency (e.g. improving response time) of responding to incident messages. Furthermore, adaptive notification is expected to prevent unwanted interruption of incident handling, leading to a positive user experience.

The remainder of this chapter elaborates the (re-)design of the notification styles from Chapter 5 to include notification timing based on user activity and relative priority. We describe when to use which notification style, summarized in a notification matrix. The redesigned notification system prototype is implemented and evaluated in a mobile surveillance task. The method and results of the evaluation study are presented and implications are discussed for notification systems' design for mobile professionals.

6.2 Design of notification styles

In this section, the design space for notifications styles is extended to include timing of notifications (McCrickard et al., 2003a). We argue that relative message priority and user activity are two relevant context factors that determine notification situations. Based on these two factors, notification rules dictate which notification style is appropriate. Finally, we describe how the notification rules and styles are implemented in an experimental prototype.

6.2.1 Priority and activity in mobile police surveillance

In police surveillance, normally the officer nearest to an incident location will be the most appropriate person to respond to this incident (Streefkerk et al., 2006; Streefkerk, van Esch-Bussemakers, & Neerincx, 2008). However, location (i.e. proximity) alone is not sufficient to predict whether an incident message is appropriate. As argued in the introduction, two other predictive factors are relative priority and officer activity. For an officer engaged in a high priority incident, 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. This illustrates *relative priority*, i.e. the priority of an incoming message compared to the priority of an incident. Relative priority can be determined from standard incident categorization in the police domain and is classified as *higher*, *equal* or *lower*.

Postponing all lower priority messages might mitigate the problem of unwanted interruption, but diminishes officers' awareness of incident messages that are relevant. For example, busy officers still need to receive a high priority message about a colleague in danger. Another example is that officers in transit to an incident need to be aware of any equal or higher priority messages to decide if a switch to another incident is necessary. These examples illustrate that officer *activity* should also be taken into account. 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, officer activity can be classified as *available* (not engaged in incident handling), *in transit* (en route to an incident) or *handling an incident*. Based on these classifications, we can distinguish nine notification situations (see also Table 6-2). The next section specifies appropriate notification styles (timing and appearance) for each of these situations, based on notification rules

6.2.2 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 6-1). Auditory signals are coupled to timing; sounds are used for directly presented notifications, whereas no sound is used for postponed messages. As in the previous chapter, the salience of the sound conveys the priority of the message (see also section 5.2.1). Notification timing is either direct (when a 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 (see Table 6-1) is not considered a useful notification style.

We can now specify five notification rules for our 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 6-2. 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, facilitating 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.

Table 6-1. Notification design space.

Timing	**Visual appearance** *Full message*	*Indicator*
Direct	F Presenting full message directly, with sound	I Presenting indicator directly, with sound
Postpone	P Postponing full message, without sound	*N/A*

Table 6-2. Matrix matching notification styles to priority and activity.

Relative priority	**Officer activity** *Available*	*In transit*	*Handling incident*
Higher	F	F	I
Equal	F	I	P
Lower	F	P	P

6.2.3 Implementation of prototype

An experimental prototype of this context-aware notification system was implemented on a PDA (Personal Digital Assistant) handheld computer. Based on the notification matrix in Table 6-2, the prototype system presented notification messages in different styles. Full messages (see Figure 6-1, left) were shown as text messages in the interface. Users could "Accept" or "Decline" a message with two buttons below the message text. Indicators (see Figure 6-1, middle) 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 check off messages in the message list (see Figure 6-1, right).

In this experimental prototype, the context-awareness of the system was simulated by having the test leader send the notification messages. Relative priority was determined by comparing the priority of a new incident to the priority of the current incident or activity. User activity was determined by the following user actions: accepting a message, arriving at the scene and finishing an incident. Based on these actions ("accepted", "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

Figure 6-1. PDA screenshots showing a full message (left), an indicator (!) in the lower right corner (middle) and the message list with two incident messages (right).

domain and that there are no technical constraints to fully implement this functionality.
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 (postpone, indicator or full message) based on a set of notification rules.

6.3 Evaluation method

To assess the effects of adaptive notification on task effectiveness and efficiency and the user experience, the prototype notification system was evaluated in a mobile surveillance task. This setting recreated core task features of police surveillance (e.g. divided attention situations, notification, navigation and incident handling). Trained student participants walked a predetermined route through a university office building while looking for targets (cf. awareness of the environment). They carried the prototype notification system, which presented messages on current incidents (cf. awareness of incident messages). When a message was presented, participants suspended the surveillance, 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.

To capture the trade-off between awareness of the environment and of incident messages, 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 6-2 to determine timing and appearance of notification presentation. We specified claims on 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 intrusiveness, workload, user preference) in Table 6-3. These claims lead to the following hypotheses on task performance and user experience, which specify the awareness trade-off for the notification styles:

1. *Full messages* will maintain awareness of incident messages, resulting in a short response time. However, using full messages will cause users sometimes to inappropriately attend to the messages, resulting in decision errors. Furthermore, this

Table 6-3. Claims for the effects of the notification styles on awareness of the environment and awareness of incident messages.

Notification styles	Awareness of environment			Awareness of incident messages	
	Number of targets	*Message intrusiveness*	*Incident handling time*	*Decision errors*	*Response time*
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

will also decrease awareness of the environment, causing a low number of targets noticed, high intrusiveness of messages and long handling time.

2. *Postponing* all messages will maintain awareness of the environment (a high number of targets noticed, low message intrusiveness 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 intrusiveness and intermediate handling time.
4. The *adaptive* notification style will balance awareness of the environment (high number of targets noticed, low message intrusiveness 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 user workload and preference differently. For example, maintaining awareness of both the environment and incident messages may come at the cost of increases in workload.

6.3.1 Participants

Thirty-two undergraduate and graduate Computer Science students (24 male, 8 female), aged between 19 and 32 years (M = 22.8, SD = 2.8), completed this study. 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.

6.3.2 Surveillance task

The surveillance 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 find 14 *targets*, consisting of 4-inch yellow paper disks, placed on the walls at various locations throughout the building (see Figure 6-2). 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 Figure 6-3). 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 surveillance. 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 three stages:

Figure 6-2. Surveillance task targets: yellow paper disks on the wall (arrow added).

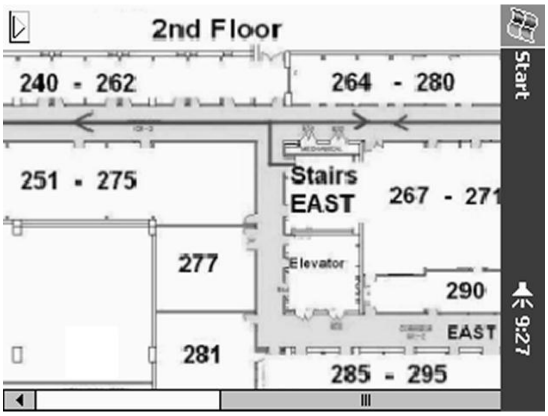


Figure 6-3. Floor overview on the PDA (rotated 90 degrees). The gray area represents a hallway, the dark gray line indicates the route.

- Reading the incident message and deciding to "Accept" or "Ignore" the incident.
- Moving to the incident location (in transit).
- Handling the incident by listening to an auditory narration of incident details.

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 Figure 6-1). After reading and accepting a message, participants moved to the incident location, indicated by a small sign on the wall. The "incident handling" consisted of playing a pre-recorded auditory narration on the PDA, relating the details of the incident. Participants memorized the details and checked the incident off, which returned their activity status to "available". They then navigated back and continued on the surveillance route where they left off.

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 6-4). Participants finished one 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

Table 6-4. Presentation order of messages (M1 to M12; priority between parentheses) during the surveillance task.

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

participants would be to suspend 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.

6.3.3 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 6-1). 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 6-2. The same set of messages and incidents was used in all conditions. Each participant participated in one experimental condition (6 male and 2 female participants per condition). A between-subjects design was employed, because the surveillance route could only be followed once without knowing the route and location of the targets. The presentation order of the route and message sets was reversed for half of the participants to avoid order effects.

6.3.4 Measures

In this experiment, individual characteristics, performance measures on the surveillance task and subjective measures were collected (see Table 6-5).

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 & Ferrano, 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* (Neerincx et al., 1999).

During the experiment effectiveness of the surveillance 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* in seconds to the second, interrupting message, timed from notification to accepting or declining the message. *Incident handling time* in seconds was calculated by subtracting the time spent on navigation from the total time to compensate for walking speed. After every message, participants rated *message intrusiveness* 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

Table 6-5. Measures and variables in the experiment.

Phase	Measure	Variable
Before	Individual characteristics	Age, Gender, Computer experience, Task switching ability, Memory score
During	Effectiveness	Number of read errors and handling errors, Number of targets noticed, Number of incident details recalled
	Efficiency	Response time, Incident handling time
	Subjective judgments	Message intrusiveness
After	Subjective judgments	Workload, System disruption, System supportiveness, Awareness of messages, Satisfaction

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.

6.3.5 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 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.

6.3.6 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 surveillance task through the building, while using a prototype notification system. They then signed an informed consent form and the first questionnaire and tests were administered. Participants familiarized themselves with the floor plans on the PDA and followed the surveillance 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 surveillance task as quickly and accurately as possible, accompanied by the test leader. Hereafter, they filled out the NASA TLX and the questionnaires.

6.3.7 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 intrusiveness 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.

6.4 Results

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

6.4.1 Surveillance task effectiveness

Effectiveness of the surveillance 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 Figure 6-4). 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 Figure 6-6). 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 Figure 6-6). 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.

Table 6-6. Means for task performance variables and message intrusiveness (MI) per condition.

Condition	Surveillance task effectiveness and efficiency						MI
	Read errors (#)	*Handling errors (#)*	*Targets (#)*	*Details (#)*	*Response time (s)*	*Incident handling time (s)*	
Full message (F)	1.5	0.3	8.6	16	10.2	178	4.6
Postpone (P)	3.1	2.0	10.8	20	12.6	172	3.1
Indicator (I)	1.4	0.4	6.8	16	15.2	176	3.7
Adaptive (A)	0.5	0.1	8.5	21	17.2	181	3.5

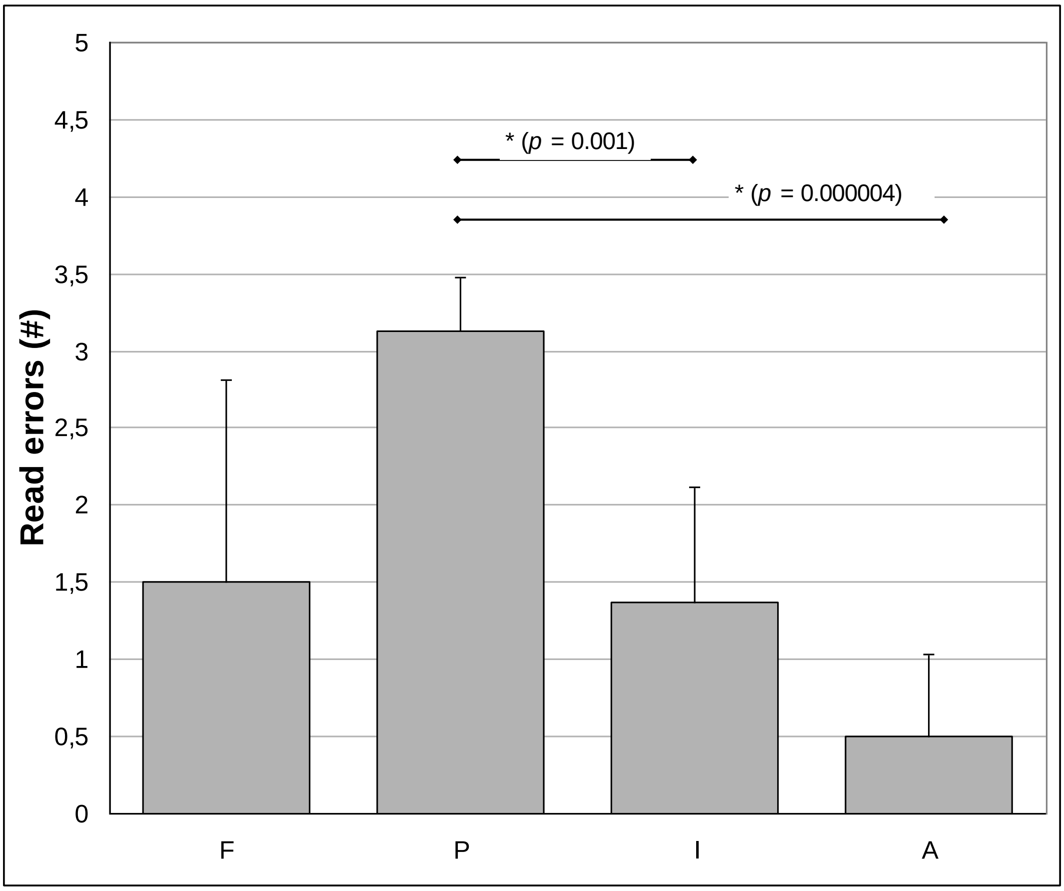


Figure 6-4. Mean number of read errors per condition.

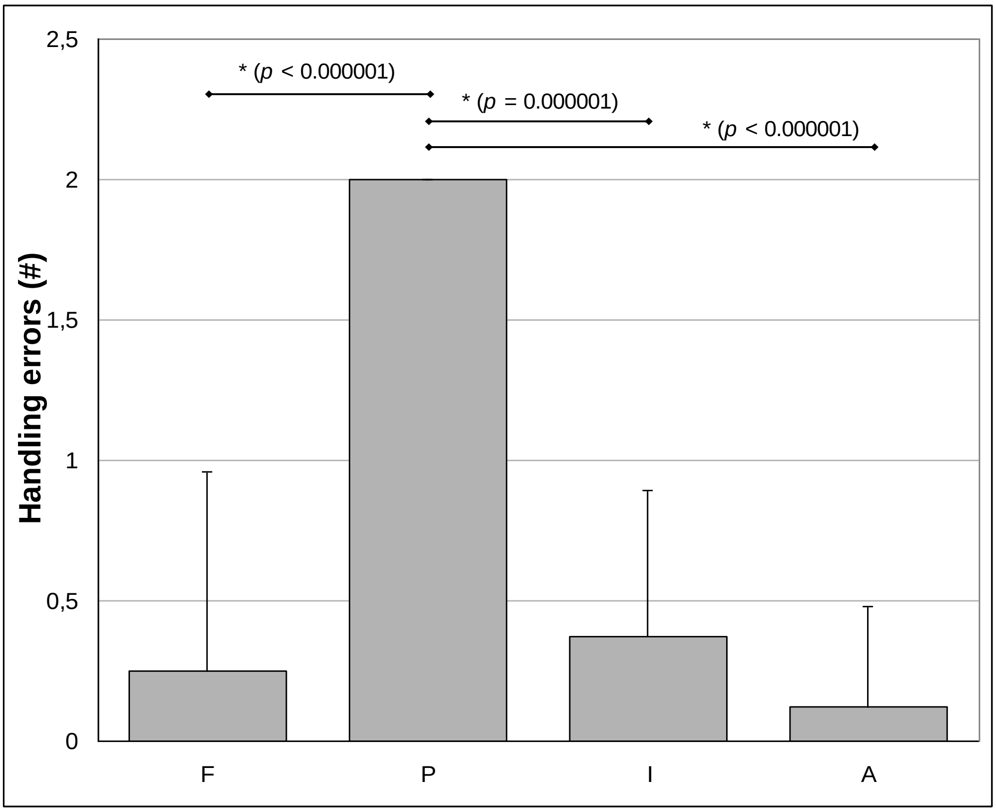


Figure 6-6. Mean number of handling errors per condition.

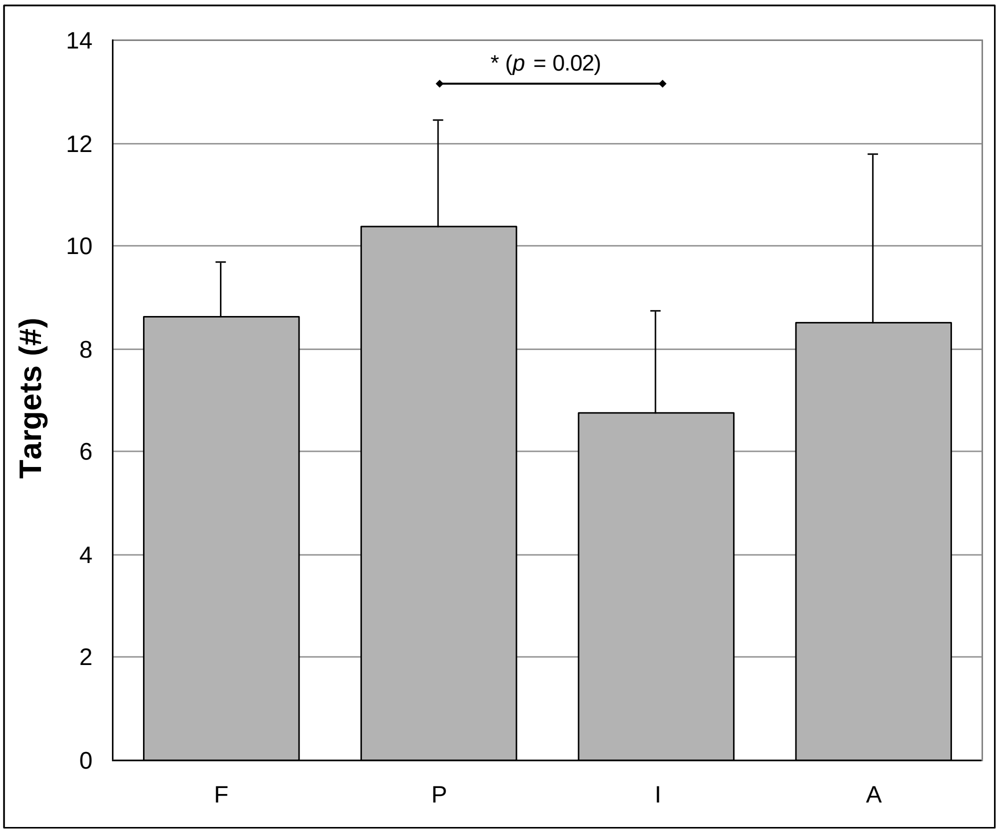


Figure 6-6. Mean number of targets noticed per condition.

6.4.2 Surveillance 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 Figure 6-7). 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

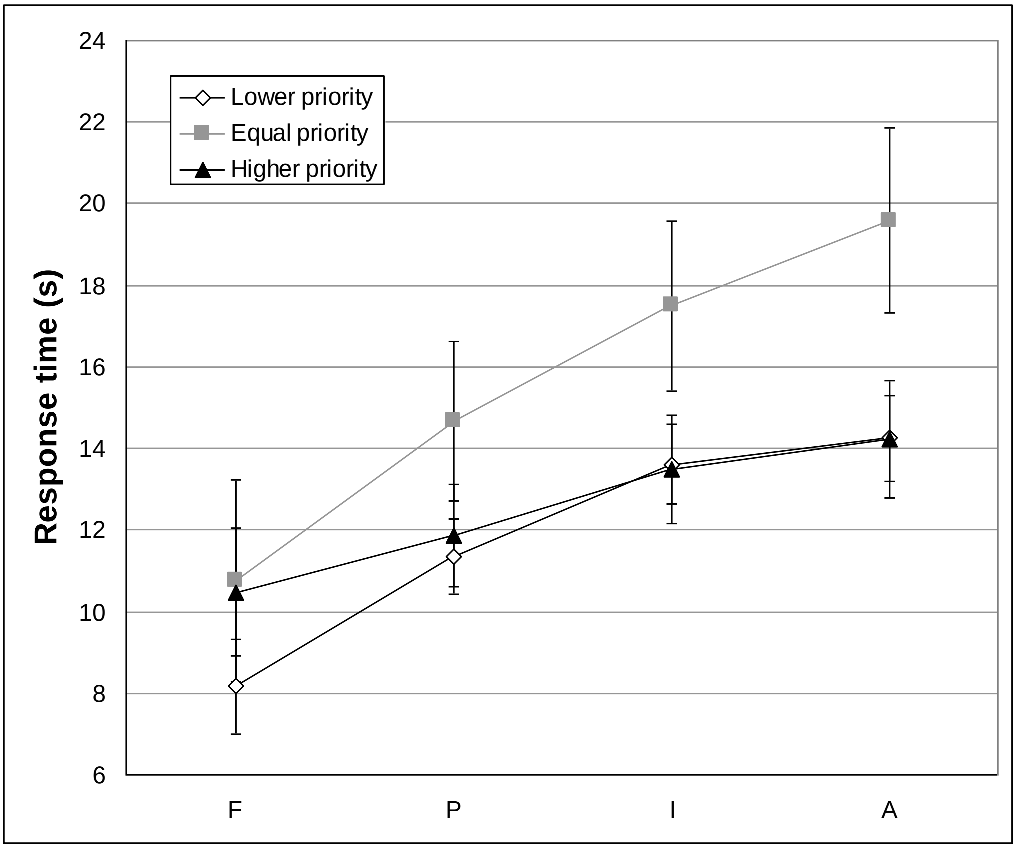


Figure 6-7. Mean response time to interrupting message per condition. Lines indicate priority.

seconds. The differences between conditions were not significant ($F(3,28) = 0.397$, $p = 0.76$; see Table 6-6). When incident handling time was analyzed per priority level, again no significant differences were found. This was contrary to what was hypothesized.

6.4.3 Message intrusiveness

Message intrusiveness scores showed a trend that approached significance ($F(3,28) = 2.63$, $p = 0.07$; see Figure 6-8) 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 intrusiveness ratings ($M_A = 3.5$ and $M_I = 3.7$) and not significantly different from the other two conditions. Although the differences in message intrusiveness 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 intrusiveness scale showed a

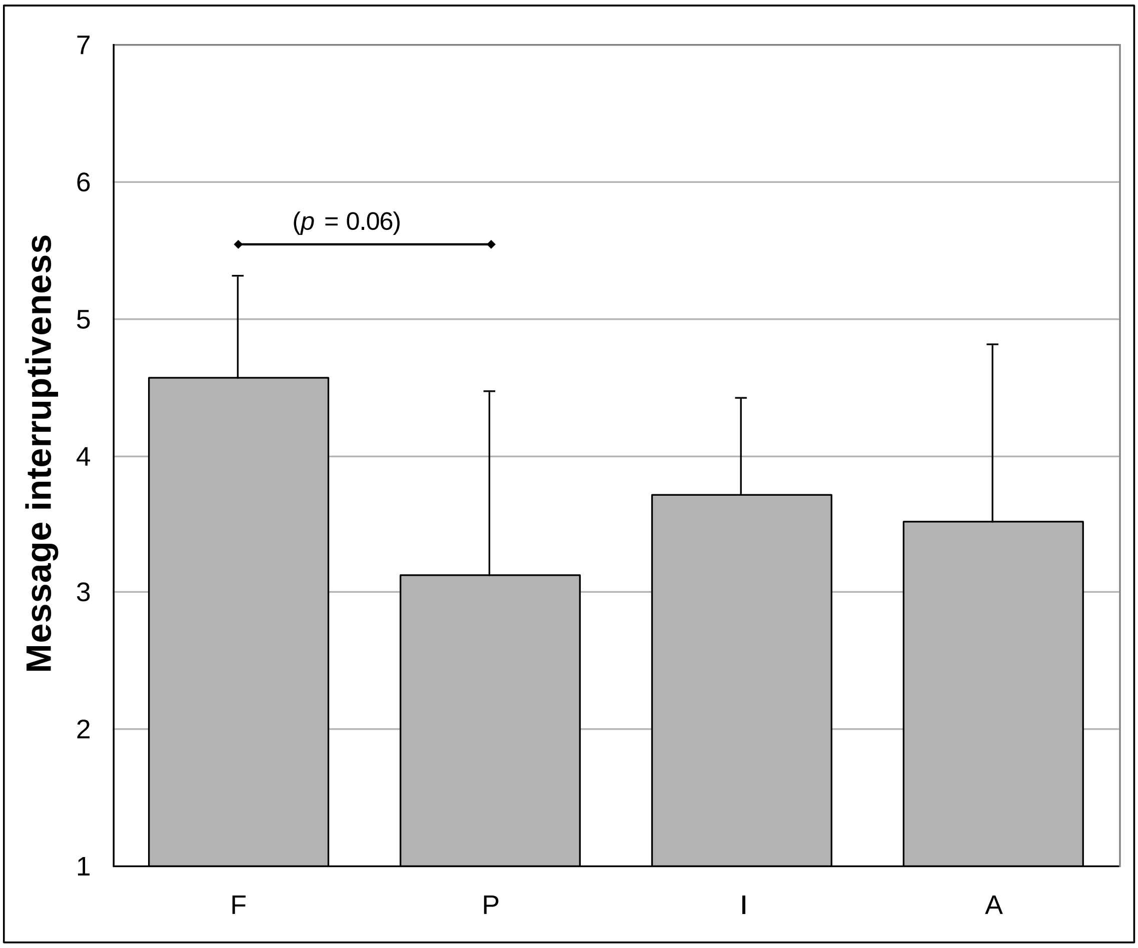


Figure 6-8. Mean message intrusiveness scores per condition.

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 ($p < 0.000001$).

6.4.4 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 ($F(3,28) = 1.35$, $p = 0.28$).

6.4.5 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 6-7, upper part). The rating scales on disruption, support and awareness showed no significant differences between conditions. Only the ratings on the satisfaction scale approached significance ($p = 0.06$; see Table 6-7, lower part).

Differences between conditions on the selected questionnaire items were further analyzed with multiple comparisons of mean ranks (see Table 6-7). The full message condition was considered significantly more interruptive ($M_F = 4.0$) than the postpone condition ($M_P = 2.6$) or adaptive condition ($M_A = 2.8$) ($p = 0.004$). Contrastingly, the full message condition also scored the highest satisfaction ratings ($M_F = 102$; not significant) and

Table 6-7. 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
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

participants were able to recognize the messages better in this condition compared to the adaptive condition.

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 intrusiveness 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.

6.4.6 Comparing the notification styles

When the different notification styles are compared across all results, the hypothesized strengths and weaknesses of each notification style become apparent (see also the hypotheses in section 6.3). 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 intrusiveness more than the other conditions. Unexpectedly, satisfaction ratings were highest for full messages.

Postponing messages maintains awareness of the environment, demonstrated by the highest number of targets noticed and lowest message intrusiveness. 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 intrusiveness 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 higher response time to incident messages.

6.5 Discussion

This study presented a refined notification system prototype that balances awareness of the environment with awareness of incoming messages in a mobile surveillance task. The notification system adapted the timing and appearance of incident messages, based on user activity and relative message priority. In a controlled mobile experiment, we evaluated the effects of adaptive notification styles on task performance, workload and user experience. 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 6-8 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.

Presenting incident messages as full messages increases unwanted interruption: they

Table 6-8. Validated claims based on 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	
	Number of targets	*Message intrusiveness*	*Incident handling time*	*Decision errors*	*Response time*
Full message (F)	ns	High	ns	Intermediate	Short
Postpone (P)	High	Low	ns	High	*N/A*
Indicator (I)	Low [a]	ns	ns	Low	ns
Adaptive (A)	ns	ns	ns	Low	Long [a]

[a] this effect is different than hypothesized.

are considered interruptive and people respond inappropriately 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 interface manipulations. 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 increased response time, presumably due to unfamiliarity with the adaptive behavior of the system (e.g. varying notification presentations). User preference for this adaptive behavior corresponds with the design choices implemented in the prototype system.

The current study contributes to the body of empirical research on mobile, context-aware notification systems. We demonstrated that a set of notification rules could determine appropriate timing and appearance of notification messages, and that this adaptive notification has positive effects on task performance and the user experience. Results from this study emphasize the positive influences of appropriate timing of interruptions found in other domains (e.g. desktop computing) (Cutrell et al., 2000; Bailey et al., 2006; Iqbal et al., 2008). It provides further evidence that postponing or deferring interruptions until users are available helps mitigate negative influences of interruptions (Iqbal et al., 2008). Additionally, the decrease in number of task errors found in earlier work is replicated here (e.g. Bailey et al., 2006). The implications of these results are that designers of context-aware notification systems should use full messages when awareness of messages needs to be high and should postpone messages when users' attention needs to be focused on the environment. Drawbacks to the use of icons on mobile devices are that they are sometimes overlooked, forgotten and require more display manipulations.

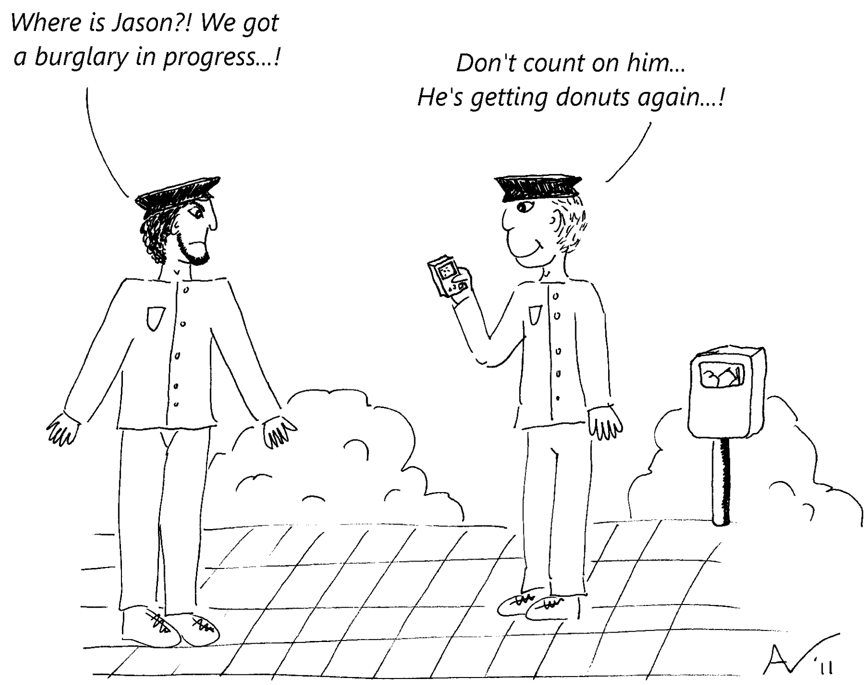
We did not find positive effects of adaptive notification on time on task (incident handling time) as reported elsewhere (Bailey et al., 2006; Iqbal et al., 2008). Nor did we find effects of adaptive notification on workload. The absence of significant differences might be explained by our manipulation: notification presentation in the adaptive condition necessarily had some overlap with the uniform conditions (see the notification matrix in Table 6-2). In addition, relatively long task durations (over 170 s) and large variability between participants in the study could have further masked differences between the conditions. This study leaves a number of questions still open. For example, is the increase in response time associated with adaptive notification a temporary effect due to unfamiliarity, or does it reflect a cost in the awareness trade-off? And why did the "full message" style receive both the highest satisfaction rating as well as the highest intrusiveness rating from the participants? Our conclusions must be carefully applied to the police domain as police officers did not participate in this study. The field study of the location-based notification system for police officers in Chapter 0 showed negative effects of interruptions on user satisfaction. Therefore we expect police end-users to be more critical of interruptions than student participants.

In conclusion of Part II of this thesis, there is great need for context-aware notification systems to be implemented in mobile, critical work domains, as demonstrated by the field study in Chapter 0. Consequently, the challenge lies in designing CAMS systems for these domains that balance the trade-off between awareness and interruption. These CAMS systems should support mobile information exchange by providing appropriately timed notifications in the appropriate modality. Both experiments in Chapters 5 and 6 contribute to a solution to this challenge, by stating design rationales on how appropriate timing and (visual and auditory) appearance of notifications can be realized, based on message priority, use context and user activity. Results from the first experiment on notification appearance show a slight improvement in task effectiveness with adaptive notification, at the cost of an increase in handling time. Users felt less interrupted and preference for the adaptive system was high. These results were echoed by the second experiment that compared four notification styles (timing and appearance) in a mobile, task-relevant setting. Presenting full messages directly maintained awareness of messages at the cost of unwanted interruption, while postponing these messages diminished interruption but also diminished awareness of messages. The adaptive notification style supported effectiveness of mobile surveillance in terms of less decision errors and a positive user experience, but at the cost of an increase in response time. We can conclude that adaptive notification supports surveillance effectiveness, but not efficiency. Results from both experiments also imply that relative message priority, the use context and user activity are three relevant context factors that can determine notification timing and appearance.

Up to now, our design effort for CAMS systems primarily focused on single user-device interaction. In addition, no professional end-users participated in the evaluation of the prototype systems. Because police officers on surveillance operate in teams, the results of the current controlled experiments need to be extended to support multiple users in teams and validated by incorporating end-users in the evaluation. Extending adaptive notification systems to teams addresses the second operational demand for CAMS: team collaboration support. For example, a networked adaptive notification system can support task allocation in police teams, by providing incident messages to the appropriate recipient using appropriate notification styles. This is expected to improve decision making between team members, team awareness and response time to incidents. The next part of this thesis focuses on designing and evaluating this team collaboration support.

PART III

TEAM COLLABORATION SUPPORT



7 Presentation Modality of Team Awareness Support

Abstract

In police surveillance, team awareness is important when asking team members for assistance. Staying aware of team members' availability helps to select the right (i.e. available) team member. This chapter investigates how the presentation modality (visual or auditory) of team information and communication influences team awareness, choice accuracy and mental effort in a surveillance task. An experimental prototype was created and implemented, which provided a visual overview of location, availability, expertise and means of transportation of team members as well as ongoing team communication. In a simulated police surveillance task employing trained participants (representative in age to police officers), this visual support prototype was compared to current auditory support, common in police practice. Results show that a visual overview of team information improves choice accuracy, while auditory information presentation improves team awareness at the level of perception, but not comprehension or projection. Mental effort did not differ between visual or auditory support. These findings suggest that presenting team information visually and communication aurally is an appropriate combination to support both team awareness and choice accuracy.

This chapter is based on the previously published material:

Streefkerk, J.W., Wiering, C., van Esch-Bussemakers, M., & Neerincx, M. (2008). Effects of Presentation Modality on Team Awareness and Choice Accuracy in a Simulated Police Team Task. In *Proceedings of the 52nd HFES Annual Meeting* (pp. 378-382). New York City, NY: HFES.

7.1 Introduction

Police officers on surveillance operate as ad-hoc, distributed teams. Team members have to keep track of the activities and whereabouts of colleagues to work effectively as a team (Convertino et al., 2005). Staying aware of team members' status and activities is referred to as team awareness (Ferscha, 2000). This ability improves team collaboration, for example in asking a colleague for assistance with an incident. In order to choose the appropriate colleague effectively, officers must monitor ongoing auditory communication to extract team information, such as location, availability, expertise and means of transportation of team members. Then, based on situational reasoning, they make a decision which colleague to approach. The accuracy of this decision depends on the situation; a high priority incident requires a rapid response, making the nearest colleague the most appropriate to ask for assistance. For efficient team collaboration between distributed team members, context-aware mobile support (CAMS) systems should support team awareness. Current mobile communication systems (e.g. radio transceivers) transmit all communication between team members via the auditory modality. Auditory information presentation is transient in nature and can sometimes interfere with task performance, especially in critical situations (see section 0). Furthermore, presenting this information aurally is cognitively demanding and can create cognitive (over)load. The current chapter focuses on which presentation modality is appropriate for CAMS to support team awareness, by comparing overviews of team information in the auditory and visual modality.

Earlier research on team awareness support (summarized in section 2.5.3) focused primarily on developing prototypes with visual overviews of team information including team member availability, location, resources, etc. The claim is that providing a visual overview of this team information as opposed to presenting it aurally will reduce cognitive load and increase team awareness which in turn improves the accuracy of team members' decisions. By nature, information via auditory modality is presented serially in time, while visual information can be presented and processed in parallel (Gaver, 1989). Using a visual overview, team members perceive relevant team information at a glance and do not have to monitor, extract and remember information from ongoing, transient communication streams. In addition, providing a visual overview of team communication allows team members to review and reread messages at an appropriate moment, in contrast to online auditory presentation. No previous studies directly compared auditory and visual presentation of team information, and tried to relate this to decision making accuracy and team awareness. Thus, it is as yet unclear how information on relevant team characteristics should be provided to improve team performance and awareness. Consequently, this study answers: *What are the effects of the visual and auditory presentation modality for team information on team awareness and decision making accuracy?*

We conducted a controlled experiment, comparing a visual support prototype to the current police practice of auditory communication, to determine the positive and negative effects of presentation modality on team awareness, choice accuracy and mental effort. The study employed a simulated police team task in a virtual environment using trained, non-professional participants, representative in age to police officers. The next sections describe the design of the team awareness support prototypes, based on a concise analysis of police team collaboration. Then, the evaluation method and results are presented and implications for the design of team awareness support for professional domains are outlined.

7.2 Design of team awareness support

7.2.1 Levels of team awareness

The concept of team awareness (TA) is related to situation awareness (SA), the ability to perceive and comprehend elements in the environment and project their status in the future (Endsley, 1988). Endsley's distinction in SA of Level 1 (perception), Level 2 (comprehension) and Level 3 (projection) can be applied to TA as well (Endsley et al., 2003). Perception is influenced by monitoring the current status of and information exchange within the team. Comprehension is aided by shared mental representations about the team, its tasks and goals. This can help to coordinate tasks to the right person. Projection is aided by reasoning about the future status of team members. Maintaining TA across the three levels will aid effective and efficient team collaboration. For example, perceiving the message that a colleague is underway to an incident and comprehending that he travels by car can help you project when he will arrive at the incident location.

7.2.2 Police team collaboration

We interviewed four Dutch police officers on police team collaboration during incidents. Specifically, we asked them how they decide which colleague to approach for assistance. Their answers provided the necessary background for the experimental design. Police officers on surveillance work together as distributed, ad hoc teams. When incidents occur, they communicate with colleagues to determine who should handle which incident. High priority incidents involve time-pressure and should be handled by two or more team members, while low priority incidents can often be handled by a single team member or postponed when no one is directly available. For this task allocation process to work fast and efficiently, team members have to 1) respond quickly to an incident message, 2) make the correct decision who to approach for assistance and 3) handle the incident quickly. For this, they have to stay aware of the location, status and activities of other team members (i.e. team awareness).

Police officers on surveillance currently have no visual overview of availability and location of team members. This makes it sometimes unclear which team member is available to handle an incident, potentially resulting in miscommunications or incidents that remain unattended for some time. Currently, all police communication proceeds via radio transceivers in the auditory modality. When asking a colleague for assistance, police officers consider four *context factors*, specifically the location, availability, means of transportation (on foot or by car) and expertise of the colleague. This decision has to be made carefully, to avoid interrupting a busy colleague or asking one who is (relatively) far away or without the right expertise to handle the incident. Which of these four context factors is most important depends on the priority of the incident (high, normal or low). The police officers in the interviews expected that a visual overview of team information would support team awareness.

Based on the interview results, we developed a concise decision model that captures the importance of the four context factors when choosing the most appropriate colleague, given the incident priority (see Table 7-1). Importance is indicated on a 3-point scale, where a value of 3 signifies the most important factor, while a value of 1 denotes the least important factor. In high priority tasks (e.g. a car crash with victims), time pressure is high, making location and means of transportation the most important factors, while expertise is considered less relevant here. In low priority tasks (e.g. interviewing a burglary victim), the expertise of the officers will be decisive in determining whom to ask for assistance, while location and transportation are less important. For normal priority tasks, both the estimated time of arrival and the expertise of the officers have to be considered. Finally, availability is always important to avoid interrupting colleagues already handling an incident. This model predicts the most accurate choice in the experimental scenarios used in this study.

Table 7-1. Importance of context factors given the task priority (3 = most important, 1 = least important).

	Task priority		
Factor	*High*	*Normal*	*Low*
Location	3	2	1
Transportation	3	2	1
Availability	2	2	2
Expertise	1	2	3

7.2.3 Hypotheses

If team information is made available in the visual modality, team members will be able to build more accurate mental models of other members' status and activities, they will have higher levels of TA, and tasks will be allocated more effectively. Furthermore, real life situations differ in choice complexity. Sometimes, the nearest officer is not available for a high priority task, or does not have the right expertise. When people maintain high levels of team awareness, choice accuracy will still be high, also for complex choice situations. In this study, choice complexity was varied systematically by distributing the context factors differently over the available team members at every choice moment (see section 7.3.3 Experimental design and manipulation). Based on these assumptions, the study will test the following hypotheses, comparing visual to auditory support.

1. Visual support will lead to higher choice accuracy scores than auditory support, especially in complex choice situations.
2. Visual support will lead to higher TA scores than auditory support.
3. Choice accuracy scores and TA scores will be correlated. High team awareness will lead to accurate choices, because users with a high level of team awareness will be able to take relevant context factors into account during task allocation and thus make more accurate choices.
4. Visual support will lead to lower mental effort than auditory support.

7.2.4 Implementation of auditory and visual prototypes

To test these hypotheses, two experimental prototypes were developed to present team information and communication, one using the auditory modality and one using the visual modality. The auditory prototype was modeled after current police practice, presenting all team communication via the auditory modality (i.e. via speakers). Team communication consisted of messages from the “dispatcher” to a specific team member, instructing to proceed to a specific incident. In the message, the incident, address, and incident priority were specified. The location of the officer could be derived from the address, the availability from the incident priority. As in police practice, a two-digit call sign for each colleague indicated their expertise (first digit) and their means of transportation (second digit).

The visual prototype showed an overview of locations of colleagues on a visual map of the environment (see Figure 7-1). This map was updated as colleagues moved through the environment. Similar to the auditory system, the call sign corresponded to their expertise and means of transportation. Availability could be derived from the color of the icons, depending on the priority of the current incident. In addition, team communication was presented visually as text messages in the lower part of the screen. All visual messages were preceded by a short auditory signal.

Both prototypes were implemented and used in a simulated police team task in a virtual environment. Using a virtual environment allows testing prototypes in a controlled setting, while still maintaining core task features of police surveillance (e.g. navigation, orientation, communication, ongoing task flow). We used Unreal Tournament (UT) software, that provides a realistic three-dimensional city environment (see Figure 7-2), and was used previously in emergency response research (Te Brake et al., 2006; Smets, te Brake, Lindenberg, & Neerincx, 2007).

7.3 Evaluation method

This study compared visual support (i.e. an overview of team information) to auditory support in a team collaboration task by measuring choice accuracy, team awareness and mental effort.

7.3.1 Participants

Twenty-six participants (15 male, 11 female), aged between 23 and 53 years (M = 34, SD = 12), completed this study. They were representative of police officers in age. All participants used computers daily and had no previous experience with police work. Eight participants played “first-person shooter” computer games occasionally.

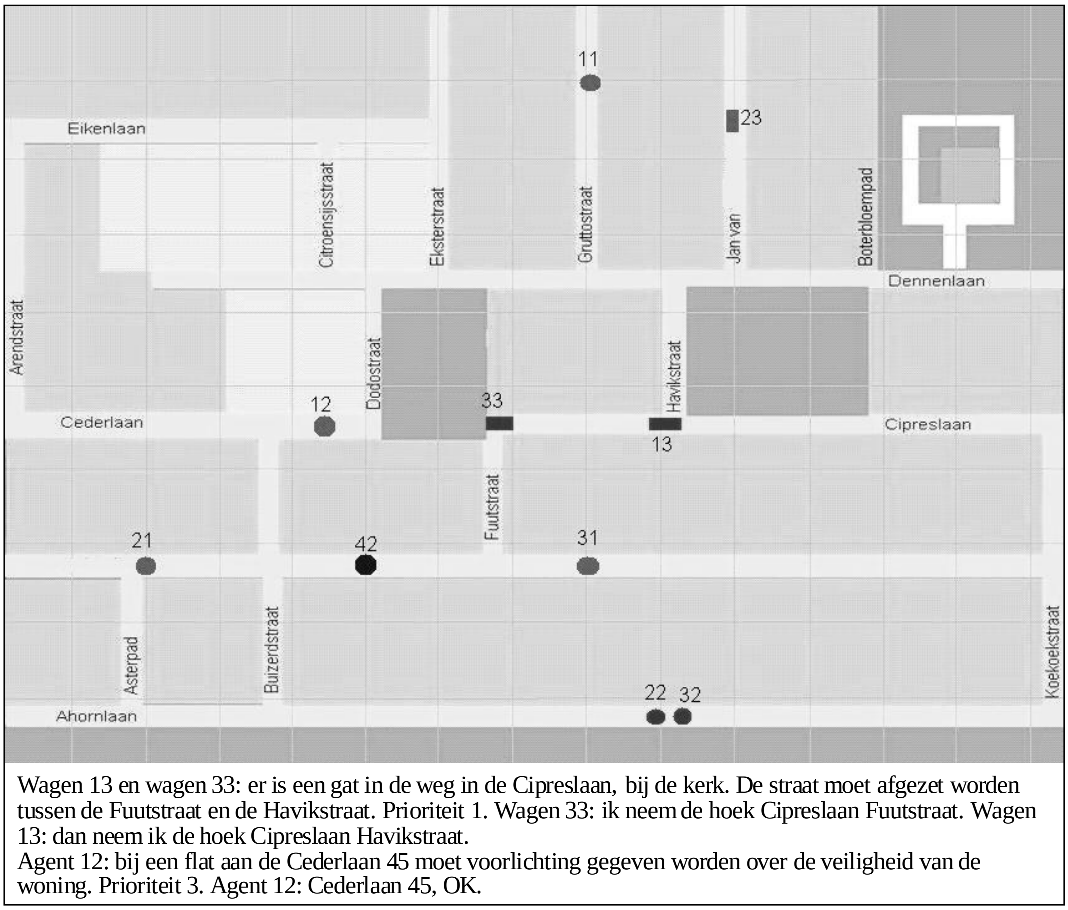


Figure 7-1. Visual geographical overview of the environment showing the position of team members and the text messages at the bottom.

Figure 7-2. Screenshot from the virtual environment.

7.3.2 Task

Participants were seated in front of two 17” desktop PC screens, one showing the virtual environment and the other the geographical overview and the incident messages. In the auditory condition, the second screen was blanked and messages were presented over speakers. Participants moved through the environment using a “gamepad” controller and could consult a paper map with street names at all times. Their task consisted of navigating along a predetermined route, while looking for specific threatening items (rifles and barrels). Nine virtual team members (controlled by *gamebots*) moved through the same environment. Participants had to keep track of these team members by monitoring communication (either visually or aurally). The “dispatcher” sent incident messages containing the two-digit call sign of the officer, the priority of the message and the address and details of the incident. The participant could hear or see communication from the team members acknowledging the message and announcing their arrival at the incident location. The position of team members in the visual overview was updated accordingly (see Figure 7-1). At six moments during each surveillance route, an incident message was directed at the participant (with call sign 42). They interrupted their surveillance to decide which colleague to ask for assistance with this incident.

7.3.3 Experimental design and manipulation

This experiment used a within-subjects, 2 (support modality: auditory or visual) x 3 (complexity: easy, intermediate or difficult) design. In one condition, the auditory prototype was used, in the other the visual prototype. All participants experienced both conditions and the same route was followed in both conditions. Two similar sets of messages were created, matched on length and content. The presentation order of the conditions and message sets was counterbalanced to avoid order effects.

Choice complexity was manipulated between the messages. By varying the distribution of the context factors (location, transport, availability and expertise) over the team members at each choice moment, three levels of complexity were created. In an easy situation, the nearest available colleague could arrive fastest at the task location, and was an expert for this particular task. When the context factors did not match the situation (mismatch), this made the decision more complex, because more colleagues had to be considered. A mismatch would be if the nearest colleague was unavailable, or if the nearest colleague would be on foot and an officer in a car could arrive faster. The intermediate level of complexity contained two or three mismatches, while difficult situations contained four or five mismatches.

7.3.4 Measures

In this experiment choice accuracy, team awareness and mental effort were measured. *Choice accuracy* was scored as the extent to which the choice for a specific team member matched the accurate choice according to the decision model (see Table 7-1). A maximum of three points could be scored for each choice. In each session, six choices had to be made. The complexity of the choice was distributed evenly; two easy, two intermediate and two complex choices were randomly presented in each session. Participants indicated their choice and reason for this choice on a paper choice questionnaire.

Team awareness (TA) was measured by interrupting the surveillance at predetermined locations in the game and presenting TA questions in the display ("freezing technique"; Salmon, Stanton, Walker, & Green, 2006). At six moments in each session, participants answered three questions on the three levels of TA. Questions were concerned with the content of the last message (perception), what expertise was necessary for an incident (comprehension) or the future availability of a team member (projection). TA scores consisted of points awarded to correct answers; a total of twelve points could be scored in every condition.

At two moments during each condition, the Rating Scale Mental Effort (RSME; Zijlstra, 1993) was used to measure *mental effort*. This paper-based rating scale ranged from 0 to 150 with accompanying descriptions of mental effort.

7.3.5 Procedure

The experiment took about two hours to complete. Participants filled out the computer experience questionnaire and informed consent form. They were trained on the decision model, the messages and on navigating through the virtual environment. When it was clear they understood the instructions, all participants followed the route once to familiarize themselves with the environment. The two experimental sessions (auditory and visual support) took about forty minutes to complete. During each session six TA questionnaires containing three questions, six choice questionnaires and two RSME rating scales were presented. After both sessions, participants were debriefed and paid.

7.4 Results

7.4.1 Choice accuracy

We investigated whether support modality had an effect on choice accuracy with a 2 (auditory or visual modality) x 3 (easy, intermediate or difficult complexity) repeated measures ANOVA. There was a significant main effect of modality; subjects made more accurate choices in the visual condition than in the auditory condition ($M_A = 1.31$, $M_V = 1.89$; $F(1,25) = 25.7$, $p < 0.001$). The main effect of choice complexity was also significant ($F(2,24) = 17.9$, $p < 0.001$; see Figure 7-3). Post-hoc analysis showed that participants had higher accuracy scores in the easy than in the intermediate complexity condition ($p < 0.01$) and higher scores in the intermediate than in the difficult condition ($p < 0.01$).

The interaction between modality and complexity was not significant; $F(2,23) = 2.09$, $p = 0.15$. There was no greater advantage of visual support over auditory support in difficult situations compared to easy situations. These results partially support hypothesis 1 in that the main effect of modality was found, but no interaction effect of modality and complexity was found.

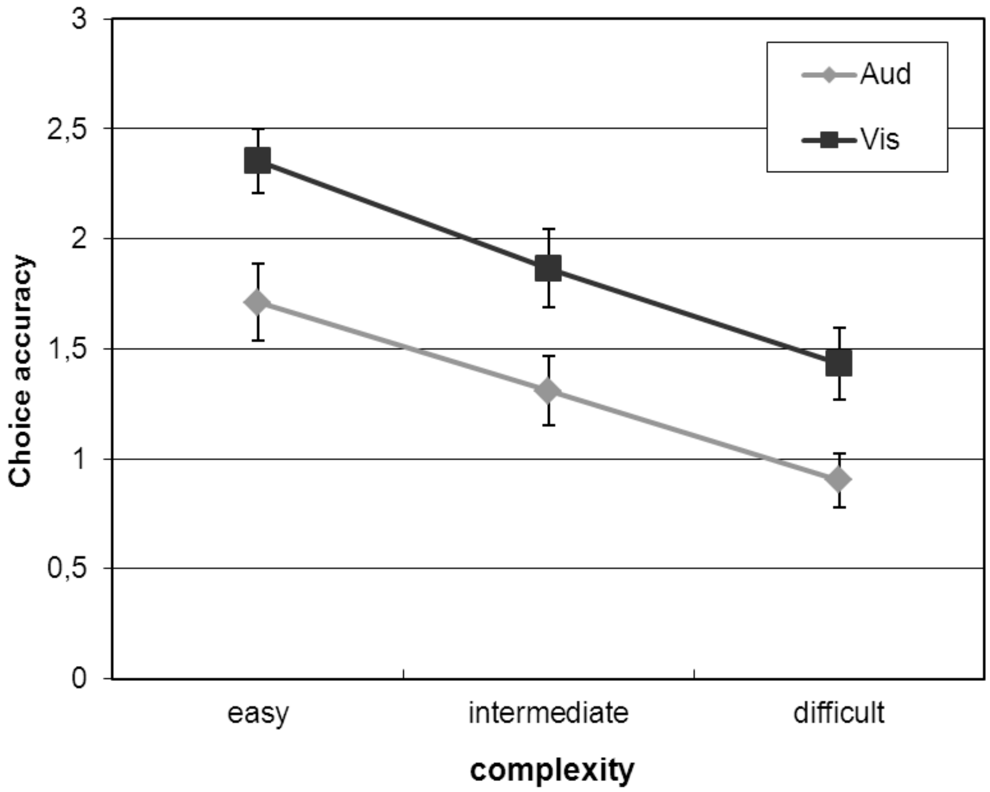


Figure 7-3. Mean choice accuracy for each complexity level. Lines indicate modality.

7.4.2 Team awareness

A 2 (auditory or visual modality) x 3 (TA levels) repeated measures ANOVA was conducted. The expected advantage of visual over auditory support, as stated in hypothesis 2, was not found; there was no significant main effect of support modality on TA ($M_A = 7.83$; $M_V = 7.88$; $F < 1$). To compare data across the different TA levels, z-scores were computed on the TA scores. The interaction between modality and TA levels was significant ($F(2,50) = 10.10$, $p < 0.001$). Post-hoc analysis showed that this difference was mainly due to higher TA scores on the first level (perception) in the auditory condition than in the visual condition ($p < 0.01$, see Figure 7-4). TA scores on the other two levels (comprehension and projection) were not significantly different.

7.4.3 Effect of TA on choice accuracy

To test whether high TA scores correlated with high choice accuracy scores, a regression coefficient was calculated on mean TA z-scores and choice accuracy scores per subject. The standardized regression coefficient was 0.08, which was not significant ($t(25) = 1.74$, $p = 0.10$). TA does not mediate the relation between modality and choice accuracy directly, since support modality does not have a significant effect on TA. Neither is there a consistent effect of TA on choice accuracy. As a result, no support could be found for hypothesis 3.

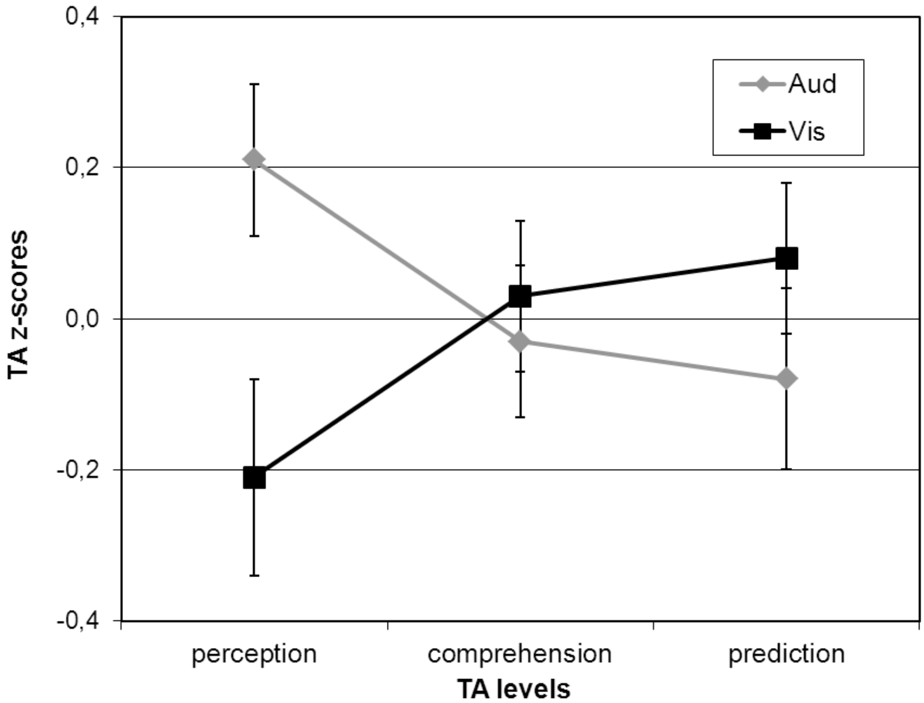


Figure 7-4. Mean team awareness scores for each TA level. Lines indicate modality.

7.4.4 Mental effort

Participants did not indicate they experienced higher mental effort in the auditory condition compared to the visual condition (M_A = 68.4; M_V = 66.4; $t(25) = 0.64$, $p = 0.53$). Although the average mental effort scores were slightly higher in the auditory than in the visual condition, this difference was not significant. Thus, hypothesis 4 was not supported.

7.5 Discussion

This study investigated whether team awareness and choice accuracy could be supported by providing a visual overview of team information, rather than relying on auditory communication only. A police team task with three levels of choice complexity was successfully simulated in a virtual environment. It was found that participants with a visual overview more often chose the appropriate team member for assistance. However, the visual support did not mediate choice complexity, nor did it influence mental effort as we expected. So, a visual overview was not more helpful (i.e. efficient) in complex situations than in easy ones. The reason might be that the participants in this study were not professionals and had only limited training on the decision model. As for team awareness, lower level team awareness (perception) improved by auditory rather than visual support, contrary to what was hypothesized. Comprehension and projection did not differ between visual or auditory support. These unexpected findings for TA can be explained by participants' remarks that the auditory messages were easier to perceive than the visually presented messages. In addition, they indicated that the visual presentation of ongoing communication required much effort and was distracting from their environment. This would also explain the lack of difference in mental effort between auditory and visual support. It seems that the benefits of the visual overview were diminished by the strenuous activity of following the ongoing communication visually.

The current study has a number of limitations. We employed a very concise decision model for experimental purposes. This model needs to be more elaborate to deal with ambiguous decision situations encountered in real life, but becomes harder to learn. In addition, no measure was taken for decision time, due to restrictions in the experimental setup. It may be that users made more accurate choices with a visual overview, but took longer to do so. Furthermore, the team awareness prototype was not implemented on a mobile device. Smaller screen size might impact team awareness negatively and decrease the speed with which users can work with the application. Finally, no police officers participated in this study. Because they are experienced in selective listening to auditory communication to keep track of team members, it will be interesting to see whether the same increase in choice accuracy is observed with actual police end-users.

To answer the research question on the appropriateness of the visual and auditory modality of information presentation, a trade-off was observed between choice accuracy and team awareness. A visual overview is more appropriate to increase choice accuracy, while auditory presentation helped users to stay more aware of messages and increase team awareness at the perception level. As a design recommendation, our research shows that when replacing current auditory systems by visual interfaces on mobile devices, designers must carefully consider this trade-off. It seems appropriate to combine visual and auditory information presentation, where team information is provided visually and the communication proceeds via the auditory modality. The high mental effort scores observed in this study suggest that integrating team information and communication for decision making is a cognitively strenuous activity. This is even more so when communication has to be perceived from a small screen. To mediate cognitive load, a support system can help to integrate this information and reason about an appropriate decision. For example, based on our decision model, a support system can provide a recommendation or advice about the most appropriate team member to approach for assistance. Such advice is expected to improve efficiency (because less time is needed to integrate information) and decrease cognitive load.

This study showed that when CAMS systems combine a visual overview of team information with auditory communication, this will improve effectiveness of team collaboration (i.e. decision making) as well as perception of messages. However, this study did not investigate whether this visual support improves *efficiency* of team collaboration (i.e. response time to and handling time of incidents). In addition, team collaboration entails interaction with team members, which was not required in the individual task in this study. To address these issues, the experimental paradigm is extended to include teams of three police officers handling incidents together. Furthermore, the team awareness support prototype is redesigned to provide advice to team members on who can handle which incident best, based on context information (user availability, location and incident priority). In the next chapter, this task allocation support is evaluated by police teams working together on a team surveillance task.

8 CONTEXT-AWARE TEAM TASK ALLOCATION SUPPORT

Abstract

To optimally distribute tasks within police teams during mobile surveillance, a context-aware task allocation system prototype is designed and evaluated with end-users. This system selects and notifies appropriate team members of current incidents, based on context information (officer availability, officer proximity to the incident and incident priority) and decision rules. Eight teams of three experienced police officers evaluated this system in a surveillance task through a virtual environment, by directly comparing it to a system without advice. Task performance (response and handling time, decision making), team communication, mental effort and user preferences were measured. Results show that advice reduces the amount of team communication and shows a trend towards less decision errors and more efficient navigation. Advice did not speed up the response to messages or handling of incidents. Three-quarters of the police officers preferred this system over the system without advice. We conclude that context-aware task allocation helps police teams to coordinate incidents efficiently.

This chapter is based on the previously published material:

Streefkerk, J.W., van Esch-Bussemakers, M., & Neerincx, M. (2009). Context-Aware Team Task Allocation to Support Mobile Police Surveillance. In *Foundations of Augmented Cognition (LNCS 5638)* (pp. 88-97). Berlin, Heidelberg: Springer.

8.1 Introduction

To work efficiently as a distributed team, mobile police officers need to coordinate their actions together. During surveillance, occurring incidents require fast and accurate responses from available team members. However, keeping track of availability and appropriately allocating tasks to team members is cognitively challenging in such a distributed work environment. To support this process, team awareness displays for mobile devices were developed that visualize team information in geographical overviews or using mobile awareness cues (Oulasvirta et al., 2007; Streefkerk, Wiering, van Esch-Bussemakers, & Neerincx, 2008). In the previous chapter we showed that a visual overview of team information and communication improved decision making, but did not improve team awareness nor mediate cognitive load. Users have to integrate team information from a small mobile display, straining their cognitive resources. Based on these findings, we suggested that context-aware mobile support (CAMS) systems should provide task allocation advice (e.g. on current incidents) to team members. Such mobile task allocation support has not been developed yet (see also section 2.5.3).

In the current chapter, we focus on how to present this advice and the effects of advice on team collaboration. Our support system prototype is redesigned to not only visualize team information, but also advise team members which officer can handle which incident best. This advice is based on officer availability, officer proximity to the incident location and incident priority. Selected team members are notified using appropriate notification styles (timing and presentation of notifications) to limit intrusiveness of notifications (see also Chapters 5 and 6). This team task allocation support is evaluated in a simulated surveillance task with police teams, addressing the following question: *What are the effects of providing task allocation advice, adapted to availability, location and incident priority, on team task performance, communication and decision making during mobile team surveillance?* We expect that this support will improve decision making on task allocation, resulting in lower response time, less team communication and positive user preferences compared to a system without advice. Based on the results of this study, implications for the design of CAMS for professionals are discussed.

8.2 Design of team task allocation support

To support the task allocation process outlined in the introduction, a mobile task allocation system is designed and implemented. This system uses a set of decision rules for task allocation (on officer availability, officer proximity and incident priority) to select team members and notify them of current incidents. Based on interviews with police officers (see section 7.2) and the decision model in Table 7-1, the decision rules were established. Because in the current study "expertise" and "mode of transport" did not differ within the team, we did not use these context factors for the decision rules.

8.2.1 Task allocation decision rules

The task allocation system uses context knowledge about availability (whether team members are handling an incident or not), proximity (how close team members are to the incident location), and whether the incident has high or low priority. This knowledge is acquired from location tracking data, user input to the system and established incident categorizations in the police domain. Based on this knowledge and a set of decision rules, the system selects the most appropriate team member(s) to handle the current incident. The selected team members receive an incident message with task allocation advice (i.e. "John and Mary can handle the burglary incident best"). As in Chapters 5 and 6, the notification style of these messages (information density and auditory salience) is adapted to limit unwanted interruptions. The system uses the following decision rules on task allocation:

1. *Proximity and priority*: high priority incidents require the nearest two available officers as soon as possible while low priority incidents require the nearest available officer.
2. *Proximity and availability*: if the nearest officers are busy with a lower priority incident, they should switch to the new incident. If they are busy with a higher priority incident, they should finish that incident first.
3. *Notification*: if officers are selected to handle an incident, the full incident message is presented with a salient notification sound. If they need to be aware that an incident is waiting for them, the system presents an indicator with a less salient sound. If they are not selected to handle the incident, an indicator is presented without sound.

8.2.2 Implementation of prototype

A support system prototype is implemented for experimental purposes using a simulated Personal Digital Assistant (PDA) on a touch screen monitor. The screen size measured 8 by

Figure 8-1. PDA screenshots showing the geographical map with the officer's location (left), an incident message (center) and the task list (right).

10.5 cm. The prototype shows a geographical north-up map with icons indicating team members' location, identity (name) and availability (red icon means busy, green icon means available) as well as the location of incidents. The map is centered on the user's location and can be panned to reveal the rest of the map. See Figure 8-1 for screenshots of the application.

Incident messages are displayed visually as full text messages with two buttons to "Accept" or "Ignore" the incident. Accepted messages move to the task list and can be checked off when the incident is finished. User actions ("Accept", "Ignore", "Finish") are used to infer user availability. Indicators are presented as small clickable icons in the lower right corner of the screen, opening the incident message when clicked.

8.3 Evaluation method

In this study, police teams performed a surveillance task through a virtual city environment. The task allocation support system presented low or high priority incident messages. At these moments, team members decided who would handle which incident, navigated to the incident location and handled the incident. Task allocation advice, notification presentation and communication were manipulated, creating two experimental conditions (with or without advice). Effects of advice on task performance, mental effort and preference were compared between the two conditions.

8.3.1 Participants

Eight teams of three police officers (20 male, 4 female) aged between 20 and 52 years (M = 33.0, SD = 9.9) completed this study. All team members had collaborated previously with each other on surveillance and were experienced police officers (average 11.2 years of experience). All participants used personal computers on a daily basis, four participants played computer games daily while fourteen had no computer game experience. Two teams used a mobile device for police work.

8.3.2 Surveillance task

Teams performed the surveillance task through a three-dimensional virtual city environment, created in Unreal Tournament (Te Brake et al., 2006). This environment simulated core task features of mobile police surveillance such as navigation, incident handling and team coordination. Each team member controlled a virtual character in this environment and they could recognize each other's virtual character by a colored uniform (red, blue or green).

The surveillance task was a time-paced, scenario-based task. It required the teams to find a maximum of 30 targets, represented by barrels that appeared at random locations throughout the environment. At predetermined moments during the surveillance round, the PDA presented in total twelve incident messages to the team members. Six incident messages indicated high priority incidents (e.g. burglary or fighting in progress), which had to be handled by two team members together. The other six indicated low priority incidents (e.g. past vandalism), which could be handled by a single team member. Team members suspended the surveillance task to read the incident message and communicated with their colleagues (using a headset) who would handle this incident. The selected team member(s) responded to it (using the "Accept" or "Ignore" buttons below the message) and navigated to the incident location as fast as possible (see Figure 8-2). Handling the incident consisted of reading and memorizing the incident description on screen. When done, they checked the incident off the task list and returned to the surveillance task. Participants could decide for themselves when to attend to each message, whether or not to accept an incident and which of their colleagues to approach for assistance.

8.3.3 Experimental design and manipulation

A within-subjects design was employed with two experimental conditions (with or without advice). In the advice condition, the system provided task allocation advice and adaptive

Figure 8-2. Screenshot from the virtual environment showing an incident location (vertical red bars; left) and a police officer behind the workstation (right).

notification based on the decision rules. Team members could choose to communicate with all team members (open channel) or selected team members only (closed channel). In the condition without advice, full incident messages were presented to all team members without task allocation advice and the communication channel was open for all team members. Two similar experimental scenarios (equal duration, number and type of incidents) were established in cooperation with two experienced police officers of the Dutch police to maximize external validity. All teams participated in both conditions and the presentation order of the conditions and scenarios was counterbalanced to avoid order effects.

8.3.4 Measures

Before the experiment, age, gender, (mobile) computer experience, game experience and police experience were assessed using a questionnaire. Furthermore, spatial ability was assessed in a computer-based spatial rotation task (Neerincx et al., 1999).

During the experimental sessions, surveillance task effectiveness was measured as the number of *targets* found by the team members and number of *errors* in decision making on task allocation. An error was counted each time a team member did not follow the decision rules. Efficiency was measured as the *response time* to incident messages in seconds, *incident handling time* in seconds, *total time* per incident in seconds and total *distance* traveled (in game units). These variables were measured per team and averaged over incidents. Furthermore, the number of *communication utterances* on task allocation between team members was counted. An utterance was defined as one complete sentence or single word (most often “Yes”, “No”, “OK”) from one team member.

After each session, *mental effort* was measured using the RSME (Zijlstra, 1993). *Subjective judgments* on team performance were measured individually with five 7-point team effectiveness scales on coordination, misunderstanding, planning, efficiency and uncertainty. After both sessions, team members were asked individually to compare both experimental sessions. On the *preference* questionnaire they indicated which of the two prototypes they would prefer in their daily police practice regarding task allocation advice, presentation of the messages and team communication.

8.3.5 Apparatus

Participants were seated behind two 17” monitors, one above another. The top monitor displayed the virtual environment and the incident details. Participants moved through the environment using a game controller. The bottom (touch-screen) monitor displayed the simulated PDA and communication interface (see Figure 8-2). They communicated via headsets with a microphone. To avoid overhearing each other in the closed channel condition, city background noise was played over the headset. While navigating through the environment, the PDA was blanked out to represent real-life walking conditions.

8.3.6 Procedure

The experiment took about three hours to complete. First, the individual characteristics questionnaire and spatial ability test were administered. Participants received instructions on the surveillance task and familiarized themselves with navigation and incident handling in two short practice scenarios (first without advice and second with advice). In the condition without advice, participants were instructed to follow the set of decision rules for task allocation (see paragraph 8.2.1), while in the advice condition the system provided task allocation advice. The two sessions took about twenty minutes each, after which the RSME and the performance questionnaires were administered. After both sessions, the preference questionnaire was administered and teams were debriefed.

8.4 Results

All task performance variables were averaged over teams and compared between the condition with advice (A) and without advice (WA). Number of decision errors, response

Table 8-1. Means for number of decision errors, response time, handling time and total time per condition and priority level.

		Decision errors		Response time (s)		Handling time (s)		Total time (s)	
Condition	**Priority**	*Low*	*High*	*Low*	*High*	*Low*	*High*	*Low*	*High*
With advice		1.8	1.5	14.4	13.2	64	66	111	136
Without advice		2.5	2.5	9.8	11.5	68	73	117	142

time, incident handling time and total time were analyzed with repeated measures ANOVA with advice (with or without) and incident priority level (high or low) as factors (see Table 8-1). Bonferroni post-hoc analyses were performed. T-tests for repeated measures were used to analyze the number of targets, distance travelled, communication and mental effort (see Table 8-2). Data from the subjective judgments questionnaires was analyzed using non-parametric Sign tests. Multiple regression analyses were performed on task performance measures, communication and mental effort with age, spatial ability, education, computer experience, game experience and police experience (averaged over teams) as predictor variables.

8.4.1 Surveillance task effectiveness

The results showed that the number of decision errors on task allocation was lower with advice (M_A = 3.4) than without advice (M_{WA} = 5.0). This effect approached significance ($F(1,6) = 4.37$, $p = 0.07$). The number of errors did not differ between high and low priority incidents, and no interaction effect was found between advice and priority (see Figure 8-3). Furthermore, slightly more targets were found in the advice condition (M_A = 18.5) compared to the condition without advice (M_{WA} = 17.4). However, this difference was not significant ($t(7) = -0.44$, $p = 0.67$). Consequently, the advice did not help teams to find more targets, but

Table 8-2. Means for number of targets, distance, communication utterances and mental effort per condition.

Condition	*Number of targets*	*Distance*	*Communication utterances*	*Mental effort*
With advice	18.5	332734	23.3	51.0
Without advice	17.4	383425	33.2	49.3

a trend towards less decision errors was observed.

Regression analysis showed that variance in targets found was explained by age (R^2 adj. = 64%, B = -0.5, $p < 0.05$); younger teams found more targets in the condition without advice. In the advice condition, no significant predictors were found on these variables.

8.4.2 Surveillance task efficiency

Response time to incident messages was slightly longer with than without advice (M_A = 13.8 s vs. M_{WA} = 10.6 s), however not significant ($F(1,6) = 3.25$, $p = 0.12$). This can be explained by the extra line of message text (with the task allocation advice) that had to be read in this condition. Response time did not differ significantly between high and low priority messages overall, but the interaction effect between advice and priority was significant ($F(1,6) = 7.24$, $p < 0.05$; see Figure 8-4). Post-hoc analyses showed that there was no difference in response time to high priority messages with or without advice, but that response time to low priority messages was longer with advice than without advice ($p < 0.05$). This is an effect of message presentation; low priority messages were presented less salient (i.e. with an indicator) to team members for whom the message was not intended. Consequently, when they were busy, these team members appropriately waited to respond to these low priority messages. Thus, with advice, response time was appropriate for the message priority.

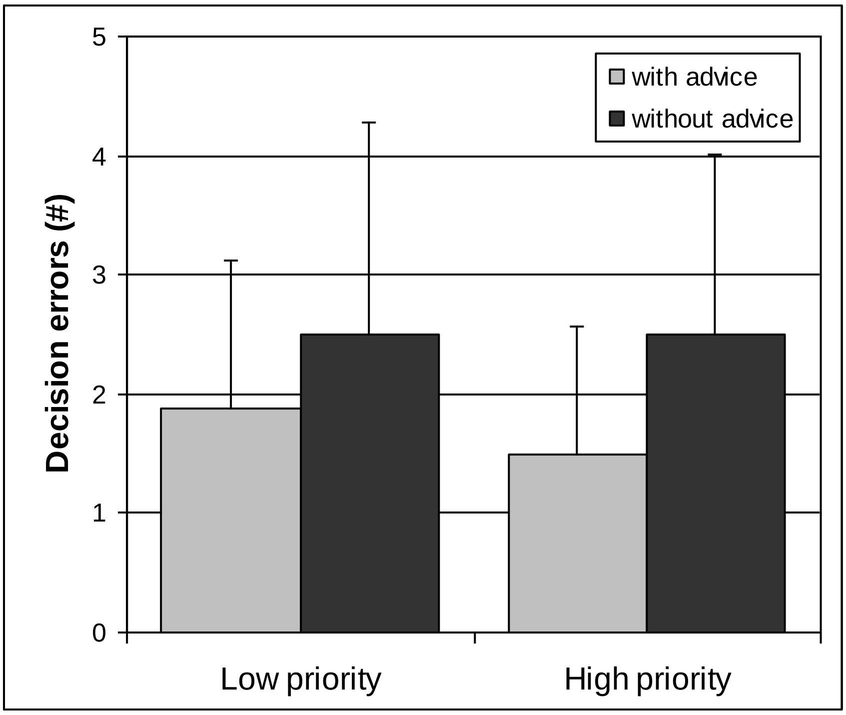


Figure 8-3. Mean number of decision errors on low and high priority incidents.

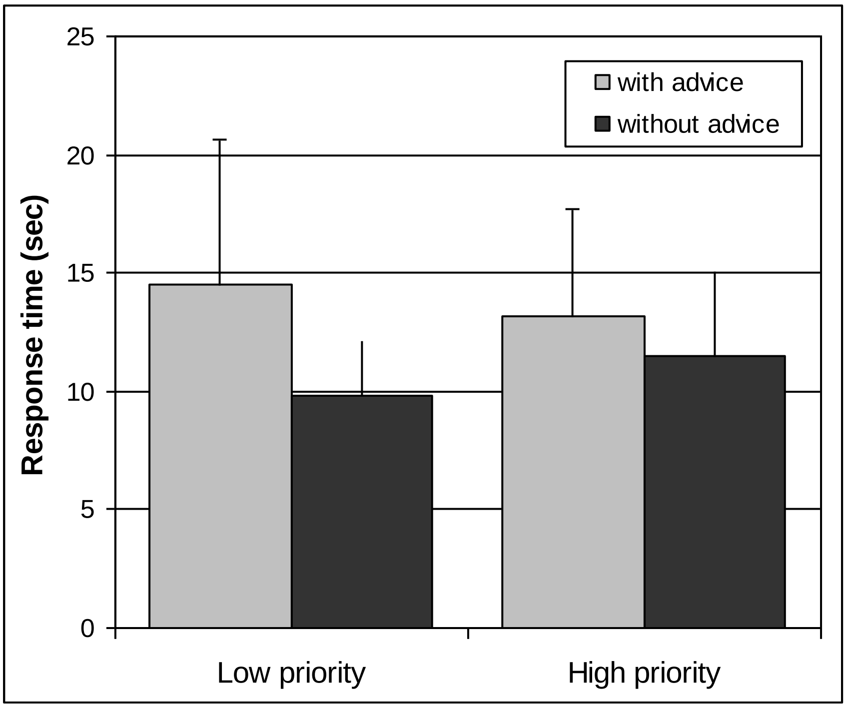


Figure 8-4. Mean response time to low and high priority incident messages.

Both incident handling time and total time were significantly predicted by the age of the team members, demonstrated by regression analysis. Variance in incident handling time was strongly predicted by age (R^2 adj. = 91%, B = -5.37, $p < 0.01$) and variance in time on task was also predicted by age (R^2 adj. = 65%, B = -4.92, $p < 0.05$); without advice, older teams were faster than younger teams. This effect was not present in the advice condition.

Consequently, we analyzed incident handling time and total time with age (grouped above or below the mean) as covariate in the ANOVA model. Incident handling time showed no significant differences with or without advice, or between high or low priority incidents. Total time showed no significant difference with or without advice, but less time was spent on low priority incidents compared to high priority incidents (M_{LOW} = 114 sec vs. M_{HIGH} = 139 sec; $F(1,6) = 33.8$, $p < 0.005$; see Figure 8-5). This is due to the fact that only one team member was required to handle low priority incidents, reducing extra time spent on task allocation, communication and navigation. The analysis showed no significant effects for the age covariate for incident handling time or total time. So, teams did not handle incidents faster with advice than without advice.

The difference in total distance traveled with or without advice approached significance ($t(7) = 2.13$, $p = 0.07$). Less distance was traveled in the advice condition. Regression analysis showed that variance in distance traveled was significantly predicted by age (R^2 adj. = 61%,

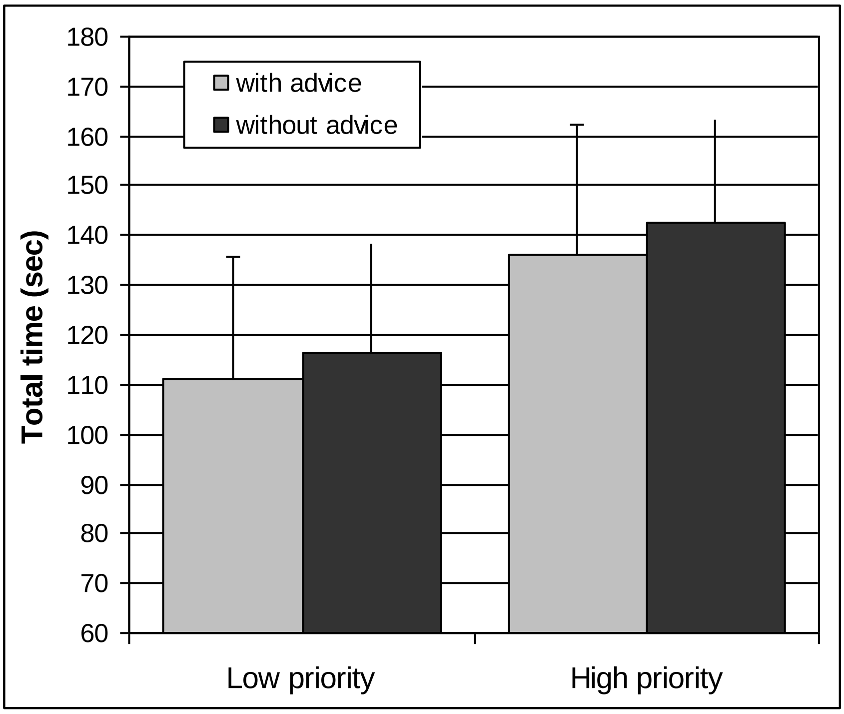


Figure 8-5. Mean total time for low and high priority incidents.

B = 8956, $p < 0.05$); younger teams traveled less distance in the condition without advice. This effect was not present in the advice condition.

8.4.3 Communication

The number of communication utterances on task allocation differed significantly between conditions ($t(7) = 4.17$, $p < 0.005$). In the advice condition, team members communicated less on task allocation than in the no advice condition (23 and 33 utterances respectively). When they had the choice between open or closed channel of communication in the advice condition, we observed that almost all teams preferred and used an open channel. An interesting observation on team communication is that without task allocation support, tasks were allocated to whoever called first or loudest. While this was not the most appropriate allocation based on the decision rules (resulting in a decision error), teams generally accepted this division of tasks.

8.4.4 Mental effort

There was no significant difference in mental effort with advice (M = 51.0) and without advice (M = 49.3). Regression analysis showed that mental effort in the condition without advice was predicted (R^2 adj. = 89%) by spatial ability (B = -5.80, $p < 0.05$), game experience (B = 14.8, $p < 0.05$) and education (B = 8.06, $p < 0.05$); participants with high spatial ability and less game experience indicated lower mental effort ratings. Mental effort in the advice condition was also predicted (R^2 adj. = 84%) by spatial ability (B = -5.17, $p < 0.05$) and game experience (B = -8.18, $p < 0.05$), but showed that participants with more game experience indicated lower mental effort.

8.4.5 Subjective judgments

The rating scales on team performance showed no significant differences between conditions. Participants did not rate their team performance differently in one of the conditions. The mean scores showed a ceiling effect (5.9 and 5.8 out of 7 for the condition with and without advice respectively).

When asked to compare both conditions, results showed that 76% of the participants preferred the advice condition in their daily police work because it supported decision making on task allocation. Half of the participants preferred the advice condition because of the lower disruptiveness of messages. However, 58% of the participants found it to be more difficult to divide attention between the PDA and the surveillance in the advice condition. This was presumably due to the use of indicator icons without sound. Team members had to focus their attention on the PDA to observe these icons.

8.5 Discussion

To support team collaboration, we expected that distributed teams could be helped by receiving task allocation advice. This study evaluated a support prototype on a mobile device that gave advice based on decision rules for three context factors (availability, proximity and priority). In a surveillance task with experienced police teams, two conditions (with and without task allocation advice and adaptive notification) were compared. As expected, with task allocation advice, significantly less team communication was required and trends towards less decision errors and more efficient navigation with advice are observed. In addition, with advice, response time to incident messages is more appropriate for the incident priority. The majority of the officers preferred the task allocation support in their daily work, although some found the adaptive system behavior hard to understand. Our results show that context-aware task allocation support helps police teams in decision making and communication on task allocation.

Contrary to our expectations, no effects of advice were found on time on task, incident handling time and mental effort; the advice does not make police officers faster nor lessen their mental effort. The time benefits of the task allocation advice may be too small compared to the total time spent on an incident (over 120 seconds). Regression analysis showed that older teams were faster in handling incidents without advice, but that with advice, no difference was found between younger and older teams. Presumably, younger teams benefit more from the task allocation advice than older teams. With regard to mental effort, learning to work with an adaptive system might have increased officers' cognitive load. As an adaptive system becomes more familiar over time, this effect is expected to decrease with prolonged system use.

The synthetic task environment allowed the teams to interact and collaborate on handling the incidents. Participants remarked they could immerse themselves in the simulation. The limitation of the current setup is that it necessarily represents a simulation of police work without dangerous situations. In fact, even without advice, teams could still handle all incidents easily and the costs from decision errors (e.g. extra navigation time) were relatively small. It can be expected that the benefits of advice are more profound in real life police practice. We will focus more on the benefits and costs in the next chapter. Furthermore, in this study only communication within the own team was broadcasted. In real life, communication from an entire district (up to 20 officers) is broadcasted over a single channel. As advice reduced the amount of communication, this may benefit the limited communication bandwidth in the police domain.

Our results have implications for the design of task allocation support systems on mobile devices. The study shows that CAMS systems can provide task allocation advice based on relatively simple decision rules on availability, proximity and priority. This advice improves team communication and shows a trend towards more accurate decision making and more efficient navigation, but does not improve the speed of incident handling. These benefits were found in a controlled lab study with police teams, in which the advice from the system was always 100% correct. However, in real life, the reliability level of context-aware systems is expected to be lower (Wickens et al., 2007). To validly state that context-aware task allocation supports task performance, we need to investigate whether *partially* reliable advice entails extra costs in terms of task performance, or that it entails the same benefits as "perfect" advice. In the next chapter, we will focus on how teams deal with partially reliable task allocation advice and how (shared) situation awareness and trust influence task performance in such situations.

9 BENEFITS AND COSTS OF TEAM TASK ALLOCATION SUPPORT

Abstract

Task allocation support from CAMS systems is at best partially reliable, because these systems are not aware of spontaneous context events that police officers run into (e.g. encountering a person in need of assistance). We need to know whether the benefits of this support outweigh the costs. Based on a use case, we specify claims regarding benefits and costs and test these claims in a team surveillance task in a synthetic task environment. Eighteen teams of three trained student participants handled incidents while using a prototype system that provided task allocation advice to half of the teams and no advice to the other half. For 50% of the incidents, context events caused this advice to be incorrect (partial reliability). We compared task performance, shared situation awareness (SA) and trust with and without advice. Overall, no difference in task performance, SA and trust was found with or without advice. Incorrect advice slows response time, but does not cause more decision errors or team communication, compared to no advice. This study shows that partially reliable advice slows down task performance, while maintaining accuracy. Consequently, in situations with high time-pressure, task allocation support could be worse than no support.

This chapter is based on the previously published material:

Streefkerk, J.W., van Esch-Bussemakers, M., & Neerincx, M. (2010). Balancing Costs and Benefits of Automated Task Allocation in Mobile Surveillance. In *Proceedings of the 2010 European Conference on Cognitive Ergonomics* (pp. 99-107). Delft, the Netherlands.

9.1 Introduction

The previous chapter demonstrated that team collaboration (e.g. in police surveillance) benefited from a context-aware mobile support (CAMS) system that provided task allocation advice. A prototype CAMS system selected team members and advised them on incident handling, based on their availability, proximity to and priority of an incident. A controlled study with police teams showed that this advice reduced team communication and that advice was preferred over no advice. However, because not all context information may be known to a CAMS system, a mismatch can occur between the actual context and what the system perceives the context to be (i.e. the system's context model) (Wickens, 2000). For example, the task allocation system monitors officers' location and availability and allocates incidents to available officers. As police work takes place on the street, police officers encounter spontaneous events, such as people on the street that need assistance. When the context model of the support system does not incorporate the knowledge that an officer is helping a person and is no longer available, the system will incorrectly allocate a new incident to him. Although these context events may not happen very frequently, they cause the system to be only partially reliable. This can negatively impact team task performance, shared situation awareness and trust in the system.

With regard to task performance, incorrect task allocation advice may lead to more decision errors on incident handling, longer response time to incident messages and more team communication (Wickens et al., 2007). Furthermore, shared situation awareness (the shared understanding of team members' status and activities) may be negatively impacted by incorrect advice (Gorman et al., 2005). Finally, trust in the system determines whether the advice will be followed. Trust needs to be tuned to the reliability of the system; too little trust leads users to disregard the advice, while too much trust leads to over-reliance on the advice (Parasuraman et al., 2004; McGuirl & Sarter, 2006). These negative effects have lead some authors to conclude that automated support systems may not be appropriate for critical domains (Sarter & Schroeder, 2001; Cummings, 2004; Wickens et al., 2007). Specifically, a literature review estimated that below a reliability of 70%, using an automated system is worse than having no automation (e.g. decision making based on raw data) (Wickens et al., 2007). Therefore, for the design of CAMS we need to study whether these potential negative effects of incorrect advice outweigh the benefits of having this support at all. This chapter will focus on the question: *What are the effects of partially* *reliable task allocation support on task performance and are these mediated by situation awareness and trust in the system?*

Based on a task allocation support use case, we specify claims in terms of task performance (decision errors, response time, handling time and communication), shared situation awareness and trust. Then, these claims are empirically tested in the same team surveillance task used in the previous study. For allocation of the incidents, half of the teams receive advice from the system, while the other half receives no advice. During half of the incidents, spontaneous context events occur, such as roadblocks or people in need of help, which are unknown to the system. These events cause the task allocation advice to be incorrect 50% of the time. We chose this reliability level to investigate whether partially reliable advice is still better than no advice and whether task performance is worse with incorrect advice than without advice. For this study, we employed two police officers to establish the experimental scenarios, and conducted the experiment with teams of trained student participants, because of the limited availability of professionals.

9.2 Operationalization

Following the Cognitive Engineering method (Neerincx et al., 2008), this section defines the expected benefits and costs of task allocation support in police surveillance. Based on previous research and police surveillance characteristics (see section 8.2.1), we specify a task allocation use case. From this use case, we derive claims on benefits and costs of task allocation support that form the basis for our hypotheses. This section concludes with a description of the prototype used in the experiment to test these hypotheses.

9.2.1 Task allocation support use case

To illustrate the claims, Table 9-1 elaborates a use case of how a CAMS system provides task allocation advice. The use case specifies the actors and two activity sequences; one without and one with the occurrence of a spontaneous context event. The primary activity sequence shows reliable task allocation advice (studied in Chapter 8). How context events (unknown to the system) cause the system to be partially reliable is demonstrated in the alternative activity sequence. Here, one of the actors is helping a panicked person when CAMS allocates a new incident to him. This incorrect advice forces the team to rethink the allocation of the incident (e.g. a team member who is further away, but not hampered by the context event can better handle the incident).

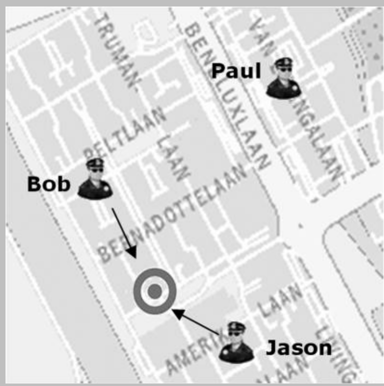
Table 9-1. Task allocation support use case.

Use case: CAMS provides task allocation advice on incident handling

Actors
Jason, Bob and Paul are three police officers on surveillance in the same city district. They are equipped with a CAMS system that provides task allocation advice based on availability, proximity to the incident location and incident priority.

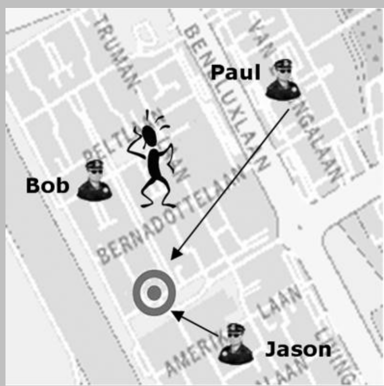
Primary activity sequence
Correct task allocation advice

Police officers Jason, Bob and Paul are on surveillance in a city district. Suddenly, they are notified by their CAMS system that a burglary is in progress. Jason and Bob are both available, and close to the incident location. CAMS advises Jason and Bob to go to the burglary, because Paul is further from the incident location. Because of the advice, Jason and Bob know that they have to handle the incident together and they can arrive quickly at the incident location. Paul remains in his part of the district.



Alternative activity sequence
Incorrect task allocation advice due to a context event

Police officers Jason, Bob and Paul are on surveillance in a city district. Suddenly, they are notified by their CAMS system that a burglary is in progress. Jason is available, but Bob is assisting a panicked person whose car was stolen. As Bob just encountered this person on the street, CAMS was unaware of this. CAMS advises Jason and Bob to go to the burglary, because Paul is further from the location. Upon receiving this advice, Bob has to contact Paul and Jason both to tell them he is currently not available. It takes Paul some time to arrive at the incident location, where Jason is impatiently waiting for him.



9.2.2 Benefits and costs of task allocation support

Based on the use case, we can now claim the expected benefits and costs of task allocation support in terms of task performance, shared situation awareness and trust (see Table 9-2). The claimed benefits of correct task allocation advice are less decision errors on task allocation, shorter response time to and handling time of incidents, and lower cognitive load due to less team communication. In effect, an automated task allocation system provides an optimal distribution of incidents among available team members. However, task allocation advice is expected to come at the cost of decreased shared situation awareness, because people tend to over-rely on advice from the system and do not have to actively keep track of the operational situation (e.g. locations of colleagues and incidents).

The claimed benefits and costs of *incorrect* task allocation advice (alternative activity sequence) are presented in the right column of Table 9-2. The benefits are that partially reliable advice helps to maintain situation awareness better because people do not over-rely on this advice. However, as argued in the introduction, we expect task performance to be worse with partially reliable advice (in our case 50% reliable) (Wickens et al., 2007). Specifically in incidents where the advice is incorrect, the costs are more decision errors on task allocation, longer response and handling time and higher cognitive load due to more team communication. Furthermore, as people notice a system sometimes provides incorrect advice (e.g. partial reliability), they will trust the system less.

These claims can now be formulated as the following hypotheses:

1. **Task performance** is overall worse with partially reliable advice than without advice.

Table 9-2. Claims on benefits and costs of task allocation.

Core function: Providing automated task allocation advice on incident handling to distributed mobile team members.			
Objective	An optimal distribution of incidents over team members, based on availability, proximity and priority.		
Claims	*Benefits*	*When advice is correct:* ✓ Less decision errors ✓ Shorter response and handling time ✓ Lower cognitive load due to less team communication ✓ Higher trust	*When advice is incorrect:* ✓ Maintain shared situation awareness better
	Costs	*When advice is correct:* × Decreased shared situation awareness due to over-reliance	*When advice is incorrect:* × More decision errors × Longer response and handling time × Higher cognitive load due to more communication × Lower trust

Incorrect advice, compared to no advice, will cause more decision errors, longer response time, and longer incident handling time. In addition, more team communication is needed to properly allocate the incidents.

2. When confronted with partially reliable advice, people will maintain shared **situation awareness** better than without advice.
3. We expect **trust** to decrease more with partially reliable advice than without advice.

9.2.3 Prototype implementation

The same task allocation prototype and underlying decision rules on task allocation are used as in the previous study (see section 8.2). Previously, the indicator icons without sound for low priority messages were not appreciated. Instead, the current prototype presents a *summary*: a button in the screen with a keyword describing the message (Figure 9-1, left), accompanied by a low salient sound. Clicking on a summary opens the full incident message. Furthermore, the functionality of the prototype is extended so that incidents can be transferred to another team member when necessary (e.g. for reallocation). Tasks can be transferred by clicking on the incident in the task list (Figure 9-1, middle) and clicking on the name of the appropriate team member in the transfer screen (Figure 9-1, right: the two buttons “Jacomi” and “Janne” in the lower part of the screen).

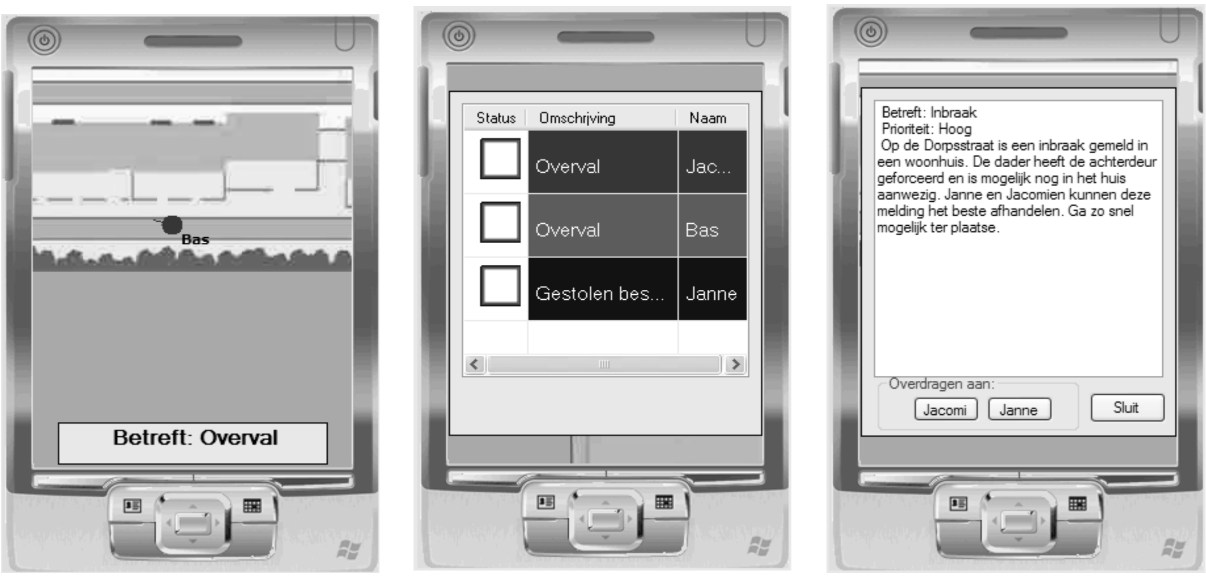


Figure 9-1. PDA screenshots showing a message summary (“*Betreft: Overval*”) presented over the map (left) , the task list (middle) and the transfer screen (right).

9.3 Evaluation method

To evaluate the claims on benefits and costs of task allocation support, teams use the prototype system in a simulated surveillance task. The task allocation system communicates incident messages to the team. Context events, such as roadblocks, cars on fire or people in need of assistance, occur in the vicinity of one team member and are not known to the task allocation system or the other team members. Due to an event, the task allocation advice for new incidents during these events is incorrect. Task performance, (shared) situation awareness and trust are compared between two conditions: a task allocation system that provides partially reliable advice and a system without advice.

9.3.1 Participants

Eighteen teams of three student participants (30 male, 24 female), aged between 18 and 29 years (M = 22.5, SD = 3.1) completed this study. They used desktop computers on a daily basis for internet browsing and email, but were inexperienced with mobile devices and with police work.

9.3.2 Experimental design and manipulation

A two-by-two mixed design was employed, with *task allocation advice* (with or without) as a between-subjects condition and *incident type* (without or with context event) as a within-subjects factor. Half of the teams received advice in the incident messages (advice condition). The other half received incident messages without advice (no advice condition) and had to follow the task allocation decision rules on proximity, priority and availability themselves.

Teams performed the same scenario-based, time-paced surveillance task as in the previous study (see section 8.3.2). The scenario consisted of five high priority incidents and seven low priority incidents (see Figure 9-2). During six of the twelve incidents, an event occurred in the vicinity of one team member. These events were timed in such a way that the next incident occurred during handling of the event (E1-6 in Figure 9-2). In the advice condition, these events caused the task allocation advice to be incorrect for six of the twelve incidents. Both conditions used the same experimental scenario. We used a between-subjects setup because the context events could only be “spontaneous” once for the participants. The number of male and female team members was balanced between conditions and teams.

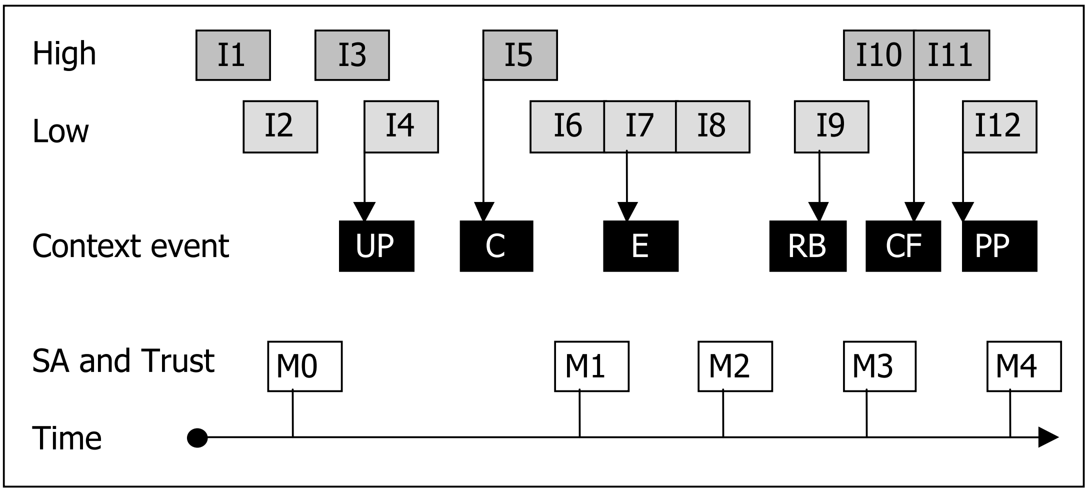


Figure 9-2. Scenario timeline indicating high and low priority incidents (I1–12), context events (E1-6) and SA and trust measurements (M0-4).

9.3.3 Measures

Effectiveness of the surveillance task was measured as the *number of incidents* handled correctly and the *number of decision errors* on task allocation. An error was counted for each team member who did not follow the decision rules on availability, proximity and priority. Efficiency was measured as *response time* to the message in seconds and *incident handling time* in seconds. As in the previous chapter, the number of *communication utterances* on task allocation between team members was counted per incident.

Five times during the scenario, the environment paused (cf. "freezing technique"; Salmon et al., 2006), and *situation awareness* (SA) was assessed with one question for each level of SA. Example questions are "On which street are you now?" (perception level), "Who is now available for a low priority incident?" (comprehension level) and "If an incident occurs on the Kerkstraat in two minutes, which colleague(s) should handle this?" (projection level). *Shared SA* was measured as the number of answers that all three team members answered correctly (maximum 15). *Trust* in the system was measured at the same time on a 7-point rating scale. After the experiment, judgments on *team performance* were collected with 7-point rating scales on coordination, misunderstanding, task allocation and efficiency.

9.3.4 Apparatus and procedure

The same workstation setup (see Figure 9-3) and procedure was used as in the previous experiment (see sections 8.3.5 and 8.3.6). The experiment took three hours to complete; half an hour per scenario. Prior to the experimental task, participants were trained on two

Figure 9-3. Workstation showing the environment and the prototype on two screens (left) and the setup of the three workstations (right).

practice scenarios consisting of four incidents, with or without advice depending on the condition. They received feedback on their performance. No events occurred during training and they were not told that events could happen or that the advice from the system could be incorrect.

9.4 Results

Task performance data was analyzed with a mixed model ANOVA, with between-subjects factor *advice* (with advice (A) vs. without advice (WA)) and within-subjects factor *incident type* (without event vs. with event). Bonferroni post-hoc analyses were performed on this data. Data on SA and trust questionnaires was analyzed with repeated measures ANOVA for independent groups and shared SA scores and rating scales were analyzed with independent group t-tests.

9.4.1 Surveillance task effectiveness

Teams handled on average 4.0 of the six incidents without event correctly compared to 2.4 of the six incidents with event. This difference was significant ($F(1,16) = 20.8$; $p < 0.001$). Advice did not seem to make a significant difference, as both with and without advice the number of incidents handled correctly was the same ($M_A = 6.7$; $M_{WA} = 6.6$; $F(1,16) = 0.01$; $p = 0.90$). No significant interaction effect was found between advice and incident type.

There was no significant difference in number of decision errors on task allocation with or without advice ($M_A = 7.0$; $M_{WA} = 6.7$; $F(1,16) = 0.05$; $p = 0.82$) or between incidents without event ($M = 3.0$) and with event ($M = 3.8$; $F(1,16) = 6.25$; $p = 0.13$). No interaction

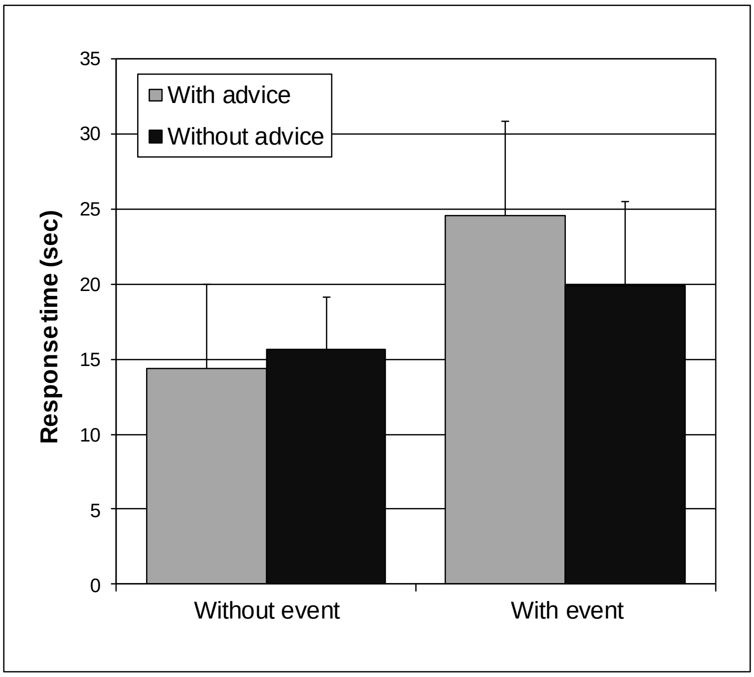


Figure 9-4. Mean response time to incidents without and with event.

effect was found between advice and incident type, and post-hoc analyses showed no significant effects.

9.4.2 Surveillance task efficiency

Response time to incident messages showed no overall significant difference with or without advice. As expected, response time to incidents with event (M = 22.2 s) was significantly higher than without event (M = 15.0 s; $F(1,16) = 19.0$; $p < 0.001$). Interestingly, an interaction effect approached significance ($F(1,16) = 3.25$; $p = 0.09$); when an event occurred, response time with advice (24.6 s) was longer than without this advice (19.9 s), an increase of about 25% on average (see Figure 9-4). Post-hoc analysis showed this effect to be significant ($p < 0.005$). Thus, response time was significantly longer due to incorrect advice.

A similar pattern emerged for incident handling time (IHT). IHT did not differ significantly with or without advice. IHT was higher for incidents with event (M = 165 s) than without event (M = 151 s; $F(1,16) = 4.37$; $p = 0.05$). Importantly, a significant interaction effect was found ($F(1,16) = 4.70$; $p < 0.05$); when an event occurred, IHT was longer with advice (M_A = 174 s) than without (M_{WA} = 155 s; see Figure 9-5). Post-hoc analyses showed however that this effect was not significant. Thus, a trend towards longer incident handling time with incorrect advice was observed.

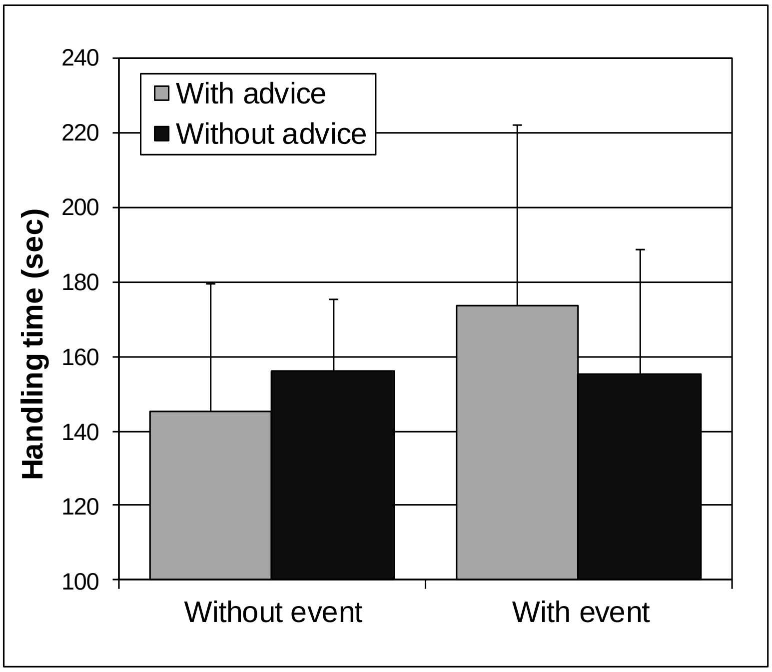


Figure 9-5. Mean handling time of incidents without and with event.

9.4.3 Communication

The number of communication utterances on task allocation showed no significant difference with or without advice. When an event occurred, significantly more utterances (M = 6.4) were needed than during incidents without event (M = 4.7; $F(1,16) = 15.7$; $p = 0.001$). In addition, a significant interaction effect was found ($F(1,16) = 8.76$; $p < 0.01$). As expected, without advice, team members communicated more during incidents with event (M = 7.1) than without event (M = 4.1). Remarkably, with advice, the number of utterances remained approximately the same during both incidents without and with event (see Figure 9-6). This last result was opposite to what was hypothesized.

So, no main benefits of partially reliable advice, compared to no advice, were found in the task performance data. The data partly confirm the first hypothesis regarding costs: incorrect task allocation advice due to events slows response time and incident handling time, but did not cause more decision errors or communication utterances.

9.4.4 (Shared) situation awareness and trust

Answers to the situation awareness (SA) questions were analyzed per condition and SA level. The number of correctly answered SA questions did not differ significantly with or

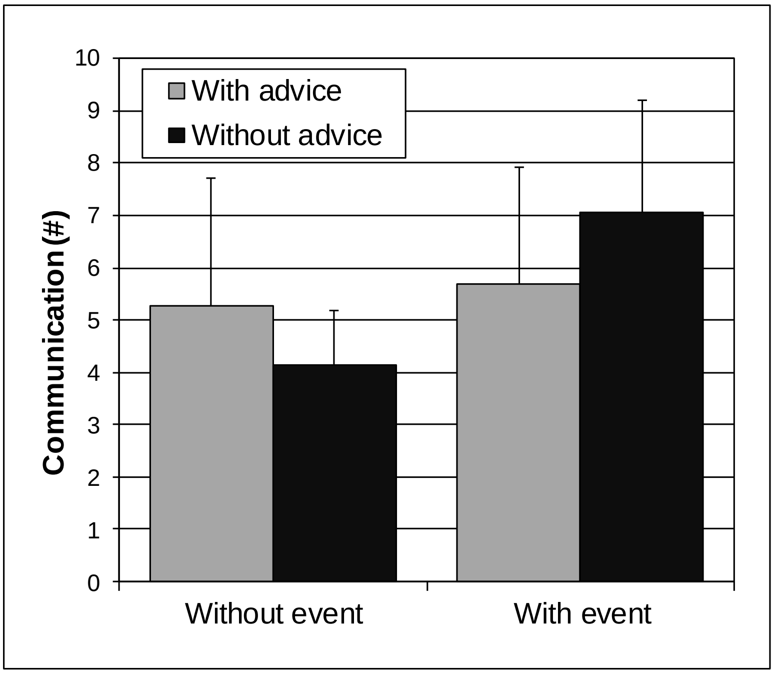


Figure 9-6. Team communication utterances during incidents without and with event.

without advice ($F(1,52) = 0.17$; $p = 0.67$). No interaction effect was found between condition and SA level ($F(2,104) = 0.16$; $p = 0.85$).

Shared situation awareness within the teams was overall low, as judged by the answers to the SA questions that all three team members answered correctly. The shared SA scores ranged from 0 to 6 (maximum possible score 15). So the best teams only provided right answers on less then half of the SA questions. There was no significant difference between shared SA with or without advice ($t(16) = 0.51$; $p = 0.61$). Correlations between shared SA scores and decision errors showed a significant negative correlation of -0.51 ($p < 0.05$) with shared SA Level 3 (projection). Thus, low SA level 3 is correlated with more decision errors. This partly confirms our second hypothesis.

Trust in the system was rated high overall (M = 5.5 out of 7; SD = 1.0). Trust did not differ significantly between the advice condition ($M_A = 5.6$) compared to the condition without advice ($M_{WA} = 5.3$; $F(1,51) = 0.96$; $p = 0.33$). This is opposite what was expected based on the third hypothesis.

9.4.5 Subjective judgments

Two of the four rating scales on team performance showed significant differences. In the condition without advice, participants agreed more with the statement “There was little misunderstanding within the team” than participants with advice ($M_{WA} = 5.6$, $M_A = 4.8$; $t(52)$

= -2.43; $p = 0.02$). In addition, they agreed more with the statement "Our team did not need to reallocate the tasks often" than participants with advice ($M_{WA} = 5.6$, $M_A = 4.7$; $t(52) = -2.52$; $p = 0.01$). Advice did not affect subjective ratings of team coordination or efficiency. So, team members stated that advice caused somewhat more misunderstanding and task reallocation within the team than no advice.

9.5 Discussion

As context-aware support systems are only partially reliable, we need to empirically establish the benefits and costs of design solutions for these systems. The final study in this thesis investigated the effects of partially reliable task allocation support on team task performance, shared situation awareness and trust. We stated claims of expected benefits and costs and tested these in the same experimental setup and surveillance task used in Chapter 8. In the current study however, spontaneous context events occurred during surveillance, such as people in need of assistance or roadblocks. These events caused task allocation advice to be incorrect for half of the incidents (50% reliability). Compared to no advice, partially reliable advice slows task performance (response time), but accuracy (decision errors) stays the same. We found no effects of advice on team communication, shared situation awareness and trust.

These results partly support the first hypothesis on task performance. Overall, partially reliable advice did not lead to worse task performance than no advice, even though the reliability level was lower than 70% (Wickens et al., 2007). We found costs of incorrect advice on task performance, specifically as longer response time and a strong trend towards longer incident handling time, compared to no advice. This corresponds to costs found in earlier work (Sarter et al., 2001). Incorrect advice did not lead to more decision errors (presumably due to a ceiling effect) or to more communication between team members.

The second hypothesis on shared situation awareness was also partly supported. Shared SA was not found to be different with partially reliable advice compared to no advice. We found a significant correlation between low shared SA (at the projection level) and a high number of decision errors. A true effect of incorrect advice on shared SA may be masked by the low scores on shared SA; some teams did not answer a single question correctly. Presumably, the surveillance task was too easy, in the sense that it did not allow for much performance degradation when people did not maintain shared SA. The costs of sub-optimal allocation may be limited in our study (e.g. traveling time between incidents was not very long), but these can be expected to increase in real-life settings.

The third hypothesis on trust was not supported; results demonstrate equally high levels of trust with partially reliable advice and without advice. Although the advice was sometimes incorrect, team members readily accepted it, leading to a high number of decision errors (compared to a “ground truth”). This is further confirmed by team communication; incorrect advice did not lead to more communication on task allocation. In addition, trust in the system remained high. This indicates that people did not ascribe failures to the system but to themselves or they did not think to criticize the system. This finding is consistent with the first stage of trust development, when system trust is often high (Muir, 1987).

This study has a number of limitations. We had to manipulate the conditions (advice / no advice) *between* teams. Presumably, this caused not all expected effects to be found. Furthermore, the reliability level of the system was set at 50%, as we wanted to gather enough data points (six events in a surveillance scenario). Of course, the actual reliability and associated benefits and costs of automated task allocation depend on how often these “spontaneous” events occur in real life. It seems reasonable to assume reliability will be higher in real life. Using field studies combined with cognitive modeling simulations will provide insight into the net effects of these events on task allocation (Kieras & Santoro, 2004). Future research should also include multiple levels of system reliability, in order to differentiate the effects of reliability on task performance. Finally, as trained students participated in this study, our findings must be tentatively applied to the police domain. Police professionals would be more experienced with communication protocols and more critical of a faulty task allocation system. This might help explain the contrary findings on team communication and trust.

This study addresses a fundamental limitation of context-aware mobile support systems: how should these systems deal with imperfections in the support they aim to provide? These imperfections will be unavoidable constraints in designing and using context-aware mobile technology. In the present study, the high number of decision errors shows that task allocation support in mobile surveillance is necessary. In addition, the study suggests that the costs of incorrect advice (in increased response and handling time) outweigh the benefits (in task effectiveness) in situations requiring a rapid response, such as during incidents with high time-pressure. For these situations, the risk of the system

Table 9-3. Validated and refined claims on benefits and costs of task allocation.

Core function: Providing automated task allocation advice on incident handling to distributed mobile team members.			
Objective	An optimal distribution of incidents over team members, based on availability, proximity and priority.		
Claims	*Benefits*	*When advice is correct:* ✓ **Less team communication** (Chapter 8) ✓ Less decision errors (Chapter 8) ✓ More efficient navigation (Chapter 8) *Trust not measured in Chapter 8*	*When advice is incorrect:* *No effects found on situation awareness*
	Costs	*When advice is correct:* *Situation awareness not measured in Chapter 8*	*When advice is incorrect:* × **Longer response time** (Chapter 9) × **Longer handling time** (Chapter 9) *No effects found on decision errors, communication and trust*

providing incorrect advice and the team having to reallocate tasks (entailing further costs) must be avoided.

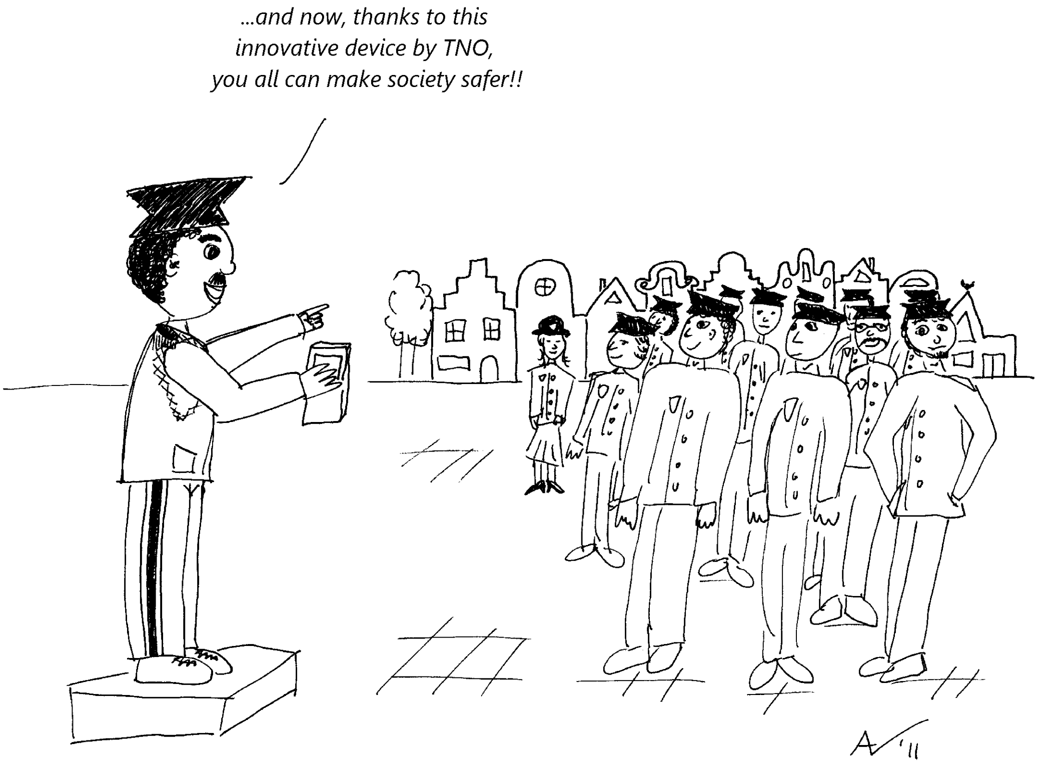
When we combine the experimental results on task performance from Chapters 8 and 9, we can validate and refine the claims on benefits and costs of task allocation support. Table 9-3 presents the validated claims (in bold) and claims that are partly supported. As found in Chapter 8, correct task allocation advice had the benefit of reducing team communication. The claims of less decision errors and more efficient navigation were partly supported. The costs of correct advice remain undecided as situation awareness was not measured in Chapter 8. In Chapter 9, we found that situation awareness is no different with incorrect advice than without advice. The claimed costs on incorrect advice are demonstrated as longer response time and a trend towards longer handling time. Task effectiveness (i.e. correctly handled incidents and decision errors), team communication and trust did not suffer costs as a result of incorrect advice.

These findings show a distinct speed-accuracy trade-off for task allocation advice. Correct advice leads to more effective, but not faster, task allocation. However, when advice is incorrect, people compensate for errors in the system and find an optimal task allocation themselves. This takes time, demonstrated by longer response and handling time of incidents. These validated claims have implications for the design of context-aware mobile support (CAMS) systems. Task allocation advice should be provided to a distributed team to facilitate decision making and reduce team communication. Designers should be aware that advice can become incorrect and that this slows task performance (speed-accuracy trade-off). When in a high time-pressure situation, advice should be used sparingly or should be accompanied by an estimation of confidence. This way, users can more appropriately decide to trust or disregard the advice. Alternatively, the design could facilitate easy reallocation of tasks (e.g. Gutwin et al., 2002). For example, users themselves can indicate their availability in a quick manner to the system (e.g. by means of a “not now!” button). Based on such user feedback, the system can revise and provide more accurate task allocation advice to the team members.

In conclusion, Part III of this thesis focused on developing team collaboration support for mobile professionals. In three empirical studies with police officers and trained (student) participants, we showed that team awareness and task allocation can be supported by a CAMS system. Specifically, Chapter 7 showed that a visual overview of team information combined with auditory communication improves team decision making and perception of messages from colleagues. Chapter 8 demonstrated that when CAMS provides advice on incident handling, this improves team communication and may improve decision making accuracy and navigation efficiency. However, advice from automated systems is at best partially reliable, for example due to context events. In Chapter 9, partially reliable advice was shown to significantly slow down task performance, while maintaining the same accuracy as in a situation without advice. No effects of partially reliable advice were found on shared situation awareness and trust in the system. The studies in Part III provide a sound basis for further evaluation of CAMS in operational settings. Future work should study multiple levels of system reliability and triangulate the benefits and costs identified in these studies with field evaluation including professional end-users.

PART IV

GENERAL CONCLUSION



10 Conclusions and Discussion

Work in mobile professional domains, such as the police domain, is challenged by limited means for mobile information exchange and inefficient team collaboration. Current mobile technology is not appropriately designed to meet the operational demands in these domains. Police officers have limited access to operational information and are not always aware of the status and availability of their colleagues. This holds for both the daily work of police surveillance as well as for large-scale incidents. The beach riot at Hoek van Holland provides a compelling example (see Table 10-1 and section 1.3.3). The Regional Support Team (RST) members were unaware whether their emergency call was received and who was tasked to help them. The commander of the RST could not reach the on-scene commander, while the on-scene commander was not fully aware what was going on and could not give orders. Human-computer interaction issues with the new C2000 communication system played an important role in most of these problems (Muller, Rosenthal, Zannoni, Ferwerda, & Schaap, 2010). So, ill-designed mobile technology in the police domain can result in decision errors in task allocation, longer response time and limited team awareness (Baber et al., 2001; Stijnman, 2004; Marcus et al., 2006; Streefkerk et al., 2008).

The main argument in this thesis is that mobile technology has to adapt information presentation to the mobile use context and Human Factors (attention, workload and individual characteristics). This will provide better support for work in these domains than non-adaptive systems (Abowd et al., 1997b; Dey et al., 2000; Shafer et al., 2001). We expect such context-aware mobile support (CAMS) systems to improve task performance (decision accuracy, response and handling time) and optimize the user experience (trust, preferences and acceptance). The central question in this thesis is: *Which features of context-aware mobile systems can support team task performance and optimize the user experience, specifically for mobile information exchange and team collaboration in professional domains?* Taking the police work environment as application domain, the research in this thesis focused on designing and evaluating CAMS system features, specifically to support mobile information exchange and team collaboration. Following a Situated Cognitive Engineering method (Neerincx et al., 2008), we conducted a field study and five experiments to test the effects of context-aware notification and task allocation support on task performance and the user experience. The main findings were that context-aware notification and task allocation support improved decision making accuracy but slowed down response time to incident messages. People trusted and preferred the CAMS system more than a non-adaptive system, leading to a positive user experience.

The rest of this chapter summarizes the main findings and implications, while the separate research questions from each chapter will be answered in the summary. Section 10.1 elaborates how the main findings contribute to CAMS requirements for mobile information exchange and team collaboration support. Methodological and theoretical explanations for the findings and directions for further research are discussed in sections 10.2 and 10.3. This thesis concludes with the practical implications for implementing CAMS systems in mobile professional domains in section 10.4.

Table 10-1. Excerpts from COT report "Beach riot at Hoek van Holland – 22 August 2009".

".... the two Regional Support Team (RST) members in the field, threatened by the crowd, do not know whether their call for emergency assistance has been received…" (p. 70)
"The RST commander tries to contact the on-scene commander to ask permission to secure the train station. When no audible response comes, he takes matters into his own hands." (p. 78)
"In the command vehicle, it is unclear for the on-scene commander what has happened, because a radio connection to the RST could not be established." (p. 74)
"… it is unclear for police officers using the C2000 communication system how to act when the system places their call on hold or gives no response." (p. 90)

10.1 CAMS requirements

The police work domain and support analysis in Chapter 2 identified challenges to operational demands in police surveillance and specified the CAMS system concept, core functions and features to support these operational demands (see also Table 2-4). Challenges to mobile information exchange consisted of limited access to operational information, interruptions and prioritisation. For example, an incident message with low priority may distract police officers from handling a high priority incident. Challenges to team collaboration consisted of limited team awareness and inefficient task allocation. For example, because police officers work dispersed over a city, they do not always know which colleagues are available for assistance on a particular incident. The CAMS system concept was envisioned as a networked mobile device that incorporated **context-aware notification**: adapting the appearance and timing of an incident message (notification style) based on use context factors. For example, when handling a high priority domestic violence incident, CAMS postpones a lower priority message about vandalism until the police officer is available again. Three context-aware notification features were designed and evaluated; adapting notification styles to 1) the use context (location), 2) the user (workload, activity), and 3) the task (message priority). To address the challenges to team collaboration, the CAMS concept incorporated two **task allocation support** features, consisting of 1) a visual overview of team information and 2) task allocation advice based on context factors (availability, proximity and priority). For example, for the domestic violence incident, CAMS selects the two nearest, available team members and advises them to handle this incident. In each separate study in this thesis, CAMS features were validated on outcome and process measures.

10.1.1 Context-aware notification to support mobile information exchange

How the three context-aware notification features addressed mobile information exchange was evaluated in Part II of this thesis. Chapter 4 studied location-based notification in the field, while Chapters 5 and 6 focused on how appearance and timing (notification styles) of incident messages should be adapted to workload, activity and message priority. The main findings are presented in Table 10-2, by listing validated benefits and costs of these features.

1. *Location awareness*: As work in the police domain takes place across a variety of locations, adapting information presentation to the current location seems a logical

design concept. When a police officer is close to a "point-of-interest" such as the street where a perpetrator lives, he is notified of this information. This way, location-based notification facilitates access to relevant operational information, wherever necessary. The police field study in Chapter 0 showed that a location-based notification system did improve officers' awareness of incidents. However, improvements in decision making and outcome measures could not be established, due to the field setting. The user experience of location-based notification was not positive: many notifications created unwanted interruptions during surveillance. We concluded that location-based notification is valuable but filtering is needed based on more context factors than location alone.

2. *Workload and activity awareness*: Unwanted interruptions by incident messages during police surveillance distract users from the environment. This feature adapted the information density and timing of incident messages to the current activity or availability of the user. For example, when an officer is not directly available, a message outline is presented instead of a full message. This improved decision making accuracy on incident messages slightly (Chapter 5), helped users to find more targets in the environment and decreased intrusiveness of messages (Chapter 6). Users were positive about this feature. Contrary to what we expected, adapting notifications to activity led to a slight increase in response time to messages (Chapter 6). Presumably, users found it hard to understand the adaptive system behavior. We concluded that activity awareness helped users to better focus attention on incoming messages while maintaining awareness of the environment.
3. *Priority awareness*: In the police domain, one incident may have higher priority than the other, entailing the need for prioritisation. This feature adapted notification salience and timing of incident messages to the priority of the incident. For example, a high priority message was presented with a more salient sound than a low priority message. This facilitated recognition of message priority and improved decision making accuracy, e.g. by ensuring individual officers were working on the task with the highest priority (Chapter 6). Users preferred this feature over a non-adaptive feature. However, priority awareness comes at the cost of postponing lower priority incidents, which slowed incident handling time overall (Chapter 5).

Concluding, context-aware notification to support mobile information exchange showed some nuanced findings. Context-aware notification had a positive influence on decision making accuracy, but on the other hand had the tendency to slow down response time. This speed-accuracy trade-off might be inherent to the domain: it is most important to handle the right incident accurately and a welcome bonus when handled quickly. Unfamiliarity with adaptive system behavior also played a role in slowing response time. Still, we found that users prefer context-aware notification over non-adaptive notification because it decreased interruption and message intrusiveness. Especially in a domain with a multitude of interruptions, less interruption is a welcome improvement.

Table 10-2. Context-aware notification to support mobile information exchange.

Challenge	Core functions	Feature	Validation of benefits and costs
Limited access to operational information	*Information support:* Adapt information presentation to current location.	**Location awareness**: Present pop-up message with operational information when near a "point-of-interest".	✓ Location-based notification increased awareness of operational information in the field (Chapter 4) × Notification based on location alone led to increased interruption (Chapter 4)
Interruptions	*Adaptation to attention, Notification*: Adapt notification styles to activity / availability of team members.	**Workload and activity awareness**: Depending on activity, postpone messages or present messages concise (e.g. icon or outline instead of full message).	✓ Context-aware notification improved task performance effectiveness slightly (less decision errors) (Chapter 5) ✓ Context-aware notification limited intrusiveness of messages, while maintaining awareness of environment (Chapter 6) × Context-aware notification slowed response time to messages slightly (Chapter 6)
Prioritisation	*Task switching support, Notification* Adapt notification styles to priority of incident.	**Priority awareness**: Depending on priority, postpone messages or present messages in interface with less salient visual and auditory signals (e.g. softer sounds).	✓ Context-aware notification improved task performance effectiveness slightly (less decision errors) (Chapter 5) ✓ Users could recognize incident priority better through context-aware notification (Chapter 6) × Context-aware notification slowed message handling time because low priority messages are postponed (Chapter 5)

10.1.2 Task allocation support for team collaboration

How the two task allocation support features address challenges in team collaboration was evaluated in Part III of this thesis. In a virtual environment, Chapter 7 studied the team information overview, while Chapters 8 and 9 studied how (partially reliable) task allocation advice should be presented to a (police) team. The main findings are presented in Table 10-3, by listing validated benefits and costs of these features.

1. *Overview of team information*: When working together across distances, team members have to be aware of each others' location and activities (team awareness) and share an understanding of the operational situation (shared situation awareness or SSA). We expected that providing a visual, geographical overview of team information on a mobile device would support team awareness and collaboration. For example, indicating the location of incidents and the availability of team members with coloured icons on a map overview. Police officers in the field preferred a visual overview of incidents because it led to better awareness of incidents (Chapter 4). In the lab, team members with the overview made more accurate decisions in who to approach for assistance. However, the overview did not make complex decisions easier, contrary to our expectation (Chapter 7). In addition, the overview did not improve team awareness or SSA (Chapters 7 and 9). On the contrary, team awareness (perception) was supported more by auditory rather than visual support. We concluded that combining a visual overview of team information with auditory communication supports team awareness and improves decision making.
2. *Task allocation advice:* One of the reasons no effect on team awareness was found may be that integrating and comparing information on a small mobile screen is cognitively challenging. CAMS can support the task allocation process by advising which team member can handle which incident best, based on context factors. For example, CAMS advises the two nearest available team members they can handle an incident best. In Chapters 8 and 9, we found this advice led to a speed-accuracy trade-off: a trend towards more accurate decisions was found, but advice did not speed up the processes of task allocation and incident handling as we expected. Moreover, incorrect advice actually slowed task performance, as people compensated for the unreliability of the system (Chapter 9). Still, advice reduced the need to communicate and participants' trust in the system was high. Importantly, three-quarters of the police participants preferred the advice (Chapter 8). No effect of advice was found on SSA within the team.

We concluded that task allocation advice might be a valuable feature, but, depending on the system reliability, might be less suited for high-time pressure situations.

Concluding, the team overview and task allocation advice are valuable and necessary features to include in CAMS systems. When studied in a virtual environment, the two features showed (relatively small) benefits on surveillance task accuracy, specifically by improving decision making. However, incorrect advice from the system slowed down response time, demonstrating the important relation between reliability and task performance. The user experience of advice was positive; people trusted and preferred advice from the system. Team awareness and shared situation awareness did not improve with or without support features. The task allocation support features should be further evaluated in operational settings to assess how costs in the lab relate to costs in the field (such as over-reliance on the advice and increased navigation time). Furthermore, the reliability of these features in the field, especially the visual overview, should receive further attention (see also 10.4).

Table 10-3. Task allocation support for team collaboration.

Challenge	Core functions	Feature	Validation of benefits and costs
Limited team awareness	*Information support:* Provide a visual overview of team information.	**Team overview**: Present a visual, geographical map in the interface with team information such as identity, location and availability.	✓ More accurate decisions with team overview (Chapter 7) ✓ Higher awareness of incidents with team overview (Chapter 4) ✗ Team overview did not make complex choices easier, nor improved team awareness or shared situation awareness (Chapter 7)
Inefficient task allocation	*Task switching support, Notification*: Provide task allocation advice based on context factors.	**Task allocation advice**: Present incident message with advice on who can handle incident best, based on proximity, availability and incident priority.	✓ Advice led to less team communication (Chapter 8) ✓ Advice might increase decision accuracy and navigation efficiency (Chapter 8) ✗ Unreliable advice slowed task performance. (Chapter 9)

10.2 Methodology

In this thesis, Situated Cognitive Engineering (SCE) helped to systematically design and evaluate the proposed CAMS system in three ways. First, the work domain and support analysis identified and elaborated the operational demands in this domain (mobile information exchange and team collaboration) and identified opportunities to support these demands. This resulted in a concise and coherent list of CAMS core functions. Second, the SCE method allowed us to specify design rationales based on which selected core functions were implemented into features. For example, the rationale behind context-aware notification was captured by the notification matrix and decision rules in Chapter 6. Third, evaluation in the SCE method helped us to validate and refine claims on the features. This allowed us to identify trade-offs between the benefits and costs of using CAMS systems (see the discussions of Chapter 6 and 9 for examples) and draw balanced conclusions.

In designing and evaluating CAMS systems for the police domain, a trade-off exists between employing naïve (student) participants and police officers (Chapter 3). Police officers possess valuable domain knowledge and experience, ensuring that system designs are appropriate for the domain. On the other hand, their experience and routine may make them reluctant to innovative work procedures and they may be limited available as well as costly to employ. Naïve participants can be employed to demonstrate that innovations (such as context-aware notification and task allocation support) provide benefits in task execution. Naturally, these benefits have to be validated with police end-users. Despite the fact that police officers were only limited available, they participated in key research activities in this thesis; in the focus groups to determine CAMS core functions, in interviews to establish the experimental scenarios and in the field and lab study as participants.

How well our research results generalize to mobile professional domains (ecological validity) has been an important methodological consideration. The evaluation framework (Chapter 3) increased ecological validity by tuning evaluation measures to criteria from the application domain (e.g. by measuring response time to an incident message in the experiments). Each experimental setting (Wizard-of-Oz lab study, mobile surveillance and synthetic task environment) captured rules and core task features from the police domain and naïve participants were trained on the evaluation tasks. For example, the surveillance task in Chapter 6 included navigation and notification and participants were trained on decision making rules. This increases the chance that evaluation results generalize to the police domain. As an example, we found similar results in handling time and preferences from Chapters 5 and 6 (with trained participants) and Chapter 8 (with end-users). Thus, employing naïve participants in addition to police officers in our evaluation process generates appropriate design proposals for the police domain.

The decision to employ a virtual environment or synthetic task environment (STE) in Part III of this thesis influenced our research results. An STE provides a controlled setting for accurate measures of performance and (S)SA in team research (Cooke et al., 2004). However, in Chapters 7, 8 and 9, we did not find effects of task allocation support on trust, mental effort and SSA. One reason might be that the surveillance task in our STE did not allow for enough performance degradation, e.g. the task did not require participants to maintain high levels of situation awareness. In addition, the participants in the studies in Chapter 7 and 9 were non-professionals. Another reason might be that participants experienced the STE differently from the real world. In comparative studies on navigation in a STE and in the real world, participants in the STE experienced worse navigation performance, decreased SA and increased mental effort (Smets, Neerincx, te Brake, Buurman, & van Oostendorp, in press). This negative influence of a STE could have masked actual effects on mental effort and SSA in our STE studies. It will be interesting and necessary to triangulate the results on team task allocation support with results from field experiments.

In conclusion, our methodology helped us to establish a requirements baseline of five CAMS features with a positive trade-off in task performance and user experience. As a result of using this iterative methodology, a substantial improvement over the state-of-the-art in location-based notification (Chapter 4) was achieved. The requirements baseline provides a starting point for further summative evaluation of CAMS systems' performance and reliability in field settings.

10.3 Implications for Human Factors in mobile HCI

CAMS systems have to be adequately geared towards Human Factors, such as attention, workload and individual differences. As stated in the introduction, support systems for the mobile professional domain have to deal with divided attention situations, interruptions and changes in workload within a team. Furthermore, interaction with a support system has to be understandable, trustworthy and optimized for individual differences (personalization).

To manage divided attention situations in the mobile domain, the current research adopted the awareness-interruption trade-off: balancing the utility of focusing attention on the environment or on an interruption (e.g. an incoming message) (McFarlane et al., 2002; McCrickard et al., 2003b; Horvitz et al., 2005; Bailey et al., 2006; Grandhi & Jones, 2010). We showed that the CAMS system could capture this utility in explicit, simple decision rules (based on context factors) and adapt notification styles of incident messages based on these rules. For example, when a user is not available, CAMS defers a low priority message to an appropriate moment. This context-aware notification led to lower message intrusiveness, slightly more accurate decisions and an appropriate focus of attention on high priority incidents. These positive effects confirmed results in earlier work (Adamczyk et al., 2004; Iqbal et al., 2008). On the other hand, we found that this approach slows down task performance overall, due to deferral of lower priority tasks. This is contrary to what other researchers found (Bailey et al., 2008; Salvucci & Bogunovich, 2010). A possible solution is to let CAMS facilitate *multitasking*, for example by letting users check the sender or subject of a message, while walking to an incident location (Nagata, 2003; Adamczyk et al., 2004) (McCrickard et al., 2003b; Grandhi et al., 2010; Salvucci et al., 2010). Of course, the costs of this solution (e.g. whether it leads to increased interruption) should be assessed in operational situations (Wickens, 1987; Nagata, 2003; Roda, 2010). In conclusion, this research contributes empirical evidence that context-aware notification helps to manage divided attention situations and interruptions, by assisting users to focus attention either on the environment or on a notification.

To manage changes in workload within a team, task allocation support from CAMS helped team members to allocate incidents appropriately based on context factors. Advice from the system improved task allocation accuracy and communication within the team, confirming earlier findings (Rovira et al., 2007; Johansson, Trnka, & Granlund, 2007; Salas et al., 2008). However, task allocation support did not mitigate mental effort, as was found earlier (Bailey et al., 2008). Presumably, the amount of effort saved in the task allocation process in these lab settings was small compared to executing the surveillance task as a whole, especially for trained (student) participants (Alty, 2003). For experienced police officers in real life situations, we expect CAMS support to decrease mental effort more. Furthermore, the expected negative effect of task allocation support on SSA (team members can “lazily” rely on the advice) was not found (Gorman et al., 2005; Wickens et al., 2007). It may be that a true effect of over-reliance did not occur because team members did not need to maintain SSA well enough for this surveillance task, as mentioned in the methodology section. The question remains how much SSA and TA is actually needed for task allocation (Nova, Girardin, Molinari, & Dillenbourg, 2006; Stanton et al., 2006). Further research should focus on the effects of over-reliance on SSA and TA in task allocation in real life situations.

Our participants found the adaptive behavior from CAMS sometimes hard to understand (Chapter 6 and 8). Why certain interface adaptations occurred (e.g. using different notification styles) was not explained to users, to avoid increasing cognitive load. However, this decreased the understandability of the system (Bellotti & Edwards, 2001; Lim & Dey, 2010). Furthermore, users had no control over the adaptations (Shneiderman, 1997; Rovira et al., 2007); CAMS determined the most appropriate notification style for the user, even if this meant deferring a notification. Low control combined with low understandability could negatively impact trust in the system (Muir, 1987; Bellotti et al., 2001; Lim et al., 2010). We found no negative effects on trust; moreover, trust in advice from the system remained high even when it was partially reliable. Still, trust was measured with a single rating scale after only a couple of hours of interaction. In our view, a more in-depth analysis is required of how understandability and control influence trust in CAMS systems over time.

Taking into account individual differences personalizes the interaction with CAMS systems for individual users (Benyon, 1993; Kramer, Noronha, & Vergo, 2000; Brusilovsky, 2001). The current research found encouraging results to include individual differences in CAMS. Specifically, the field study in Chapter 4 showed that police roles (e.g. emergency officer or district officer) influence the relevance of location-based notifications. Expertise and mode of transport influenced the appropriateness of task allocation within a team (Chapter 7). Finally, age was found to affect the use of a task allocation advice; younger teams profited more from advice than older teams (Chapter 8). Further research should assess whether personalized support based on role, expertise, mode of transport and age optimizes the user experience of CAMS systems.

10.4 Implementing CAMS in mobile professional domains

CAMS systems are designed specifically for the mobile professional domain, which is characterized by dynamic operational situations, distributed actors and resources, limited communication bandwidth and high variations in time pressure and workload. There is a strong need for communication and flexible task coordination within distributed, ad-hoc teams. In addition to the police domain, CAMS support can be extended to similar mobile professional work environments such as fire fighting (Jiang et al., 2004b; Chen, Li, & Wang, 2010), crisis management and emergency response (Cai, 2005; Smets et al., 2007; Sapateiro et al., 2008; Straus, Bikson, Balkovich, & Pane, 2010; Te Brake & van der Kleij, 2010) and military patrol (Adkins, Kruse, & Younger, 2002; Smets, Streefkerk, & Neerincx, 2010).

CAMS systems support work processes in these domains by adapting notification and task allocation based on context factors such as location, availability and priority. This will benefit mobile professionals in three ways: by reducing the number of interruptions, supporting team awareness and supporting the task allocation process. First, CAMS makes communication less interruptive in all these domains by considering context factors such as location, user activity and message priority. For example for military group commanders, a low priority message about supplies will be presented less salient than a high priority message about hostile activity (Smets et al., 2010). Second, CAMS optimizes team awareness within (ad-hoc) teams. In emergency response, when police officers and fire-fighters have a shared team information overview of each others activities, collaboration at an incident site will improve (Convertino et al., 2005; Carroll et al., 2007). Third, CAMS allocates tasks to team members based on context factors such as current location or mission. For example, CAMS can advise rescue workers at an incident site which high-risk areas to evacuate first. Or CAMS can advise three military team members to check a possible disturbance in a village near them, while informing the rest of the team of this action. The present research indicates that supporting team awareness and task allocation will improve team collaboration accuracy and ensure the highest priority tasks are attended.

Implementing CAMS systems in these domains requires dealing with technological constraints in sensor technology, context representation and reliability. First, sensor technology (such as GPS location tracking) to enable context-awareness is not always accurate and robust. Some sensors are difficult to use in real operational contexts, because they require contact with the body (e.g. motion and gaze detection) (Kern et al., 2003; Fogarty et al., 2005a). As sensing technology matures, the inaccuracies will diminish and practical operability will increase. Second, the context representation in CAMS is necessarily a simplified, and not entirely accurate, representation of the actual context. To improve the richness of this representation, it needs to include more context factors such as role, expertise, means of transportation and history of user actions. Users themselves can provide input for the context representation as well, such as police officers indicating which briefing focal points are most important this month. To improve the accuracy of the context representation, combinations of modelling techniques have to be sought, such as ontologies, statistical reasoning and agent technology (Dey et al., 2001; Fogarty et al., 2005b; Petersen, Cassens, Kofod-Petersen, & Divitini, 2008; Riboni & Bettini, 2009). Third, due to unpredictable events in the world and information unknown to the system, CAMS can at best only be partially reliable. The reliability should be appropriately reflected in the design of CAMS, to ensure appropriate levels of trust in the system. One solution is to present confidence information to users about the reliability of the support (Antifakos, Kern, Schiele, & Schwaninger, 2005). This way, users can more accurately decide whether or not to trust the system (Van der Kleij et al., 2009; Gunawan, Alers, Brinkman, & Neerincx, 2010).

In a broader perspective, implementation of CAMS systems requires a change in the work organization in mobile professional domains. Mobile technology provides workers with better capabilities in information processing and decision making while on the move (Alberts & Hayes, 2003). Mobile devices, because they display adaptive behavior and provide suggestions about work distribution, become virtual team members or e-partners (Marsh et al., 2000; Neerincx et al., 2006). In the police domain, this will cause a shift from central control by the emergency room to distributed task allocation by a team of police officers (cf. "power to the edge"). Both police officers as well as management need to accept new mobile technology and its influence on work organization (Venkatesh et al.,

Table 10-4. Future incident handling with a context-aware mobile support system.

"... the two Regional Support Team members, threatened by the crowd, call for emergency assistance. CAMS determines that four police officers are nearby and advises them to move to the two team members in need of assistance. Auditory directions combined with location icons on a visual display help these colleagues to move quickly. One of the RST members sees on his CAMS display the four closing in on them and feels safer now they will be assisted. After meeting up, the six team members retreat towards the train station, while the crowd follows them…"

"In the command vehicle, reports come in that the crowd of hooligans is moving to the train station. The on-scene commander sees on his team overview that the RST commander and two team members are available and allocates the task of securing the train station to them. As CAMS notifies the RST commander, the on-scene commander can turn his attention to other matters. CAMS advises the RST commander with a salient notification message to move directly to the train station, and informs him that his two team members will join him…"

2000). In addition, privacy and security issues arise when introducing context-aware systems (Ackerman, Darrell, & Weitzner, 2001; Karat, Karat, Brodie, & Feng, 2005). Mobile devices used on the street need to secure access to police databases and storage of sensitive information (such as the identity of suspects). Also, sensing and storing (private) information based on user preferences, actions and history should be managed well. Which information is stored in the user model? How long will it be accessible and who has access?

In conclusion, CAMS systems are no "silver bullet" for all information exchange and team collaboration challenges in mobile professional domains. They are designed to support one or more specific tasks in these domains, both in the dynamics of daily use as well as in large-scale incidents. How the core functions and features are implemented and how well the technological constraints are met, determines to what extent users accept, trust and use these systems (Marcus et al., 2006; McFarland & Hamilton, 2006; Streefkerk et al., 2008). In the future, incidents such as the beach riot at Hoek van Holland will still occur. In our view, CAMS systems can provide benefit for the teams of professionals who have to handle these incidents, by facilitating mobile information exchange and optimizing team collaboration. This way, the vision of future work in the police domain in Table 10-4 will become reality.

SUMMARY

Work in mobile professional domains, such as the police domain, is challenged by limited means for mobile information exchange and inefficient team collaboration. Police officers on surveillance have limited access to operational information on the street and are not always aware of the current status and availability of their colleagues. Current mobile technology is not appropriately designed to meet these operational demands. This increases the risk of distraction, information overload, lowered team awareness and inefficient distribution of tasks within a team. To mitigate these risks, the design of support systems for mobile police officers must take into account the mobile use context, technological constraints and human cognitive capacities. The research in this thesis is concerned with designing and evaluating context-aware mobile support (CAMS) systems for mobile information exchange and team collaboration. The central question is: *Which features of context-aware mobile systems can support team task performance and optimize the user experience, specifically for mobile information exchange and team collaboration in mobile professional domains?*

Taking mobile police surveillance as the application domain, the first part of this thesis elaborated our Situated Cognitive Engineering design and evaluation method. Chapter 2 focused on which task, context and user factors are relevant in the police domain to design context-aware support. First, a work domain and support analysis identified challenges and support opportunities in mobile police surveillance. We found that mobile information exchange was challenged by limited access to operational information, interruptions and unclear prioritization. Team collaboration was challenged by limited team awareness and

inefficient task allocation. Next, to address these challenges, the concept of context-aware mobile support (CAMS) was specified. We envisioned a networked mobile device that has knowledge about the current context of use, such as location of incidents and police officers, availability of officers and incident priority. To support mobile information exchange, CAMS should notify police officers of operational information using context-aware notification: adapting the appearance and timing of incident messages (notification style) based on location, availability and priority. Furthermore, to support team collaboration between distributed team members, CAMS should provide a geographical overview of team information and task allocation advice on who can best handle an incident, also based on location, availability and priority.

Chapter 3 focused on which methods and measures should be used to evaluate context-aware mobile support systems. We employed a test-plan that specified a field study and five experiments to evaluate the effects of CAMS system features on surveillance task performance and the user experience. The methods and techniques used in these experiments (Wizard-of-Oz lab study, mobile experiment, and evaluation in synthetic task environments) were selected and combined based on a framework of evaluation constraints. In an iterative process of assessment and refinement, CAMS system features were evaluated on domain-relevant outcome and process measures.

Context-aware notification to support mobile information exchange

We studied context-aware notification in Part II of this thesis. Chapter 4 studied the effects of a location-based notification system on work process efficiency and effectiveness and the user experience. A field study with police officers showed that using police officers' current location to trigger notification presentation (i.e. *location-based* notification) increased officers' awareness of incidents. However, the current implementation caused too much interruption due to too many or non-relevant notifications. We concluded that intelligent filtering of notifications is needed, based on more context factors than location alone.

Next, we wanted to know whether adapting notification appearance to message priority and the use context (*context-aware* notification) actually improved surveillance task performance. Chapter 5 studied the effects of notification style appearance on task performance, notification intrusiveness and the user experience. In a Wizard-of-Oz lab experiment, trained participants performed a simulated surveillance task by watching videos, while using a notification system prototype. When message priority was high, the prototype used salient notification signals and when used in a high workload context, it presented more condensed information (e.g. a message summary). The results showed that in high workload situations, context-aware notification improved the number of targets found slightly compared to uniform notification. This came at a cost of an increase in message handling time, because users postponed reading low priority messages until they had time to do so. Still, context-aware notification was generally preferred although participants mention it was sometimes hard to understand.

The results from the first experiment showed that *when* notifications are presented within an ongoing task influences task performance. Chapter 6 studied the effects of a context-aware notification system, which adapts notification timing and appearance to message priority and user activity, on task performance and the user experience. We redesigned the notification system prototype from Chapter 5 and conducted a mobile experiment to investigate the effects of different notification styles (full message, postpone, indicator or adaptive). The adaptive notification style tuned notification timing to user activity and notification appearance to message priority. Trained student participants performed a building walkthrough while the prototype notified them of occurring incidents. The results showed that full messages led to high message intrusiveness, more interruption and an intermediate number of decision errors. Postponing all messages led to the highest number of decision errors, while indicators still caused interruption and were sometimes overlooked or forgotten. The adaptive notification style maintained awareness of incident messages, leading to a low number of decision errors, without decreasing awareness of the environment. This in turn led to a high number of targets found. The drawback was that this style came at the cost of increased response time.

Summarizing the three studies in the second part of this thesis, we demonstrated that location-based notification in the field led to improved awareness of incidents. This was echoed by two lab studies, in which context-aware notification caused more targets to be found in high workload situations, slightly less decision errors, and lower notification intrusiveness compared to uniform notification. The cost of context-aware notification is an increase in response time and message handling time. No effects of context-aware notification were found on mental effort. Thus, we conclude that context-aware notification improves awareness of the environment, leads to less interruption and makes users' decisions better, but not faster or less effortful.

Task allocation support for team collaboration

Part III of this thesis focused on team awareness and task allocation support for team collaboration. Team awareness is an important prerequisite for efficient collaboration. Factors that negatively influence team awareness are geographical distribution, changing locations and changing activities of team members. Furthermore, extracting team information from auditory communication alone (cf. the common radio transceiver) is cognitively demanding and will negatively impact team awareness. Accordingly, we wanted to know whether a visual overview of team information could support team awareness (staying aware of status and activities of team members) and decision making (when asking team members for assistance). Chapter 7 studied the effects of a visual or auditory presentation modality for team information on team awareness and decision making accuracy. We compared a visual overview of team information and communication with auditory team communication only. A lab study required individual, trained participants to decide which (virtual) team member to ask for assistance, based on the context factors location, availability, mode of transport and expertise. The results showed that a visual overview improved decision making accuracy and that auditory support improved perception of communication messages. No effects were found on mental effort. From this, we can conclude that providing a complementary visual overview to auditory team communication is an appropriate option to improve both team awareness and decision making.

The study on team awareness showed decision accuracy improved, but did not focus on whether decisions were made faster. We expected task allocation advice (advising which team member can handle which incident best) to have a positive effect on speed of decision making. Chapter 8 studied the effects of providing task allocation advice, adapted to location, availability and incident priority, on team task performance, communication and decision making. A prototype task allocation system was designed for a mobile device and evaluated with professional end-users. This context-aware prototype provided a geographical overview of team information (i.e., status and location of team members and incidents). Based on proximity, availability and priority rules, the system selected appropriate team members and advised them on incident handling via appropriate notification styles. In a synthetic task environment (a virtual city), teams of three police officers performed a surveillance task with this prototype, compared to a prototype without advice. The results showed that with task allocation advice, teams made less decision errors and travelled less distance. Also, team members communicated less on task allocation. The advice did not make teams faster nor had an effect on their experienced mental effort. From these results, we can conclude that providing advice on task allocation improves decision accuracy, but does not increase decision speed in mobile police surveillance.

Because context information may be missing or outdated, automated task allocation systems will never be 100% reliable. Chapter 9 studied the effects of partially reliable task allocation support on task performance, situation awareness and trust in the system. We conducted a final lab study with the prototype from the study in Chapter 8. In the same virtual city environment, teams of trained student participants handled incidents while using the prototype system that provided task allocation advice to half of the teams and no advice to the other half. Context events, such as encountering a panicked person or a roadblock, caused this advice to be incorrect for 50% of the incidents (partial reliability). Partially reliable advice increased response time and handling time of incidents but did not lead to more decision errors. Situation awareness in the team was generally low and users' trust in the system generally high; both were not influenced by partially reliable advice. This led us to conclude that, depending on the reliability, automated task allocation may be less appropriate when time-pressure is high.

Summarizing the three studies in the third part of this thesis, they showed that team collaboration can be improved by a visual overview of team information, complementary to auditory communication. This overview increases choice accuracy in selecting who to approach for assistance, but does not improve team awareness per se, nor mental effort. This improvement in decision making is echoed by results from the task allocation experiments. With task allocation advice, less decision errors were made, but teams were not faster in their decision making. Similarly, when task allocation support failed, not more decision errors were made, but response time and handling time increased even further. We can conclude that team awareness and task allocation support generally improve the accuracy, but not the speed of decision making in mobile police surveillance.

In conclusion, the research in this thesis focused on context-aware mobile support (CAMS) systems for police surveillance. Our iterative methodology helped to systematically design CAMS system features and evaluate them with trained participants and police end-users. Support for mobile information exchange was designed as context-aware notification, based on location, activity and priority. Support for team collaboration was designed as a team overview and task allocation advice. Evaluation of these support features in different settings found a speed-accuracy trade-off: the features improved decision making accuracy but slowed down response time and handling time of incidents. People trusted and preferred the CAMS system more than a non-adaptive system, indicative of a positive user experience. Further research should address technological constraints in CAMS systems (e.g. accuracy and reliability of context representation) and focus on summative evaluation of CAMS systems in operational settings. This way, CAMS systems can provide benefit for teams of mobile professionals by facilitating mobile information exchange and optimizing team collaboration.

SAMENVATTING

Werk in mobiele professionele domeinen, zoals het politiedomein, wordt negatief beïnvloed door beperkte middelen voor informatie-uitwisseling en inefficiënte teamsamenwerking. Politieagenten op surveillance hebben op straat beperkt toegang tot databases en informatie omtrent incidenten en zijn niet altijd op de hoogte van de huidige status en beschikbaarheid van collega's. De huidige mobiele technologie is niet optimaal ontworpen om aan de operationele eisen te voldoen. Dit verhoogt het risico op informatieoverlast, ongewenste afleiding, beperkt teambewustzijn en inefficiënte verdeling van taken binnen een team. Om deze risico's te voorkomen, dient het ontwerp van een mobiel ondersteuningssysteem voor politieagenten rekening te houden met de gebruikscontext, de technologische mogelijkheden en menselijke cognitieve capaciteiten. Dit proefschrift richt zich op het ontwerpen en evalueren van contextafhankelijke mobiele ondersteuning (*context-aware mobile support* of CAMS) voor informatie-uitwisseling en teamsamenwerking. De centrale vraag is: *Welke kenmerken van contextafhankelijke mobiele ondersteuning kunnen de taakprestatie en gebruikerservaring verbeteren, specifiek voor mobiele informatie-uitwisseling en teamsamenwerking?*

Het eerste deel van dit proefschrift beschrijft onze *Situated Cognitive Engineering* aanpak, toegepast op het taakdomein van politiesurveillance. Hoofdstuk 2 richt zich op welke taak-, context- en gebruikersfactoren relevant zijn in het politiedomein voor het ontwerp van CAMS systemen. Een domeinanalyse van de werkprocessen in politiesurveillance toonde kansen en uitdagingen voor het ontwerp van deze ondersteuning. Bijvoorbeeld dat mobiele informatie-uitwisseling negatief beïnvloed wordt door beperkte toegang tot operationele informatie, interrupties en onduidelijke prioriteitsstelling. Of dat teamsamenwerking negatief wordt beïnvloed door beperkt teambewustzijn en inefficiënte taakverdeling. Om deze uitdagingen aan te gaan, specificeerden we het CAMS systeemconcept. CAMS werd voorgesteld als een mobiel apparaat, opgenomen in een netwerk, dat kennis heeft over de gebruikscontext, zoals de locatie van incidenten en politieagenten, beschikbaarheid van agenten en de prioriteit van incidenten. Om mobiele informatie-uitwisseling te ondersteunen en interruptie te verminderen, moet CAMS politieagenten attenderen op operationele informatie door contextafhankelijke notificatie. Dit houdt in dat de presentatie en timing van incidentberichten (de stijl van attendering) aangepast worden aan locatie, beschikbaarheid en prioriteit. Verder moet CAMS teamsamenwerking tussen teamleden op afstand ondersteunen. Hiervoor biedt CAMS een visueel geografisch overzicht van teaminformatie en advies over taakverdeling (welk teamlid het beste een incident kan afhandelen), gebaseerd op locatie, beschikbaarheid en prioriteit.

Hoofdstuk 3 richt zich op welke methoden en maten gebruikt moeten worden om CAMS systemen te evalueren. Er werd een testplan opgesteld van een veldstudie en vijf experimenten om de effecten van CAMS systemen op taakprestatie en de gebruikerservaring te evalueren. We kozen de methoden en technieken in deze studies (Wizard-of-Oz laboratorium studie, mobiel experiment, en evaluatie in virtuele omgevingen) op basis van een raamwerk van evaluatie beperkingen. In een iteratief proces werden CAMS prototypes geïmplementeerd, beoordeeld op domeinrelevante uitkomst- en procesmaten en vervolgens verfijnd.

Contextafhankelijke attendering om informatie-uitwisseling te ondersteunen

Deel II van dit proefschrift richt zich op contextafhankelijke attendering. We onderzochten de effecten van een locatie-gebaseerd attenderingssysteem op efficiëntie en effectiviteit van het politiewerk en de gebruikerservaring van agenten in Hoofdstuk 4. Een veldstudie met dertig politieagenten toonde aan dat door attenderingen dichtbij een incidentlocatie te presenteren (*locatie-gebaseerde* attendering), agenten beter op de hoogte zijn van incidenten. Echter, het systeem presenteerde soms te veel of irrelevante attenderingen. We concludeerden dat intelligente filtering van attenderingen nodig is op basis van meerdere contextfactoren.

Om dit te bereiken wilden we eerst weten of het aanpassen van de visuele en auditieve presentatie van een attendering op de prioriteit en de gebruikscontext (*contextafhankelijke* attendering) de taakprestatie verbetert en de storing door berichten vermindert. Hoofdstuk 5 beschrijft de studie naar de effecten van attenderingsstijl op taakprestatie, storendheid van de attendering en de gebruikerservaring. In een Wizard-of-Oz lab experiment lieten we twintig getrainde deelnemers een gesimuleerde surveillancetaak uitvoerden door het kijken naar politievideo's, terwijl een prototype attenderingssysteem hen berichten aanbood. Wanneer de prioriteit van een bericht hoog was, gebruikte het prototype een opvallende attenderingsstijl (bijv. een hard geluid). Wanneer het in een situatie van hoge werklast gebruikt werd, presenteerde het prototype samengevatte informatie (bijv. een kernwoord uit het bericht). De resultaten lieten zien dat contextafhankelijke attendering leidde tot meer gevonden targets uit de video's tijdens hoge werklast situaties. Ook werden de berichten als minder storend ervaren ten opzichte van uniforme attendering. Deelnemers waren echter langer bezig met de berichten, omdat ze het lezen van berichten met lage prioriteit uitstelden. Er werd geen effect gevonden op mentale werkbelasting. Deelnemers hadden een voorkeur voor contextafhankelijke attendering, ook al vond een klein aantal de attenderingsstijlen moeilijk te begrijpen.

De resultaten van het vorige experiment lieten zien dat de taakprestatie ook bepaald wordt door de *timing* van een attendering. Om dit te bestuderen, werd het prototype uit Hoofdstuk 5 herontworpen om de timing van attenderingen aan te passen aan de activiteit van gebruikers en de presentatie aan te passen aan berichtprioriteit. In Hoofdstuk 6 voerden we een mobiel experiment uit om de effecten te meten van verschillende attenderingsstijlen (volledig bericht, uitgesteld bericht, icoon of adaptief) op de taakprestatie en de gebruikerservaring. We trainden tweeëndertig deelnemers op het prototype en lieten ze een mobiele surveillance uitvoeren door een gebouw terwijl het prototype hen attendeerde op incidenten. De resultaten lieten zien dat volledige berichten als storend werden ervaren, en leidden tot meer interruptie en een gemiddeld aantal beslisfouten. Het uitstellen van alle berichten leidde tot het hoogste aantal beslisfouten, terwijl iconen nog steeds leidden tot interruptie en soms werden gemist of vergeten. Met de adaptieve stijl bleven gebruikers op de hoogte van berichten, wat leidde tot een laag aantal beslisfouten, zonder dat berichten hun aandacht afleidden van de omgeving, te zien aan het hoge aantal gevonden targets. Een minpunt was dat deze stijl leidde tot een verhoogde reactietijd op berichten.

Samenvattend, de eerste studie in het tweede deel toonde aan dat locatie-gebaseerde attendering in het veld het bewustzijn van incidenten verbetert. Dit positieve resultaat werd ondersteund door twee lab experimenten waarin contextafhankelijke attendering zorgde voor meer gevonden targets, minder beslisfouten en minder storende berichten vergeleken met uniforme attendering. Contextafhankelijke attendering gaat echter ten koste van de reactiesnelheid en afhandelingtijd van berichten. We kunnen dus concluderen dat contextafhankelijke attendering het omgevingsbewustzijn van gebruikers verbetert, interruptie vermindert en beslissingen verbetert, maar niet versnelt.

Ondersteuning voor taakverdeling om teamsamenwerking te verbeteren

Deel III van dit proefschrift richt zich op ondersteuning voor teambewustzijn en taakverdeling om teamsamenwerking te verbeteren. Teambewustzijn is een belangrijke voorwaarde voor efficiënte samenwerking. Factoren die teambewustzijn negatief beïnvloeden zijn geografische verspreiding, veranderende locaties en veranderende activiteiten van teamleden. Daarnaast is het bijhouden van teaminformatie via een auditief kanaal (bijv. via de portofoon) cognitief belastend en dit kan teambewustzijn verder beperken. Daarom wilden we nagaan in hoeverre het hebben van een visueel overzicht van teaminformatie teambewustzijn verbetert (het beter op de hoogte blijven van status en activiteiten van teamleden) en beslissingen kan verbeteren (bij het vragen van het juiste teamlid om assistentie). Hoofdstuk 7 beschrijft de studie naar de effecten van visuele of auditieve presentatie van teaminformatie op teambewustzijn en besluitvorming. We vergeleken een visueel overzicht van teaminformatie en communicatie met auditieve communicatie alleen. In een lab studie lieten we zesentwintig getrainde, individuele deelnemers beslissen welk (virtueel) teamlid ze om assistentie zouden vragen, gebaseerd op de contextfactoren: locatie, beschikbaarheid, vervoer en expertise. De resultaten lieten zien dat het visuele overzicht leidde tot betere beslissingen maar dat auditieve communicatie zorgde dat deelnemers de berichten beter konden waarnemen. Er werd geen effect gevonden op mentale werkbelasting. Op basis hiervan lijkt het hebben van een visueel overzicht gecombineerd met auditieve communicatie een goede oplossing om zowel teambewustzijn als besluitvorming te ondersteunen.

De studie naar teambewustzijn toonde dus aan dat juistheid van beslissingen verbeterde, maar richtte zich niet op de snelheid van besluitvorming. We verwachtten dat het geven van advies over taakverdeling (welk teamlid het beste welk incident kan afhandelen) op basis van locatie, beschikbaarheid en prioriteit een positieve invloed heeft op beslissnelheid. Hoofdstuk 8 beschrijft de studie naar de effecten van dit advies op taakprestatie, communicatie en besluitvorming van een team van politieagenten. Een prototype adviessysteem werd ontworpen op een mobiel apparaat en geëvalueerd met acht teams van drie agenten. De interface van dit prototype toonde een geografisch overzicht van teaminformatie (d.w.z. de status en locatie van teamleden en incidenten). Als er een incident plaatsvond, selecteerde het prototype de juiste teamleden volgens regels over nabijheid, beschikbaarheid en prioriteit en berichtte hen over het incident in de juiste attenderingsstijl. In een virtuele stadsomgeving voerden dezelfde teams een surveillancetaak uit met dit prototype en met een prototype zonder advies. De resultaten lieten zien dat teams met advies minder beslisfouten maakten en minder afstand aflegden. Ook hoefde er met advies minder gecommuniceerd te worden over taakverdeling. Het advies maakte de teams echter niet sneller, noch had het effect op mentale werkbelasting. We concludeerden dat advies over taakverdeling tot betere, maar niet snellere beslissingen leidt in mobiele politiesurveillance.

Omdat informatie uit de omgeving niet altijd beschikbaar of up-to-date is, zal een geautomatiseerd adviessysteem nooit 100% betrouwbaar zijn. Hoofdstuk 9 beschrijft de studie naar de effecten van gedeeltelijk betrouwbaar advies op taakprestatie, situationeel bewustzijn, en vertrouwen in het systeem. We voerden een laatste labstudie uit met het herontworpen prototype uit Hoofdstuk 8. In dezelfde virtuele stadsomgeving lieten we achttien teams van getrainde studenten incidenten afhandelen, waarbij de helft van de teams advies kreeg en de andere helft niet. Onverwachte gebeurtenissen tijdens de taak, zoals een persoon in paniek of een obstakel, maakten dat het advies incorrect was voor de helft van de incidenten (gedeeltelijke betrouwbaarheid). De resultaten toonden aan dat het gedeeltelijk betrouwbare advies de reactietijd en afhandelingtijd van incidenten verhoogde, maar niet leidde tot meer beslisfouten. Situationeel bewustzijn in het team was over het algemeen beperkt en het vertrouwen in het systeem hoog; er werd op beiden geen effect gevonden van advies. Op basis van deze resultaten concludeerden we dat, afhankelijk van de betrouwbaarheid, geautomatiseerd advies mogelijk minder geschikt zal zijn voor situaties met hoge tijdsdruk.

Samenvattend toonden de drie studies in het derde deel van dit proefschrift aan dat teamsamenwerking verbeterd kan worden door een visueel overzicht van teaminformatie, aangevuld met auditieve communicatie. Dit overzicht verbeterde de juistheid van teambeslissingen, maar niet het teambewustzijn perse, noch de werkbelasting. Deze verbetering in besluitvorming werd ondersteund door de resultaten van de adviesexperimenten. Met advies over taakverdeling werden minder beslisfouten gemaakt, maar waren teams niet sneller in hun besluitvorming. Sterker, wanneer het advies fout was werden niet meer beslisfouten gemaakt, maar vertraagde de reactietijd en taaktijd nog meer. We kunnen concluderen dat ondersteuning voor teambewustzijn en taakverdeling over het algemeen de juistheid, maar niet de snelheid van beslissingen verbetert in mobiele politiesurveillance.

Conclusie

Dit onderzoek richtte zich op contextafhankelijke mobiele ondersteuning (CAMS) systemen voor politie surveillance. Onze iteratieve aanpak hielp ons om systematisch de kenmerken van CAMS te ontwerpen en te evalueren met getrainde deelnemers en politieagenten. Ondersteuning voor mobiele informatie-uitwisseling werd ontworpen als contextafhankelijke attendering, gebaseerd op locatie, activiteit en prioriteit. Ondersteuning voor teamsamenwerking werd ontworpen als een overzicht van teaminformatie en advies over taakverdeling. Evaluatie van deze ondersteuningsvormen met verschillende methoden liet een genuanceerd beeld zien: de besluitvorming op incidenten verbeterde maar de reactietijd en afhandelingtijd verhoogde. Men vertrouwde CAMS en verkoos het boven een niet-adaptief systeem, wat wees op een positieve gebruikerservaring. Verder onderzoek zal zich moeten richten op de technologische uitdagingen van een CAMS systeem (bijv. de accuraatheid en betrouwbaarheid van de context representatie) en op summatieve evaluatie van dit systeem in operationele situaties. Op deze manier kunnen CAMS systemen teams van mobiele professionals effectief ondersteunen door het faciliteren van mobiele informatie-uitwisseling en het optimaliseren van teamsamenwerking.

ACKNOWLEDGEMENTS

It is done, it is over. And what a sweet feeling it is! I don't think anything in my life up to now has taken as much determination as finishing this thesis. It is often sa(i)d that carrying out PhD research is a solitary and arduous endeavor. My PhD certainly was not easy and I admit doubting whether I would finish it. However, looking back over the past six years I see I've developed just because of the hard times. Crisis as an opportunity for growth. I was fortunate to receive support in various forms (moral, financial and loving) from the following people, which helped me to carry on and finish this thesis.

First and foremost, my two supervisors Mark en Myra have provided the scaffolding that allowed me to build this work. Mark, your scientific view, critical feedback and never-ending questions have given this thesis its quality. Myra, your practical solutions and empathic abilities assisted me on a day-to-day basis. Thank you both for helping me to develop myself as a scientist.

Olivier, Marc, Peter-Paul, Tjerk and Nanja, my fellow PhD's and roommates at TNO and Delft: thank you for being there and sharing the burden. Three down, three to go! Jasper, Kim, Maaike, Aletta, and all colleagues at TNO Soesterberg: thanks for your valuable input and assistance in experiments, great "*Personeelsvereniging*" activities or for just partying in Utrecht. A special thanks to Bert Bierman for your never-ending support in programming the prototypes.

This research aimed to help police officers do a better job. However, this was only possible because of interesting discussions with, input from and participation of police professionals. Special thanks to Roy Mente and Peter Duin at the VTS Politie Nederland and

Elle de Jonge at police district Groningen. My gratitude goes to all participating police officers in the lab-experiments and the officers of district Korrewegwijk in Groningen who participated in the field evaluation. I hope this work will lead to improvements in the ICT systems you use.

Carrying out the research in this thesis was made possible by the MultimediaN program and funded by the Dutch Ministry of Economic Affairs. Thanks to Nellie Schipper, John Schavemaker and Marcel Worring for their involvement in MultimediaN. I want to thank the TNO organisation for employing and supporting me as a PhD. In 2007, I undertook a very interesting and memorable research visit to Virginia Tech in Blacksburg, USA. A warm thank-you to Scott, Shahtab, Jason, Laurian, Stacey, Edgardo and Doug for making me part of the VT family. I'll never pick up a Nerf-gun without thinking about you! Special thanks to the Fulbright Scholarship Program and the TNO Defence Research Scholarship Program for making this visit possible.

The loving support from my family, Fenna, Dick, Marieke en Job was very important to me. Thanks for your ongoing involvement, interest and compassion. My rowing team at the Tromp boatclub in Hilversum provided me with a physical outlet whenever I suffered from PhD stress. Hendrik, Floris and the whole Utrecht-group: thanks for making Utrecht the greatest place to live! A special thanks to Patrick for being a true friend and my "paranimf". Never before have I earned two new titles in two years: "Doctor" and "Father". Although becoming a father is like doing your PhD in reverse: first you get your title, and then the research starts… Anna and Jonah, I hope our "research" never ends. Your love means the world to me!

CURRICULUM VITAE

Jan Willem Streefkerk (born Sept. 1, 1978 in Heemskerk) received his secondary education (VWO) at the Gemeentelijk Gymnasium in Hilversum from 1990 until 1997. After a year at the Vrije Hogeschool in Driebergen, he studied Cognitive Psychology at Maastricht University. Specializing in Cognitive Ergonomics, he received his Masters' title in 2003. He conducted two successful internships at Groningen University and TNO Human Factors, focusing on traffic psychology and the role of emotion and trust in mobile user experiences. From 2003 to 2005, Jan Willem combined working as a Human Factors researcher at TNO with lecturing at the Hague University. His PhD research was conducted at TNO from 2005 until 2009, in cooperation with the VTS Police Netherlands. During this time, he was granted a Fulbright Scholarship for a research visit to Virginia Tech, USA. Currently, he is employed at the Perceptual and Cognitive Systems department at TNO in Soesterberg. His research focuses on the design of mobile support and netcentric operations for military and civil security professionals. Jan Willem lives in Utrecht with his partner Anna and their son Jonah.

References

Abowd, G., Atkeson, C., Hong, J. I., Long, S., Kooper, R., & Pinkerton, M. (1997a). Cyberguide: A Mobile Context-Aware Tourist Guide. *ACM Wireless Networks, 3,* 421-433.

Abowd, G., Dey, A., Orr, R., & Brotherton, J. (1997b). Context-awareness in wearable and ubiquitous computing. In *Proceedings of the First International Symposium on Wearable Computing* (pp. 179-180).

Abowd, G., Mynatt, E., & Rodden, T. (2002). The Human Experience. *Pervasive Computing, 1,* 48-57.

Ackerman, M., Darrell, T., & Weitzner, D. (2001). Privacy in context. *Human-Computer Interaction, 16,* 167-176.

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).

Adkins, M., Kruse, J., & Younger, R. (2002). Ubiquitous computing: omnipresent technology in support of Network Centric Warfare. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences* (pp. 551-559). IEEE.

Alberts, D. & Hayes, R. (2003). *Power to the Edge: Command... Control... in the Information Age*. DOD CCRP.

Alpert, S., Karat, J., Karat, C.-M., Brodie, C., & Vergo, J. (2003). User attitudes regarding a user-adaptive ecommerce web site. *User Modeling and User-Adapted Interaction, 13,* 373-397.

Alty, J. (2003). Cognitive Workload and adaptive systems. In E.Hollnagel (Ed.), *Handbook of Cognitive Task Design* (pp. 129-146). Mahwah: Lawrence Erlbaum Associates.

Anhalt, J., Smailagic, A., Siewiorek, D., Gemperle, F., Salber, D., Weber, S. et al. (2001). Toward Context-Aware Computing: Experiences and Lessons. *IEEE Intelligent Systems, 16,* 38-46.

Antifakos, S., Kern, N., Schiele, B., & Schwaninger, A. (2005). Towards Improving Trust in Context-Aware Systems by Displaying System Confidence. In *Proceedings of Mobile HCI 2005* (pp. 9-14).

Arciszewski, H., de Greef, T., & van Delft, J. (2009). Adaptive Automation in a Naval Combat Management System. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 39,* 1188-1199.

Aziz, Z., Anumba, C., Ruikar, D., Carrillo, P., & Bouchlaghem, N. (2006). Intelligent wireless web services for construction—A review of the enabling technologies. *Automation in Construction, 15,* 113-123.

Baber, C., Haniff, D., Sharples, M., Boardman, M., & Price, A. (2001). A requirements analysis of personal mobile computers for police officers. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics, Volume Six: Industrial Ergonomics, HCI, and Applied Cognitive Psychology* (pp. 21-28). Aldershot: Ashgate.

Bailey, B., Konstan, J., & Carlis, J. (2000). Measuring the effects of interruptions on task performance in the user interface. In *Proceedings of IEEE Int.Conference on Systems, Man and Cybernetics* (pp. 757-762).

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,* 1-28.

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.

Bardram, J. & Hansen, T. (2004). The AWARE Architecture: Supporting Context-Mediated Social Awareness in Mobile Cooperation. In *Proceedings of CSCW'04* (pp. 192-201). Chicago, IL: ACM.

Bartram, L., Ware, C., & Calvert, T. W. (2003). Moticons: detection, distraction and task. *International Journal of Human-Computer Studies, 58,* 515-545.

Begole, J. & Tang, J. C. (2007). Incorporating Human and Machine Interpretation of Unavailability and Rhythm Awareness Into the Design of Collaborative Applications. *Human-Computer Interaction, 22,* 1-2.

Bellotti, V. & Edwards, K. (2001). Intelligibility and accountability: human considerations in context-aware systems. *Human-Computer Interaction, 16,* 193-212.

Benyon, D. (1993). Accommodating individual differences through an adaptive user interface. In M. Schneider-Hufschmidt, T. Kühme, & U. Malinowski (Eds.), *Adaptive User Interfaces - Results and Prospects.* Amsterdam, North-Holland: Elsevier.

Bouwman, D., Haaker, T., & de Vos, H. (2008). Mobile Applications for Police Officers. In *Proceedings of the 21st Bled eConference.* Bled, Slovenia.

Brusilovsky, P. (2001). Adaptive Hypermedia. *User Modeling and User-Adapted Interaction, 11,* 87-110.

Cai, G. (2005). Extending Distributed GIS to Support Geo-Collaborative Crisis Management. *Geografic Information Science, 11,* 4-14.

Camp, P. J., Hudson, J. M., Keldorph, R. B., Lewis, S., & Mynatt, E. (2000). Supporting Communication and Collaboration Practices in Safety-Critical Situations. In *Proceedings of CHI 2000* (pp. 249-250).

Carroll, J. M. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.

Carroll, J. M. (2000). *Making use - scenario-based design of human-computer interactions.* Cambridge: MIT Press.

Carroll, J. M., Mentis, H., Convertino, G., Rosson, M. B., Ganoe, C. H., Sinha, H. et al. (2007). Prototyping Collaborative Geospatial Emergency Planning. In B. van der Walle, P. Burghardt, & C. Nieuwenhuis (Eds.), *Proceedings of ISCRAM 2007* (pp. 105-113). Delft, the Netherlands.

Carroll, J. M., Rosson, M. B., Convertino, G., & Ganoe, C. H. (2006). Awareness and teamwork in computer-supported collaborations. *Interacting with computers, 18,* 21-46.

Carroll, J. M., Neale, D. C., Isenhour, P. L., Rosson, M. B., & McCrickard, D. S. (2003). Notification and awareness: synchronizing task-oriented collaborative activity. *International Journal of Human-Computer Studies, 58,* 605-632.

CBS (2008). Integral Safety-monitor 2008 [Integrale Veiligheidsmonitor 2008, Landelijke rapportage]. The Hague, the Netherlands: Centraal Bureau voor de Statistiek.Available online in Dutch (November 7, 2009) at http://www.cbs.nl/nl-NL/menu/themas/veiligheid-recht/publicaties/publicaties/archief/2009/2008-integrale-veiligheidsmonitor-pub.htm. [On-line].

Chen, G. & Kotz, H. (2000). *A survey of context-aware mobile computing research* (Technical Report TR2000-381). Dartmouth.

Chen, Y., Li, D., & Wang, C. (2010). Map Synchronization and Alternatives Optimization for Firefighters Cooperative Decision Support in Ad Hoc Networks. *Journal of Networks, 5,* 39-46.

Cheverst, K., Davies, N., Mitchell, K., Friday, A., & Efstratiou, C. (2000). Developing a Context-aware Electronic Tourist Guide: Some Issues and Experiences. In *Proceedings of CHI 2000* (pp. 17-24).

Cheverst, K., Mitchell, K., & Davies, N. (2002). Exploring Context-aware Information Push. *Personal and Ubiquitous Computing, 6,* 276-281.

Convertino, G., Ganoe, C. H., Schafer, W. A., Yost, B., & Carroll, J. M. (2005). A multiple view approach to support common ground in distributed and synchronous geo-collaboration. In *Proceedings of CMV* (pp. 121-132).

Conway, A. & Kane, M. (2001). Capacity, Control and Conflict: An Individual Differences Perspective on Attentional Capture. In C. Folk & B. Gibson (Eds.), *Attraction, Distraction, and Action: Multiple Perspectives on Attentional Capture* (pp. 349-372). Amsterdam: Elsevier.

Cooke, N. & Shope, S. (2004). Designing a synthetic task environment. In S. Schiflett, L. Elliot, E. Salas, & D. Coovert (Eds.), *Scaled Worlds: Development, Validation, and Application* (pp. 263-278). Surrey: Ashgate.

Cremers, A., Engels, J., & Vlaskamp, F. (2003). *Service at a distance using a mobile terminal [Dienstverlening op afstand via een mobiele terminal].* Hoensbroek: IRV.

Cremers, A. & van der Pijl, D. (2005). *Individual travelinformation for users with a mental handicap. [Individuele reisinformatie voor personen met een verstandelijke beperking: Inventarisatie en raamwerk]* (Rep. No. TNO DV3 2005-C 005). Soesterberg, the Netherlands: TNO Defence, Security and Safety.

Cummings, M. (2004). Automation Bias in Intelligent Time Critical Decision Support Systems. In *Proceedings of AIAA Intelligent Systems.* Chicago, IL, USA.

Cutrell, E. B., Czerwinski, M., & Horvitz, E. (2000). Effects of Instant Messaging Interruptions on Computing Tasks. In *Proceedings of CHI 2000* (pp. 99-100).

Dahlback, N., Jonsson, A., & Ahrenberg, L. (1993). Wizard of Oz Studies: Why and How. In *Proceedings of the 1st international conference on Intelligent User Interfaces* (pp. 193-200). ACM Press.

De Greef, T., van Dongen, K., Grootjen, M., & Lindenberg, J. (2007). Augmenting Cognition: Reviewing the Symbiotic Relation Between Man and Machine. *Lecture Notes in Computer Science (LNCS), 4565,* 439-448.

Derryberry, D. & Reed, M. (2001). A Multidisciplinary Perspective on Attentional Control. In C. Folk & B. Gibson (Eds.), *Attraction, Distraction, and Action: Multiple Perspectives on Attentional Capture* (pp. 325-347). Amsterdam: Elsevier.

Dey, A. K. & Abowd, G. (2000). CybreMinder: A Context-Aware System for Supporting Reminders. *Lecture Notes in Computer Science (LNCS), 1927,* 172-186.

Dey, A. K., Abowd, G., & Salber, D. (2001). A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications. *Human-Computer Interaction, 16,* 97-166.

Dix, A., Finlay, J., Abowd, G., & Beale, R. (1998). *Human-Computer Interaction*. New York: Prentice Hall.

Dourish, P. & Bly, S. (1993). Portholes: Supporting awareness in a distributed work group. In *Proceedings of CHI 1993* (pp. 541-547). ACM Press.

Duh, H., Tan, G., & Chen, V. (2006). Usability Evaluation for Mobile Device: A Comparison of Laboratory and Field Tests. In *Proceedings of Mobile HCI '06* (pp. 181-186). Helsinki, Finland.

Eijk, R. v., Koning, N., Steen, M., & Reitsma, E. (2006). Developing a We-Centric Service for Mobile Police Officers to Support Spontaneous Communication. https://doc.freeband.nl/dscgi/ds.py/Get/File-67631 [On-line].

Endsley, M. (1988). Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 97-101).

Endsley, M., Bolte, B., & Jones, D. (2003). *Designing for Situation Awareness: An Approach to User-Centered Design*. New York: Taylor & Francis.

Endsley, M. R. & Garland, D. J. (2000). *Situation awareness analysis and measurement*. Mahwah, NJ: Erlbaum.

FBI (2008). Uniform Crime Report "2008 Crime in the US". Federal Bureau of Investigation, Washington DC. Available online (November 7, 2009) at http://www.fbi.gov/ucr/cius2008/index.html. [On-line].

Ferscha, A. (2000). Workspace awareness in mobile virtual teams. In *Proceedings of IEEE 9th International Workshops on Infrastructure for Collaborative Enterprises (WET ICE 2000)* (pp. 272-277). Gaithersburg, MD, USA.

Feteke, A. & Kun, A. (2008). Handheld computing in law enforcement; a pilot study. In *Proceedings of 2008 IET 4th International Intelligent Environments Conference.* Seattle, WA.

Fischer, G. (2001). User Modeling in Human-Computer Interaction. *User Modeling and User-Adapted Interaction, 11,* 65-86.

Fogarty, J., Hudson, S. E., Atkeson, C., Avrahami, D., Forlizzi, J., Kiesler, S. et al. (2005a). Predicting Human Interruptibility With Sensors. *ACM Transactions on Computer-Human Interaction, 12,* 119-146.

Fogarty, J., Ko, A. J., Aung, H. H., Golden, E., Tang, K. P., & Hudson, S. E. (2005b). Examining Task Engagement in Sensor-Based Statistical Models of Human Interruptibility. In *Proceedings of CHI 2005* (pp. 331-340).

Gajos, K., Everitt, K., Tan, D., Czerwinski, M., & Weld, D. S. (2008). Predictability and Accuracy in Adaptive User Interfaces. In *Proceedings of CHI 2008* (pp. 1271-1274). Florence, Italy: ACM.

Gaver, W. W. (1989). The SonicFinder: An Interface That Uses Auditory Icons. *Human-Computer Interaction, 4,* 67-94.

Gievska, S. & Sibert, J. (2004). A Framework for Context-Sensitive Coordination of Human Interruptions in Human-Computer Interaction. *Lecture Notes in Computer Science (LNCS), 3196,* 418-425.

Goodman, J., Brewster, S., & Gray, P. (2004). Using Field Experiments to Evaluate Mobile Guides. In *Proceedings of the 3rd Annual Workshop on HCI in Mobile Guides*. Glasgow, Scotland.

Gorman, J., Cooke, N., Pederson, H., Connor, O., & De Joode, J. (2005). Coordinated Awareness of Situation by Teams (CAST): Measuring Team Situation Awareness of a Communication Glitch. In *Proceedings of the HFES Annual Meeting* (pp. 274-277). HFES.

Grandhi, S. & Jones, Q. (2010). Technology-mediated interruption management. *International Journal of Human-Computer Studies, 68,* 288-306.

Grootjen, M., Neerincx, M., & Veltman, J. (2006). Cognitive task load in a naval ship control centre : from identification to prediction. *Ergonomics, 49,* 1238-1264.

Gunawan, L., Alers, H., Brinkman, W. P., & Neerincx, M. (2010). Effect of map sharing and confidence information in situation-map making. In W.P. Brinkman & M. Neerincx (Eds.), *Proceedings of ECCE 2010* (pp. 41-48). Delft, the Netherlands: ACM.

Gutwin, C. & Greenberg, S. (2002). A Descriptive Framework of Workspace Awareness for Real-Time Groupware. *Computer Supported Cooperative Work (CSCW), 11,* 411-446.

Hart, S. & Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North Holland.

Hofer, T., Schwinger, W., Pichler, M., Leonhartsberger, G., & Altmann, J. (2003). Context-Awareness on Mobile Devices - the Hydrogen Approach. In *Proceedings of the 36th Hawaii International Conference on System Sciences*. IEEE Computer Society.

Hollnagel, E. & Woods, D. (1983). Cognitive systems engineering: New wine in new bottles. *International Journal of Man-Machine Studies, 18,* 583-600.

Hong, J. I. & Landay, J. A. (2001). An infrastructure approach to context-aware computing. *Human-Computer Interaction, 16,* 287-303.

Höök, K. (2000). Steps to take before intelligent user interfaces become real. *Interacting with computers, 12,* 409-426.

Hopp, P. J., Smith, C. A., 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.), *Proceedings of the 10th International Conference on User Modeling* (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 Application. *Communications of the ACM, 46,* 52-59.

Ioimo, R. & Aronson, J. (2004). Police Field Mobile Computing: Applying the Theory of Task-Technology Fit. *Police Quarterly, 7,* 403-428.

Iqbal, S. T., Adamczyk, P. D., Zheng, X. S., & Bailey, B. P. (2005). Towards an Index of Opportunity: Understanding Changes in Mental Workload during Task Execution. In *Proceedings of CHI 2005* (pp. 311-320).

Iqbal, S. T. & Bailey, B. P. (2008). Effects of Intelligent Notification Management on Users and Their Tasks. In *Proceedings of CHI 2008* (pp. 93-102).

Iqbal, S. T. & Horvitz, E. (2007). Disruption and Recovery of Computing Tasks: Field Study, Analysis, and Directions. In *Proceedings of CHI 2007* (pp. 677-686).

Isbell, C. L., Omojokun, O., & Pierce, J. S. (2004). From devices to tasks: automatic task prediction for personalized applicance control. *Personal and Ubiquitous Computing, 8,* 146-153.

ISO 9241-11 (1998). *Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability* (Rep. No. 9241-11).

Jameson, A. (2001). Modeling Both the Context and the User. *Personal and Ubiquitous Computing, 5,* 29-33.

Jameson, A. (2002). Usability Issues and Methods for Mobile Multimodal Systems. In *Proceedings of the ISCA Tutorial and Research Workshop on Mulit-Modal Dialogue in Mobile Environments*. Kloster Irsee, Germany.

Jameson, A., Schäfer, R., Weis, T., Berthold, A., & Weyrath, T. (1999). Making systems sensitive to the user's time and working memory constraints. In M.T. Maybury (Ed.), *Proceedings of the 4th international conference on Intelligent User Interfaces* (pp. 79-86). New York: ACM.

Jiang, X., Chen, N. Y., Hong, J. I., Wang, K., Takayama, L., & Landay, J. A. (2004a). Siren: Context-aware Computing for Firefighting. In *Proceedings of Pervasive 2004*. Vienna, Austria..

Jiang, X., Hong, J. I., Takayama, L., & Landay, J. A. (2004b). Ubiquitous computing for firefighters: field studies and prototypes of large displays for incident command. In *Proceedings of CHI 2004* (pp. 679-686). Vienna, Austria: ACM Press.

Johansson, B., Trnka, J., & Granlund, R. (2007). The Effect of Geographical Information Systems on a Collaborative Command and Control Task. In B. van der Walle, P. Burghardt, & C. Nieuwenhuis (Eds.), *Proceedings of ISCRAM 2007* (pp. 191-201). Delft, The Netherlands.

Kaber, D. B. & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science, 5,* 113-153.

Karat, J., Karat, C.-M., Brodie, C., & Feng, J. (2005). Privacy in information technology: Designing to enable privacy policy management in organizations. *International Journal of Human-Computer Studies, 63,* 153-174.

Kellar, M., Inkpen, K., Dearman, D., Hawkey, K., Ha, V., MacInnes, J. et al. (2004). *Evaluation of Mobile Collaboration: Learning from our Mistakes* (Rep. No. CS-2004-13). Canada: Dalhousie University, Faculty of Computer Science.

Kerer, C. & Kirda, E. (2001). Layout, Content and Logic Separation in Web Engineering. In *Web Engineering* (pp. 135-147). Berlin / Heidelberg: Springer-Verlag.

Kern, N. & Schiele, B. (2003). Context-aware notification for wearable computing. In *Proceedings of IEEE International Symposium on Wearable Computers*. IEEE Computer Society.

Kieras, D. & Santoro, T. (2004). Computational GOMS modeling of a complex team task: Lessons learned. In *Proceedings of CHI 2004* (pp. 97-104). Vienna, Austria: ACM.

Kjeldskov, J. & Graham, C. (2003). A Review of Mobile HCI Research Methods. *Lecture Notes in Computer Science (LNCS), 2795,* 317-335.

Kjeldskov, J., Graham, C., Pedell, S., Vetere, F., Howard, S., Balbo, S. et al. (2005). Evaluating the usability of a mobile guide: the influence of location, participants and resources. *Behavior & Information Technology, 24,* 51-65.

Kjeldskov, J., Skov, M. B., Als, B. S., & Hoegh, R. T. (2004). Is It Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field. *Lecture Notes in Computer Science (LNCS), 3160,* 61-73.

Kjeldskov, J. & Stage, J. (2004). New techniques for usability evaluation of mobile systems. *International Journal of Human-Computer Studies, 60,* 599-620.

Klapp, S. (1987). Short-term memory limits in human performance. In P.A. Hancock (Ed.), *Human Factors Psychology* (pp. 1-28). North-Holland: Elsevier.

Kolasinski, E. M. (1996). *Prediction of Simulator Sickness in a Virtual Environment.* [Unpublished Doctoral Thesis] University of Central Florida, Orlando.

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.

Kramer, J., Noronha, S., & Vergo, J. (2000). A User-Centered Design Approach to Personalization. *Communications of the ACM, 43,* 45-48.

Krug, K., Mountain, D., & Phann, D. (2003). Webpark. Location-based services for mobile users in protected areas. *GeoInformatics, 6,* 26-29.

Kun, A., Miller III, W., & Lenharth, W. (2004). Computers in Police Cruisers. *Pervasive Computing, 3,* 34-41.

Lewis, M. & Jacobson, J. (2002). Game engines in scientific research: Introduction. *Communications of the ACM, 45,* 27-31.

Lim, B. & Dey, A. (2010). Toolkit to Support Intelligibility in Context-Aware Applications. In *Proceedings of UbiComp 2010* (pp. 13-22). Kopenhagen, Denmark: ACM.

Marcus, A. & Gasperini, J. (2006). Almost Dead on Arrival: A Case Study of Non-User-Centered Design for a Police Emergency-Response System. *Interactions, 13,* 12-18.

Marsh, S. & Meech, J. (2000). Trust in Design. In *Proceedings of CHI 2000: Extended Abstracts* (pp. 45-46). New York, NY: ACM.

Masthoff, J. (2002). The evaluation of adaptive systems. In N.V.Patel (Ed.), *Adaptive evolutionary information systems*. Hersey, PA: IDEA Group.

McCrickard, D. S., Catrambone, R., Chewar, C., & Stasko, J. (2003a). 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. (2003b). Attuning Notification Design to User Goals and Attention Costs. *Communications of the ACM, 46,* 67-72.

McCrickard, D. S., Chewar, C., Somervell, J., & Ndiwalana, A. (2003). A Model for Notification Systems Evaluation - Assessing User Goals for Multitasking Activity. *ACM Transactions on Computer-Human Interaction, 10,* 312-338.

McFarland, D. & Hamilton, D. (2006). Adding contextual specificity to the technology acceptance model. *Computers in Human Behavior, 22,* 427-447.

McFarlane, D. C. & Latorella, K. A. (2002). The scope and importance of human interruption in human-computer interaction design. *Human-Computer Interaction, 17,* 1-61.

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.

McGuirl, J. M. & Sarter, N. B. (2006). Supporting trust calibration and the effective use of decision aids by presenting dynamic system confidence information. *Human Factors, 48,* 656-665.

Mente, R. (2004). *Internal Analysis Notification Services CIP [Interne verkenning Attenderings services CIP]* (Rep. No. Internal report CIP). Driebergen, The Netherlands: CIP Dutch Police Organisation.

Miner, T. & Ferrano, F. R. (1998). The Role of Speed of Processing, Inhibitory Mechanisms, and Presentation Order in Trail-Making Test Performance. *Brain and Cognition, 38,* 246-253.

Miyata, Y. & Norman, D. (1986). Psychological issues in support of multiple activities. In D. Norman & S. Draper (Eds.), *User Centered Systems Design: New Perspectives on Human-Computer Interaction.* Hillsdale: Lawrence Erlbaum Associates.

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies, 27,* 527-539.

Muller, E. R., Rosenthal, U., Zannoni, M., Ferwerda, H., & Schaap, S. (2010). *Beach riots at Hoek van Holland.* The Hague, the Netherlands: COT Institute for Safety and Incident Management.

Munson, J. & Gupta, V. (2002). Location-based notification as a general-purpose service. In *Proceedings of the 2nd International Workshop on Mobile Commerce (WCM '02)* (p. 44). ACM Press.

Murray, J. (2000). Wearable Computers in Battle: Recent Advances in the Land Warrior System. In *Proceedings of the Fourth International Symposium on Wearable Computers* (pp. 169-170). IEEE.

Nagata, S. (2003). Multitasking and Interruptions During Mobile Web Tasks. In *Proceedings of the HFES Annual Meeting* (pp. 1341-1346). St. Louis, Missouri: Mira Digital Publishing.

Neerincx, M. (2004). Cognitive Task Load Analysis: Allocating Tasks and Designing Support. In E. Hollnagel (Ed.), *Handbook of Cognitive Task Design* (pp. 283-306). Mahwah: Lawrence Erlbaum Associates.

Neerincx, M. (2010). Situated Cognitive Engineering for Crew Support in Space. *Personal and Ubiquitous Computing,* 12 pages-DOI 10.1007/s00779-010-0319-3.

Neerincx, M., Grootjen, M., & Veltman, J. (2004). How to manage cognitive task load during supervision and damage control in an all-electric ship. *IASME Transactions, 1,* 253-258.

Neerincx, M. & Lindenberg, J. (1999). Supporting individual situation awareness in web-environments. In R.Pikaar (Ed.), *Ergonomie in Uitvoering: De Digitale Mens* (pp. 144-155). Nederlandse Vereniging voor Ergonomie.

Neerincx, M. & Lindenberg, J. (2008). Situated cognitive engineering for complex task environments. In J.M.C. Schraagen, L. Militello, T. Ormerod, & R. Lipshitz (Eds.), *Naturalistic Decision Making and Macrocognition* (pp. 371-387). Aldershot, UK: Ashgate.

Neerincx, M., Lindenberg, J., Smets, N., Grant, T., Bos, A., Olmedo Soler, A. et al. (2006). Cognitive Engineering for Long-Duration Missions: Human-Machine Collaboration on the Moon and Mars. In *Proceedings of the 2nd IEEE International Conference on Space Mission Challenges for Information Technology* (pp. 40-46). Los Alamitos, CA: IEEE.

Neerincx, M. & Streefkerk, J. W. (2003). Interacting in Desktop and Mobile Context: Emotion, Trust and Task performance. In E. Aarts, R. Collier, E. van Loenen, & B. de Ruyter (Eds.), *Proceedings of EUSAI 2003: Lecture Notes in Computer Science* (pp. 119-132). Berlin etc.: Springer.

Neerincx, M., van Doorne, H., & Ruijsendaal, M. (2000). Attuning computer-supported work to human knowledge and processing capacities in ship control centres. In J.M.C. Schraagen, S. Chipman, & V. Shalin (Eds.), *Cognitive task analysis* (pp. 341-362). Mahwah, New Jersey: Erlbaum.

Norman, D. (1986). Cognitive engineering. In D. Norman & S. Draper (Eds.), *User-Centered System Design: New perspectives on human-computer interaction* (pp. 31-62). Hillsdale: LEA.

Nova, N., Girardin, F., Molinari, G., & Dillenbourg, P. (2006). The Underwhelming Effects of Location-Awareness of Others on Collaboration in a Pervasive Game. *Frontiers in Artificial Intelligence and Applications, 137,* 224-238.

Nulden, U. (2003). Investigating police patrol practice for design of IT. In *Proceedings of CHI 2003* (pp. 820-821).

Oulasvirta, A., Petit, R., Raento, M., & Tiitta, S. (2007). Interpreting and Acting on Mobile Awareness Cues. *Human-Computer Interaction, 21,* 97-135.

Oulasvirta, A., Tamminen, S., Roto, V., & Kuorelahti, J. (2005). Interaction in 4-Second Bursts: The Fragmented Nature of Attentional Resources in Mobile HCl. In *Proceedings of CHI 2005* (pp. 919-927).

Paramythis, A., Totter, A., & Stephanidis, C. (2001). A modular approach to the evaluation of adaptive user interfaces. In *Proceedings of Eight International Conference on User Modeling 2001.* Freiburg.

Parasuraman, R. & Miller, C. (2004). Trust and Etiquette in High-Criticality Automated Systems. *Communications of the ACM, 47,* 51-55.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs. *Journal of Cognitive Engineering and Decision Making, 2,* 140-160.

Paymans, T. & Lindenberg, J. (2004). Usability trade-offs for adaptive user interfaces: ease of use and learnability. In *Proceedings of the IUI conference 2004* (pp. 301-303). New York: ACM Press.

Petersen, S., Cassens, J., Kofod-Petersen, A., & Divitini, M. (2008). To be or not to be Aware: Reducing Interruptions in Pervasive Awareness Systems. In *Proceedings of the Second International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies* (pp. 327-332). IEEE.

Pica, D. & Sørensen, C. (2004). On Mobile Technology in Context: Exploring Police Work. *Journal of Computing and Information Technology, 12,* 287-296.

Police Union (2009). Quickscan Beach Riot Hoek van Holland [Resultaten Quickscans Strandrellen Hoek van Holland]. Available online in Dutch (November 7, 2009) at http://www.politiebond.nl/downloads/96dfe4bef686ade2b69ee4e4d0c09d54.pdf. [On-line].

Preece, J., Abras, C., & Maloney-Krichmar, D. (2004). Designing and evaluating online communities: research speaks to emerging practice. *International Journal of Web-based Communities, 1,* 2-18.

Riboni, D. & Bettini, C. (2009). Context-aware Activity Recognition through a Combination of Ontological and Statistical Reasoning. In D. Zhang (Ed.), *Ubiquitous Intelligence and Computing (LNCS 5585)* (pp. 39-53). Berlin Heidelberg: Springer.

Roda, C. (2010). Attention support in digital environments. Nine questions to be addressed. *New Ideas in Psychology, 28,* 354-364.

Rogers, Y., Brignull, H., & Scaife, M. (2002). Designing Dynamic Interactive Visualisations to Support Collaboration and Cognition. In *Proceedings of Symposium on Collaborative Information Visualization Environments*. London: IEEE.

Rosson, M. B. & Carroll, J. M. (2002). *Usability engineering: scenario-based development of human-computer interaction*. San Francisco: Academic Press.

Roto, V., Oulasvirta, A., Haikarainen, T., Kuorelahti, J., Lehmuskallio, H., & Nyyssonen, T. (2004). *Examining mobile phone use in the wild with quasi-experimentation* (Rep. No. HIIT Technical Report 2004-1). Helsinki, Finland: HIIT.

Rovira, E., McGarry, K., & Parasuraman, R. (2007). Effects of Imperfect Automation on Decision Making in a Simulated Command and Control Task. *Human Factors, 49,* 76-87.

Rypkema, J., van der Lee, M., van der Meiler, P.P., & Weitenberg, A.I.M. (2005). *Situation Awareness – Information needs of soldiers during non-combattasks [Informatiebehoefte van soldaten tijdens niet-gevechtstaken]*. (Technical Report TNO-DV3 2005 B092). Soesterberg: TNO Defence, Security and Safety.

Safety Region Kennemerland (2009). Reconstruction C2000 during Poldercrash [Reconstructie C2000 tijdens Poldercrash]. Available online in Dutch (November 7, 2009) at http://www.ggdkennemerland.nl/client/1/?websiteid=1&contentid=2840&hoofdid=0 [On-line].

Salas, E., Cooke, N., & Rosen, M. (2008). On Teams, Teamwork, and Team Performance: Discoveries and Developments. *Human Factors, 50,* 540-547.

Salas, E., Prince, C., Baker, D., & Shrestha, L. (1995). Situation awareness in team performance: Implications for measurement and training. *Human Factors: Special issue on situation awareness, 37,* 123-136.

Salmon, P., Stanton, N., Walker, G., & Green, D. (2006). Situation awareness measurement: A review of applicability for C4i environments. *Applied Ergonomics, 37,* 225-238.

Salvucci, D. & Bogunovich, P. (2010). Multitasking and Monotasking: The Effects of Mental Workload on Deferred Task Interruptions. In *Proceedings of CHI'10* (pp. 85-88). Atlanta: ACM.

Sapateiro, C., Antunes, P., Zurita, G., Vogt, R., Infante, P., & Baloian, N. (2008). Supporting Unstructured Activities in Crisis Management: A Collaboration Model and Prototype to Improve Situation Awareness. In F. Friedrich & B. van der Walle (Eds.), *Proceedings of ISCRAM 2008*. Washington DC.

Sarter, N. B. & Schroeder, B. (2001). Supporting Decision Making and Action Selection under Time Pressure and Uncertainty: The Case of In-Flight Icing. *Human Factors, 43,* 573-583.

Sawhney, N. & Schmandt, C. (2000). Nomadic Radio: Speech and Audio Interaction for Contextual Messaging in Nomadic Environments. *ACM Transactions on Computer-Human Interaction, 7,* 353-383.

Shafer, S., Brumitt, B., & Cadiz, J. (2001). Interaction Issues in Context-Aware Interactive Environments. *Human-Computer Interaction, 16,* 363-378.

Shneiderman, B. (1997). Direct Manipulation for Comprehensible, Predictable, and Controllable User Interfaces. In *Proceedings of Intelligent User Interfaces Conference 1997* (pp. 33-39). Orlando, Florida.

Smets, N., Neerincx, M., te Brake, G., Buurman, T., & van Oostendorp, H. (in press). Evaluation of navigation interfaces in virtual versus real environments. *Entertaiment Computing, submitted to Special Issue*.

Smets, N., Streefkerk, J. W., & Neerincx, M. (2010). Commanders Dashboard: Overview of tactical changes to improve situated decision making in the field. In W.P. Brinkman & M. Neerincx (Eds.), *Proceedings of ECCE 2010* (pp. 193-200). Delft, the Netherlands: ACM.

Smets, N., te Brake, G., Lindenberg, J., & Neerincx, M. (2007). Influence of mobile map size and user capacities on situation awareness tested in a virtual environment. In B. van der Walle, P. Burghardt, & C. Nieuwenhuis (Eds.), *Proceedings of ISCRAM 2007* (pp. 557-564).

Sorensen, C.-F., Wu, M., Sivaharan, T., Blair, G. S., Okanda, P., Friday, A. et al. (2004). A context-aware middleware for applications in mobile Ad Hoc environments. In *Proceedings of the 2nd workshop on Middleware for pervasive and ad-hoc computing* (pp. 107-110). Toronto, Canada: ACM.

Sorensen, C. & Pica, D. (2005). Tales from the police: Rhythms of interaction with mobile technologies. *Information and organization, 15,* 125-149.

Stanton, N., Stewart, R., Harris, D., Houghton, R., Baber, C., McMaster, R. et al. (2006). Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. *Ergonomics, 49,* 1288-1311.

Stijnman, G. (2004). *Evaluation Pilot "Mobiel Blauw" [Evaluatie Pilot Mobiel Blauw]* (Rep. No. Internal report). Amsterdam: Police Amsterdam Amstelland.

Straus, S., Bikson, T., Balkovich, E., & Pane, J. (2010). Mobile Technology and Action Teams: Assessing BlackBerry Use in Law Enforcement Units. *Computer Supported Cooperative Work (CSCW), 19,* 45-71.

Streefkerk, J. W., van Esch-Bussemakers, M., & Neerincx, M. (2006). Designing personal attentive user interfaces in the mobile public safety domain. *Computers in Human Behavior, 22,* 749-770.

Streefkerk, J. W., van Esch-Bussemakers, M., & Neerincx, M. (2007). Context-Aware Notification for Mobile Police Officers. In D. Harris (Ed.)*, Engineering Psychology and Cognitive Ergonomics, LNAI 4562* (pp. 436–445). Berlin: Springer.

Streefkerk, J. W., van Esch-Bussemakers, M., & Neerincx, M. (2008). Field evaluation of a mobile location-based notification system for police officers. In *Proceedings of Mobile HCI 2008* (pp. 101-108). Amsterdam, The Netherlands: ACM.

Streefkerk, J.W., van Esch-Bussemakers, M., & Neerincx, M. (2009). Context-Aware Team Task Allocation to Support Mobile Police Surveillance. In *Foundations of Augmented Cognition (LNCS 5638)* (pp. 88-97). Berlin, Heidelberg: Springer.

Streefkerk, J.W., van Esch-Bussemakers, M., & Neerincx, M. (2010). Balancing Costs and Benefits of Automated Task Allocation in Mobile Surveillance. In *Proceedings of the 2010 European Conference on Cognitive Ergonomics* (pp. 99-107). Delft, the Netherlands.

Streefkerk, J. W., van Esch-Bussemakers, M., Neerincx, M., & Ait-Yaiz, R. (2006). *Attentive services: relevant factors for design* (Rep. No. TNO DV3 2006 A056). Soesterberg: TNO Defence, Security and Safety.

Streefkerk, J.W., van Esch-Bussemakers, M., Neerincx, M., & Looije, R. (2008). 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.

Streefkerk, J. W., Wiering, C., van Esch-Bussemakers, M., & Neerincx, M. (2008). Effects of Presentation Modality on Team Awareness and Choice Accuracy in a Simulated Police Team Task. In *Proceedings of the 52nd HFES Annual Meeting* (pp. 378-382). New York City, NY: HFES.

Tamminen, S., Oulasvirta, A., Toiskallio, K., & Kankainen, A. (2004). Understanding mobile contexts. *Personal and Ubiquitous Computing, 8,* 135-143.

Te Brake, G., de Greef, T., Lindenberg, J., Rypkema, J., & Smets, N. (2006). Developing adaptive user interfaces using a game-based simulation environment. In B. van Der Walle & M. Turoff (Eds.), *Proceedings of ISCRAM 2006*. Newark, NJ.

Te Brake, G. & van der Kleij, R. (2010). Map-Based Support for Effective Collaboration in Micro-mobile Virtual Teams. *Interactive Collaborative Information Systems, 281,* 183-203.

van der Kleij, R., de Jong, A., te Brake, G., & de Greef, T. (2009). Network-aware support for mobile distributed teams. *Computers in Human Behavior, 25,* 940-948.

van Loon, M. (2004). Digital Briefing [Digitale Briefing]. *APB, 22,* 16-17.

Venkatesh, V. & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science, 46,* 186-204.

Vermeulen, B. (2004). *Freeband SPEARS Final Report: A scalable personal area services platform* (Rep. No. Freeband/SPEARS/D1.13). Enschede: Telematica Instituut.

Vertegaal, R. (2002). Designing attentive interfaces. In *Proceedings of the ACM Symposium on Eye tracking research and applications* (pp. 23-30). New York: ACM Press.

Vetere, F., Howard, S., Pedell, S., & Balbo, S. (2003). Walking through mobile use: novel heuristics and their application. In *Proceedings of OzCHI 2003* (pp. 24-32). Brisbane, Australia.

Vicente, K., Hayes, B., & Willeges, R. (1987). Assaying and isolating individual-differences in searching a hierarchical file system. *Human Factors, 29,* 349-359.

Wahlster, W. (2003). Smartkom: Symmetric Multimodality in an Adaptive and Reusable Dialogue Shell. In R. Krahl & D. Gunther (Eds.), *Proceedings of the Human Computer Interaction Status Conference* (pp. 47-62). Berlin.

Wang, Y. & Emurian, H. (2005). An overview of online trust: Concepts, elements, and implications. *Computers in Human Behavior, 21,* 105-125.

Weibelzahl, S. (2005). Problems and Pitfalls in Evaluating Adaptive Systems. In S. Chen & G. Magoulas (Eds.), *Adaptable and Adaptive Hypermedia Systems* (pp. 285-299). Hershey, PA: IRM Press.

Weld, D. S., Anderson, C. R., Domingos, P., Etzioni, O., Gajos, K., Lau, T. A. et al. (2004). Automatically Personalizing User Interfaces. In *Proceedings of the International Joint Conference on Artificial Intelligence 2003* (pp. 1613-1619). Acapulco, Mexico.

Wickens, C. D. (1987). Attention. In P.A. Hancock (Ed.), *Human factors psychology* (pp. 29-80). North-Holland: Elsevier.

Wickens, C. D. (2000). The trade-off for design for routine and unexpected performance: implications of situation awareness. In M.Endsley & D. J. Garland (Eds.), *Situation Awareness and Measurement* (pp. 190-202). Mahwah: LEA.

Wickens, C. D. (2008). Situation Awareness: Review of Mica Endsley's 1995 Articles on Situation Awareness Theory and Measurement. *Human Factors, 50,* 397-403.

Wickens, C. D. & Dixon, S. (2007). The benefits of imperfect diagnostic automation: a synthesis of the literature. *Theoretical Issues in Ergonomics Science, 8,* 201-212.

Wiredu, G. (2007). User appropriation of mobile technologies: Motives, conditions and design properties. *Information and Organization, 17,* 110-129.

Wood, S., Cox, R., & Cheng, P. (2004). Attention Design: Seven issues to consider. In *Proceedings of the workshop "Designing for Attention" during the 18th Human Computer Interaction Group Annual Conference.* Leeds, UK.

Yeh, Y. Y. & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors, 30,* 111-120.

York, J. & Pendharkar, P. C. (2004). Human-computer interaction issues for mobile computing in a variable work context. *International Journal of Human-Computer Studies, 60,* 771-797.

Young, M. S. & Stanton, N. (2002). Malleable Attentional Resources Theory: A New Explanation for the Effects of Mental Underload on Performance. *Human Factors, 44,* 365-375.

Zhai, S. (2003). What's in the eyes for attentive input. *Communications of the ACM, 46,* 34-39.

Zhang, D. & Adipat, B. (2005). Challenges, Methodologies, and Issues in the Usability Testing of Mobile Applications. *International Journal of Human-Computer Interaction, 18,* 293-308.

Zijlstra, F. (1993). *Efficiency in work behavior. A design approach for modern tools.* Delft University of Technology.

Zipf, A. & Jost, M. (2006). Implementing adaptive mobile GI services based on ontologies: Examples from pedestrian navigation support. *Computers, Environment and Urban Systems, 30,* 784-798.