# Handbook of Research on User Interface Design and Evaluation for Mobile Technology

## Volume II

Joanna Lumsden
*National Research Council of Canada*
*Institute for Information Technology – e-Business, Canada*

Information Science REFERENCE

INFORMATION SCIENCE REFERENCE

Hershey · New York

Acquisitions Editor: Kristin Klinger
Development Editor: Kristin Roth
Senior Managing Editor: Jennifer Neidig
Managing Editor: Sara Reed
Copy Editor: Joy Langel, Katie Smalley, and Angela Thor
Typesetter: Jeff Ash
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

and in the United Kingdom by
Information Science Reference (an imprint of IGI Global)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 0609
Web site: http://www.eurospanonline.com



Library of Congress Cataloging-in-Publication Data

Handbook of research on user interface design and evaluation for mobile technology / Joanna Lumsden, editor.

p. cm.

Summary: "This book provides students, researchers, educators, and practitioners with a compendium of research on the key issues surrounding the design and evaluation of mobile user interfaces, such as the physical environment and social context in which a device is being used and the impact of multitasking behavior typically exhibited by mobile-device users"--Provided by publisher.

Includes bibliographical references and index.

ISBN 978-1-59904-871-0 (hardcover) -- ISBN 978-1-59904-872-7 (ebook)

1. Mobile computing--Handbooks, manuals, etc. 2. Human-computer interaction--Handbooks, manuals, etc. 3. User interfaces (Computer systems)--Handbooks, manuals, etc. I. Lumsden, Joanna.

QA76.59.H36 2008

004.165--dc22

2007024493

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

# Chapter XLV
# Evaluating Context-Aware Mobile Interfaces for Professionals

**Jan Willem Streefkerk**
*TNO Defense, Security and Safety, The Netherlands*
*Delft University of Technology, The Netherlands*

**Myra P. van Esch-Bussemakers**
*TNO Defense, Security and Safety, The Netherlands*

**Mark A. Neerincx**
*Delft University of Technology, The Netherlands*
*TNO Defense, Security and Safety, The Netherlands*

**Rosemarijn Looije**
*TNO Defense, Security and Safety, The Netherlands*

## ABSTRACT

*Evaluation refines and validates design solutions in order to establish adequate user experiences. For mobile user interfaces in dynamic and critical environments, user experiences can vary enormously, setting high requirements for evaluation. This chapter presents a framework for the selection, combination, and tuning of evaluation methods. It identifies seven evaluation constraints, that is, the development stage, the complexity of the design, the purpose, participants, setting, duration, and cost of evaluation, which influence the appropriateness of the method. Using a combination of methods in different settings (such as Wizard-of-Oz, game-based, and field evaluations) a concise, complete, and coherent set of user experience data can be gathered, such as performance, situation awareness, trust, and acceptance. Applying this framework to a case study on context-aware mobile interfaces for the police resulted in specific guidelines for selecting evaluation methods and succeeded to capture the mobile context and its relation to the user experience.*

## INTRODUCTION

In designing mobile support systems, 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 the dynamic use context, specific constraints of devices and risks of negative transfer from desktop experiences to mobile experiences (Nagata, 2006).

Due to these three issues, the user experience of mobile user interfaces is still an important bottleneck for services in the professional domain (Marcus & Gasperini, 2006). Realizing adequate user experiences is done by selecting the right method, based on specific constraints for evaluation of mobile, context-aware applications. Combining evaluation methods should capture the dynamic context aspects and their relations to the user experience in a complete, concise, and coherent way (cf. Neerincx & Lindenberg, in press). Finally, tuning of techniques and measures should ensure that the obtained results are relevant to the application domain.

### 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. In these domains, context-aware mobile devices have potential to support specific tasks such as notification to relevant information in context or facilitating communication with specific team members. Example domains include the police, ambulance and firefighter services, Urban Search and Rescue (USAR) teams, and the armed forces. Context-aware mobile devices have not yet become widespread in these domains.

Evaluation for the professional domain is distinguished from other domains by the following 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) to ask police officers about their emotional response toward the interaction with a mobile device. It seems more relevant to measure how many more criminal cases get solved in less time than before the introduction of the device. Secondly, 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 simulators may prove useful. Finally, access to professional end-users for evaluation purposes may be limited due to busy schedules and limited resources. The following case study is used to focus the discussion of evaluation methods for mobile context-aware interfaces and to provide an example from the professional application domain.

### Case Study: Evaluating a Mobile Support System for Police Officers

For mobile police officers, increasingly more (multimedia) information becomes available to perform their tasks. In addition, both the interaction possibilities with devices and the momentary user needs for information or services continuously change over time and place (Baber, Haniff, Sharples, Boardman, & Price, 2001). Finally, shared situation awareness (SA) and communication within or between teams are vital for task execution, but may be diminished due to distributed persons and locations. Both theory and police practice show a clear need for interfaces that attract and guide the attention of individual officers or teams to relevant, high priority information or objects in a mobile setting (Streefkerk, Van Esch-Bussemakers, & Neerincx, 2006). The PAUI (Personal Attentive User Interface) project aims at designing and evaluating an adaptive user interface to support mobile police officers.

Throughout this chapter we will use the police officers' surveillance task as an example. On surveillance, police team members have to detect criminal incidents or respond to incoming calls and take fast and effective action. This requires notification of relevant information at the right time and place and optimal situation awareness, for instance, by knowing exactly where they are and where their colleagues are. A context-aware system can support these processes. In order to do this, the support system needs to have knowledge about the user, his use context, and the task he is working on. Subsequently, the system can adapt the interaction and communication to this knowledge. In this case study, a context-aware mobile system is designed that notifies police officers to incidents, based on their location and task. The main question here is to which context, user, and task factors this system should adapt and how it should adapt the interaction. Evaluation plays an important part in this project. At various moments, evaluating design solutions and concepts checks their validity for end-users, the use context, and the application domain.

## Mobile Use Context

Context-aware mobile user interfaces are developed to improve the user experience by adapting the system behavior, based on a model of relevant use context factors. User experience is a term used to describe cognitive, affective, and social responses that are induced by the use of a product or service. However, the actual effect on user performance and acceptance has been assessed insufficiently for current applications (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 evaluation in real use contexts that give validity to the research results (Kjeldskov & Graham, 2003). Use context is especially important for mobile devices as it can change constantly, in contrast with the use context of desktop applications. Although recreating central aspects of the mobile use context in the lab is sufficient to identify usability problems (Kjeldskov, Skov, Als, & Hoegh, 2004), the added value of field evaluation lies primarily in a deeper insight into the user experience in a dynamically changing context. In addition, field evaluation provides insights into effects of environmental factors, such as distractions, lighting conditions, body movement, and unreliable wireless networks (Duh, Tan, & Chen, 2005; Zhang & Adipat 2005).

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 outside the use context, although a field study was used to guide the initial design (Jiang, Chen, Hong, Wang, Takayama, & Landay, 2004). 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. A related project, Freeband FRUX, aims to design mobile applications for police and rescue workers (Van Eijk, De Koning, Steen, & Reitsma, 2006) by incorporating end-users in the analysis stage and a 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 (Kellar, Inkpen, Dearman, Hawkey, Ha, et al., 2004; Vetere, Howard, Pedell, & Balbo, 2003; Zhang & Adipat, 2005). A shift can be seen towards employing new techniques to sample the user experience within the context of use. 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). However, these solutions are still in the development stage. Concluding from the discussion, evaluation of mobile context-aware systems is lacking a coherent and concise set of methods and techniques 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.

In the remainder of this chapter, first, the constraints in the evaluation of mobile adaptive systems for professionals are described. Then, following a user-centered design approach, the framework will be applied to the case study described. This framework is not intended to fully capture all existing evaluation methods, but to provide a practical approach for evaluation of professional mobile systems and present a "core" set of methods. Best practices, problems, and lessons learned are described in depth as they apply to the case study. It should be noted that this framework is general and can be applied to other evaluation methods than the ones mentioned in this chapter. Finally, specific guidelines for evaluation of mobile, adaptive systems are presented.

## FRAMEWORK OF EVALUATION CONSTRAINTS

An effective and efficient use of evaluation methods is aimed 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, for example, by complementing and cross-validating results between methods. Further tuning of methods should ensure that results are relevant to the application domain. The framework distinguishes the following constraints that influence which methods, techniques, and measures can be employed.

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 changes due to the changing context and this emerging, adaptive behavior should be tested in a proper way. 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 an MP3 player. Thus, the *setting* of the evaluation is important. Access to representative end-users and situations in the professional domain can be limited, begging the question *who* to include in your evaluation and for *how long*. Finally, different methods for evaluating mobile devices involve different *costs* in both time and resources. Concluding, seven constraints were identified that can be summarized as the following questions:

- Which stage of the development process are you currently in?
- What is the purpose of the evaluation?
- How complex is the design?
- Who are your participants?
- In which setting will the evaluation take place?
- What is the duration of the evaluation?
- What are the costs of the evaluation?

## Stage in the 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 & Adipat, 2005). Here, evaluation provides an important proof of concept that the adaptation model and application result in meaningful support.

## 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. It is distinguished between formative methods, used to generate design solutions, and summative methods, used to measure acceptance of designs. Within the 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 this case study, evaluation of revolutionary new concepts is adapted to specific police contexts and tasks. This purpose is in contrast with redesigning or improving existing applications.

## 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 & Jöst, 2005) that is, the adaptive system behavior. Design solutions with different degrees of complexity need different evaluation approaches. 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 there should always be a fit to the dynamic context of use. This point is closely related to the innovativeness of the system. A factor that further increases the complexity is the fact that users themselves also show adaptive behavior. A system that dynamically adapts to dynamic user characteristics 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 hard 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, Totter, & Stephanidis, 2001).

## Participants

A 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 & Sørensen, 2004). Determining these requirements, user characteristics and needs 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. End-users are particularly necessary during the analysis and implementation stages because of their knowledge of the mobile and dynamic use context and their 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, 2006). When no actual end-users can be involved, a careful selection of participants has to ensure they are representative of end-users.

## 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 & Graham, 2003). Environment independent methods are not situated in the use environment. 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 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 essential elements from the use environment 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 environment 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 and environmental factors (such as low bandwidth) are taken into account during the evaluation, providing more realistic results (Zhang & Adipat, 2005).

## 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, that is, 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 context of use (Kort & De Poot, 2005). Tuning measures to the application domain can be done by relating them to performance criteria for professionals (cf. Neerincx & Lindenberg, in press; see Box 1).

## Evaluation Cost

Finally, 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, Graham, Pedell, Vetere, Howard, Balbo, & Davies, 2005). Recent comparisons between methods show that rapid reflection 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.

## Conclusion

From the discussion, it is clear that the stage, purpose, complexity, participants, setting, duration, and cost each impose constraints on which evaluation technique to use. When the constraints to the goal of evaluating context-aware professional user interfaces are specifically applied, 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 the final design.
- Actual end-users must be involved as *participants* 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 *setting* is restricted, simulation yields a realistic yet controlled evaluation environment.
- 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 can only be 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 participants as it makes little sense to evaluate a support for a specific police task using students in an artificial setting.

However, each constraint has its own unique contribution to the selection of techniques.

## APPLYING THE FRAMEWORK

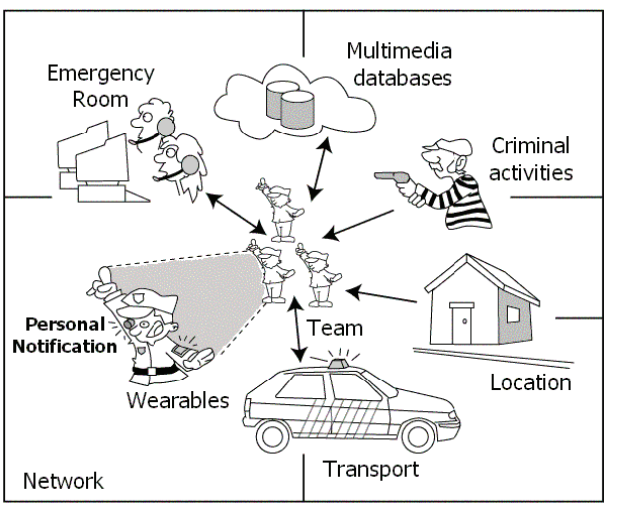
The framework of constraints to the case study of the PAUI project is now applied (described in the first section) on designing a context-aware notification user interface for police officers. In this case study, a user-centered design approach was followed, described in Box 1. Based on this approach, it is illustrated how the evaluation methods and techniques at each step in the UE method were selected, combined, and tuned and their benefits and limitations are discussed. The advanced stage of the project enables all methods to be addressed. However, it must be noted that this research is still ongoing and not all results are obtained yet. Theoretical considerations and relevant alternative techniques are presented in separate boxes (Box 2, 3, & 4). These techniques are not unique to evaluation of mobile context-aware applications. The boxes are categorized on the constraint of evaluation setting (environment independent, artificial, and natural).

The PAUI project started off with a definition of the concept. This concept was based on literature research of relevant HCI literature and domain research including participatory observation of the work domain. The concept for a support system for mobile police officers is shown in Figure 1. 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 is presented to a team of police officers, each having their own characteristics, task, and transport. Based on these factors, the individual officer will receive personal notification at the right time and place, possibly via a wearable or handheld device.

### Focus Group Evaluation

- **Stage:** Early analysis stage
- **Purpose:** Innovative design; formative
- **Complexity:** High; concept of adaptive system
- **Participants:** Thirty; police personnel with diverse backgrounds and experience
- **Setting:** Within and outside use environment
- **Duration:** Short; one half day
- **Costs:** High in resources; low in time

*Figure 1. Concept of the attentive user interface for police officers*



## *Box 1. User-centered design method*

The Usability Engineering (UE) method provides an empirical design and evaluation approach where knowledge about user needs and design solutions are refined until they meet the usability requirements. Because these cannot be predicted from the onset of design, involving users in the design process is regarded as an important necessity for a successful design. This method has been developed and applied for the design of user-interface support for space missions, ship control centers, and mobile services.

**Approach**
The UE method starts with the definition of a concept (see Figure 2, left side), which is a broad description of the proposed system. Scenarios are then drafted from the relevant application domain and describe users, their tasks, and context in a comprehensive, narrative style. Especially describing and understanding the dynamic nature of use contexts for mobile devices is crucial for applying UE. From the scenarios, the process of requirements analysis results in a requirements specification. These requirements describe in detail the user needs with respect to their work practice and the role the system fulfills in addressing these needs. User requirements form the basis for the system features. Features can be considered solutions to user needs and describe what functionality the system should have. As the method progresses from concept to features, the level of detail increases.

**Evaluation**
Evaluation of concept, scenarios, user requirements, and features is done by validating them to objective and subjective quality criteria, such as established Human Computer Interaction (HCI) metrics and new HCI metrics specifically adapted for mobile devices (see Figure 2, right side). Effectiveness, efficiency, and satisfaction are established criteria for evaluating HCI (ISO 9241-11, 1998) but should be matched to domain-specific performance criteria. Furthermore, professionals have to be able to rely on a system while performing critical tasks. Therefore, trust and user acceptance are important aspects that influences actual use and develop over time (Marsh & Meech, 2000). Finally, mobile context-aware systems should maintain or heighten situation awareness (SA) by providing the right information at the right time (Endsley & Garland, 2000).
It is important to note that this UE method is an iterative process, with a full cycle including the assessment of the proposed features on HCI metrics, and further specification of these features based on this assessment. Parts of the system can also be evaluated, it is not necessary to evaluate the whole system at once. The end products of this cycle are generic guidelines, models and prototypes that are validated in their respective context and use domain.

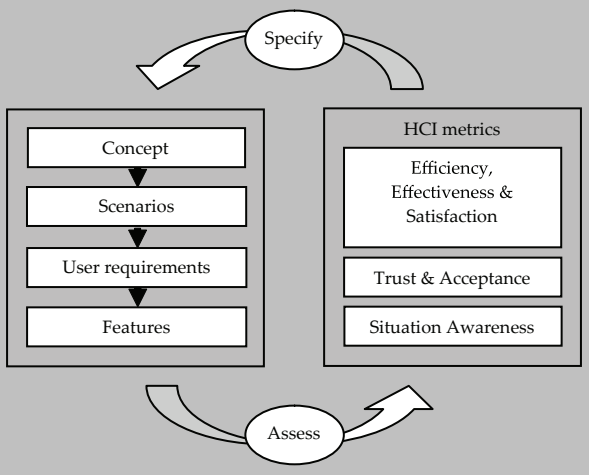
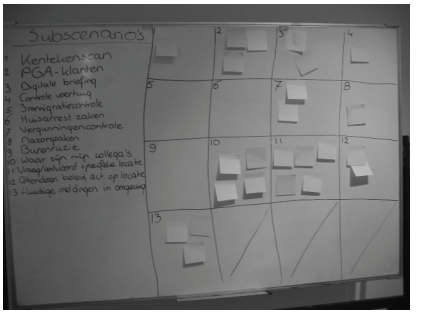



*Figure 2. The Usability Engineering Method (adapted from Streefkerk et al., 2006)*

**Further Reading**

- Neerincx and Lindenberg (2005)
- Lindenberg, Nagata and Neerincx (2003)
- Gorlenko and Merrick (2003)
- Vetere et al. (2003)

*Figure 3. Illustrations from the focus groups showing the participants (left) and the plenary voting system (right)*




## Selection and Application of Method

To evaluate the high-level concept, involvement of end-users to capture the mobile police context and domain knowledge was needed. Furthermore, a brainstorm discussion of innovative use scenarios and context of use from different perspectives was wanted because the purpose was to gather as much and diverse information as possible. Finally, no concrete previous examples were available that could serve as a reference point. Based on these considerations, a focus group setting with police officers and management personnel was selected. Alternative methods such as questionnaires or survey research (see Box 2) did not satisfy these requirements.

During this focus group, the concept presented was explained and the participants brainstormed about possible situations where the context-aware system would have an added benefit (see Figure 3). In four small groups (separate for police officers and management personnel) of five to six participants, ideas were written down in the form of short stories with illustrations. The participants were stimulated to "think out of the box" and to give room for new and refreshing ideas. The groups were moderated by a researcher who guided and stimulated the discussion with examples and recorded comments. At the end of the session, a plenary vote was taken on which stories illustrated maximum benefit 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 was selected for further use.

The resulting scenario clearly illustrated problems police officers experience when on surveillance. Examples include quickly assessing when a call is relevant to them and what priority a call has. The participants considered these problems to result in unnecessary distraction from their surveillance task and to hamper optimal task execution. They also believed the concept to be a possible solution to these problems.

## Combination of Techniques

To validate the results of the focus group, they were combined with rapid ethnography (see Box 4), conducted on surveillance with three police officers. This ethnography focused on the work context, communication tools, location information, and support for procedures that officers use. The results of the study identified relevant moments in the work of police officers that could use support from a context-aware system. For example, sometimes a call is only relevant during a specific shift or for a specific group of officers. Based on these moments, relevant task and context aspects (such as location, task priority, and officer task history) were identified. By charting similarities and consistencies between the focus group scenario and the results from the ethnography study, the validity of these aspects was supported by actual work practice.

In addition, another focus group was used to evaluate the resulting usage scenario from the first focus group. This second focus group was similar in setting and participants to the first. Its purpose was more restricted because participants were instructed to think of realistic work situations where a context-aware system could provide benefits. From the scenario and discussion with participants, a list of requirements resulted. Again, a plenary vote decided which requirements was either indispensable, necessary, or merely worth considering (see Figure 3). This technique resulted in a prioritized list of domain-specific requirements, validated by end-users.

## Tuning of Techniques

Tuning of the focus groups and ethnography to the application domain was done in four ways: First, by involving a police officer in the organization and execution of the focus group. Second, by specifically focusing on problems that police officers on the move would encounter and could be solved by a context-aware system. For example, a context-aware system presents only calls that are in the direct vicinity of the police officer. Third, by relating the benefits of a context-aware system explicitly to performance criteria for police officers, such as amount of time spent on surveillance or response times to calls. Finally, during the focus group, by dividing the officers and management into two separate groups. This resulted in separate usage scenarios for officers and management and allowed relating scenario elements to specific user characteristics and roles.

## Benefits and Limitations

This evaluation resulted in a usage scenario and list of requirements, validated by end users and domain analysis. By analyzing and observing the surveillance task of the police officers in context, relevant moments for notification were identified. In addition, relevant context aspects (such as time,

*Box 2. Environment independent evaluation techniques*

*Focus group evaluation*
For designing mobile, context aware systems, this evaluation method is best used at 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.

*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.

**Further Reading**

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

location, and history) were captured. Employing end-users helped to determine priorities for the requirements. The combination of focus groups and rapid ethnography resulted in a unique and validated list of requirements for the design of mobile, context-aware applications for the police. Alternative approaches to evaluating high-level concepts are discussed in Box 2.

Some limitations to the quality of the focus groups were observed. The participants found it sometimes difficult to relate the scenario to specific, realistic work situations and had selective recall for some situations. They focused quickly on established procedures and found it hard to integrate possibilities offered by new technologies. In addition, for the moderators, it was difficult at times, to interpret insights into the working practice of police officers.

## Designing the Context-Aware System

After the second focus group, a list of requirements for the attentive user interface for police officers was compiled (Streefkerk et al., 2006). From this list, the requirement of notification was the first and most important focus. The attentive user interface should notify police officers to relevant information in their environment, without distracting them unnecessarily from their primary surveillance task. This was done by adapting the salience and information density of the notification (i.e., the notification style) based on two rules. First, when user workload was high, information was presented more concisely. For example, a short summary of the message was presented, prior to the actual message. Second, when message priority was high, the salience of the audiovisual signals was increased, for example, by using loud auditory signals. This adaptive notification principle was evaluated by simulating it on a handheld computer (PDA). The simulated system presented messages to the user in the different notification styles, based on their workload and the message priority (see Figure 4). High priority messages were presented with red visual flashing bars or icons and a sharp sound. Medium priority messages were presented with a soft sound and low priority messages without sound. In low workload situations the full message text was presented at once, but in high workload situations, first a summary of the message was presented.

## Wizard of Oz Evaluation

- **Stage:** Intermediate design stage
- **Purpose:** Validate innovative design solution; formative
- **Complexity:** Moderate; (simulated) adaptive functionality
- **Participants:** Twenty; representatives

*Figure 4. Screenshots from the PDA with a summary (left) and a high priority message (right)*

- **Setting:** Artificial; lab experiment employing Wizard of Oz setup
- **Duration:** Short; 2 hours
- **Costs:** Low in both time and resources

## Selection and Application of Method

To evaluate the innovative support concept of adaptive notification, a Wizard of Oz setup was chosen (see Box 3) based on the following considerations. The purpose of the evaluation was to guide the further design effort. A simulated setting that allowed recreated basic aspects of the police officers surveillance task was needed. Furthermore, a flexible environment was necessary because the influence of changing context (e.g., workload and message priority) on interaction with a mobile device was being tested. Finally, the Wizard of Oz setup allowed the concept to be empirically tested by systematically comparing two conditions. As the concept dealt with general instead of task-specific abilities, a representative participant group was used.

Twenty participants were involved in this study, representative to end-users in age and education. They had to perform a simulated police surveillance by watching videos, recalling targets, and answering questions on these videos (see Figure 5). Simultaneously, the researcher sent low, medium, or high priority messages at predefined moments to the PDA. Participants had 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 2 hours, including training, two scenarios, and debriefing.

## Combining and Tuning of Techniques

This evaluation combined both qualitative and quantitative techniques. Performance data (time on task, number of errors in task, and questions) were collected using event-logging on the PDA and questionnaires. Subjective judgments (notification intrusiveness, preference for condition) were 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 recall and describe different "targets" from the videos, which is an important surveillance skill. Results from the evaluation indicated that the adaptive notification is positively evaluated. Participants preferred the adaptive notification over non-adaptive and a trend towards better performance with the adaptive system was observed.

*Figure 5. Screenshots from the videos used in the Wizard-of-Oz evaluation*

*Box 3. Evaluation in artificial settings*

*Wizard of Oz*
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 human intervention. This technique is appropriate when no time-critical system performance is required.

*Game based evaluation*
Game-based evaluation provides 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.

**Further Reading**

- Dahlback, Jonsson, and Ahrenberg (1993)
- Lewis and Jacobson (2002)

## Benefits and Limitations

The Wizard of Oz setup managed to create the illusion of a working, adaptive support concept. Participants indicated they could compare the conditions easily, which improved accuracy and validity of their subjective judgments. However, additional training is necessary to facilitate the distinction between notification styles. This study delivered new insights into the user experience with an adaptive, context-aware system. It resulted in specific, validated notification styles, matched to user workload and message priority. Based on this evaluation, the concept is redesigned to employ more distinctive sounds and visual signals. Finally, it was found that the notification style has to match the task that has to be performed.

# Game-Based Evaluation

- **Stage:** Intermediate design stage
- **Purpose:** Innovative; summative; validate adaptive notification for teams
- **Complexity:** High; (simulated) advanced functionality
- **Participants:** Twenty-four participants in teams of three will be included, depending on availability
- **Setting:** Artificial; lab experiment employing game-based environment
- **Duration:** Extended; 4 hours
- **Cost:** Low in both time and resources

## Selection and Application of Method

In the previous Wizard of Oz study, participants could only make a limited set of decisions in a task they did not directly control. The next planned evaluation aims to evaluate the redesigned adaptive notification principle and model in a richer yet controlled environment. Game-based evaluation allows flexibility in recreating task-specific aspects of the use context, such as team tasks with multiple actors. The simulated reality of the task environment requires using end-users as participants. Furthermore, it allows measuring performance and shared situation awareness by accurately logging participants' behavior. Based on these considerations, a lab experiment in a game-based simulation environment was selected (see Box 3).

A surveillance environment will be created within the PC game Unreal Tournament (see Figure 6; for a description, see Te Brake et al., 2006). Including 24 participants in teams of three is aimed at, depending on availability. The team navigates through this environment on surveillance, including reconnaissance, gathering information, and communicating with team members. In addition, participants receive assignments for additional tasks (finding locations or items) via the context-aware system, simulated on the PDA. By modeling user workload, location, and task, it decides which participant to present with which task. Both performance and the appropriateness of the adaptive behavior are subjects of evaluation. Therefore, an experimental condition with the adaptive system will be compared to a non-adaptive system. The duration of the evaluation is approximately 4 hours to allow thorough training on using the environment.

## Combining and Tuning of Techniques

During this evaluation a combination of qualitative and quantitative measures is collected. Performance data include time on task, number of errors, and distance traveled. In addition, a measure for effectiveness of the system would be the number of tasks solved. Trust, acceptance, and preference were measured using questionnaires and rating scales. Situation awareness (SA) is measured with a technique called "freezing" (Endsley & Garland, 2000) where the workflow is 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 uses a think-aloud protocol to collect both positive and negative incidents in using the context-aware system.

The evaluation setup is tuned to the police environment by using a diverse set of tasks that are representative of police surveillance. Furthermore, the critical incidents reported by the police officers participating in this evaluation are analyzed carefully. These incidents may suggest the appropriateness of the context-aware system in the field.

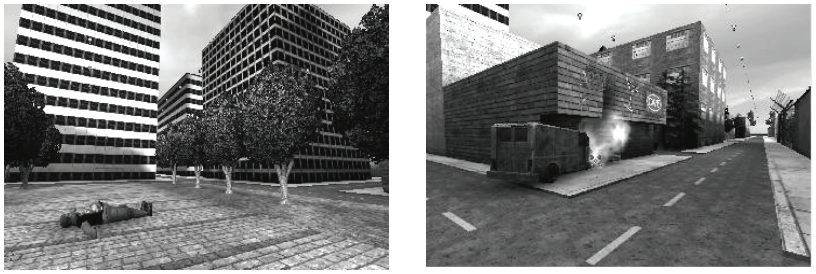
This evaluation is expected to result in a validation of the adaptive notification principle in a rich task-based setting. Furthermore, appropriate moments support is necessary, based on context factors such as task and location, are expected to be identified.

## Benefits and Limitations

The game-based technique allows for accurate quantitative measures of performance data and SA, because the behavior and navigation path of the participant are recorded. In addition, the appropriateness of the adaptive behavior can be measured as well. Furthermore, multiple participants can work collaboratively on one task in the same environment, allowing evaluation with teams. Finally, critical events can be pre-programmed into the scenario running in the game simulation.

Some factors negatively influence a game-based evaluation. Asking participants to fill out

*Figure 6. Screenshots from the game-based environment Unreal Tournament, showing a victim (left) and a car accident (right)*



SA questions and rating scales interferes with the task flow at certain moments. In addition, prior gaming experience should be well documented, as this influences participants' performance. Finally, some participants are susceptible to simulator sickness, which can occur in game-based simulation (Kolasinski, 1996).

## Field Evaluation in the Professional Domain

- **Stage:** Final implementation stage
- **Purpose:** Summative; validate functioning of final system in context
- **Complexity:** High; full system functionality
- **Participants:** Thirty; end-users
- **Setting:** Natural use environment
- **Duration:** Longitudinal; 3 months
- **Cost:** High in both time and resources

### Selection and Application of Method

The final evaluation planned in the PAUI project will be field evaluation of the adaptive notification system in the natural work setting (see Box 4). The purpose is to validate the full functioning of the innovative context-aware system with end-users and to provide the final "proof of concept" in the application domain. This system is evaluated in a longitudinal study to measure impact on work processes, trust, acceptance, and learning effects. The costs in both time and resources are relatively high compared to other methods, as the police

*Box 4. Evaluation in natural settings*

Field evaluation is conducted in natural environments, often during the final phase 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 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. A benefit of the technique is 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. 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*
Within natural setting evaluation, a distinction can be made between ethnographic field studies and field experiments. An ethnographic study would describe the functioning of the mobile context-aware system in the work context and require participation of the researcher in the work activities. Contrastingly, field experiments would test two versions of a context-aware system under different conditions to evaluate the influence on task performance. Field experimentation allows for more control but can only be used for restricted evaluation purposes, such as a usability evaluation.

**Further Reading**

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

organization and personnel have to participate. In addition, collecting and analyzing field data is necessary, further increasing the costs.

The final prototype of the context-aware user interface will be implemented on a mobile device, integrated with existing police infrastructures and made available to police officers. The context-aware system notifies police officers to relevant information based on their location, provides location of colleagues, and supports task switching and police procedures. Thirty officers with different roles, such as emergency aid, district surveillance, and prevention, participate in this evaluation. After an initial training phase, the system is used during daily work for a period of 3 months.

### Combining and Tuning of Techniques

This evaluation focuses on the user experience in context, integration of the system in work practice, and acceptance within the organization. Techniques include participatory observation, interviews, and questionnaires. To evaluate the user experience, critical incidents in task execution with the system are reported weekly by the officers. These reports are then related to the specific context variables logged by the system. Finally, researchers conduct a monthly participatory observation session on surveillance with officers. This technique aims at getting deeper insight into the system's impact on work processes.

The system is evaluated in a pre and post-test setup, thereby giving insight into changes caused by the system. Prior to evaluation, the expected effects of the system are captured in specific criteria. These performance criteria are 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 is collected by recording events from police databases and analyzing system events on the PDA.

### Benefits and Limitations

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 can the system be assessed in the actual and diverse work situations that occur naturally. This is a necessary and valuable step before actual implementation of the finished system, as it allows final changes and tuning of the system.

However, it is also a costly method as an advanced prototype, a mobile evaluation setup and the participation of end-users are necessary. In addition, at this implementation stage, it is difficult to make thorough changes in design when needed. Finally, 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.

## EVALUATION GUIDELINES

In the third section, four specific methods were described to evaluate mobile context-aware technology for professionals. Based on the benefits and limitations of these techniques and the lessons learned in the PAUI project, the key findings are summarized in the form of guidelines. In Table 1 the appropriateness of the focus group, Wizard of Oz, game-based and field evaluation with respect to the seven evaluation constraints is presented. A plus sign (+) means the technique is appropriate considering that particular constraint. A minus sign (-) means the technique is less appropriate.

As can be concluded from the preceding discussion, evaluation of mobile, context-aware applications differs from desktop evaluation. Therefore, guidelines specific for evaluation of mobile, context-aware applications are printed bold in Table 1. For example, game-based evaluation techniques are less appropriate or necessary in desktop evaluation. The focus on the use context is more important in early stages for mobile

*Table 1. Guidelines for the appropriateness of evaluation techniques (vertical) based on the constraints (horizontal)*

| | | Focus group | Wizard of Oz | Game-based | Field |
|---|---|---|---|---|---|
| **Stage** | Analysis | + | - | - - | ++ |
| | Design | + | + | ++ | ++ |
| | Implementation | - | - | ++ | + |
| **Purpose** | Formative | + | + | ++ | + |
| | Summative | - | - | ++ | ++ |
| **Complexity** | Low | + | - | - - | - |
| | Medium | - | + | ++ | - |
| | High | + | - | ++ | + |
| **Participants** | Representatives | - | + | ++ | - |
| | End-users | ++ | ++ | ++ | + |
| **Setting** | Independent | + | - | - - | - |
| | Natural | - | ++ | - - | ++ |
| | Artificial | - | + | ++ | - |
| **Duration** | Short | + | + | ++ | + |
| | Longitudinal | - | - | - - | ++ |
| **Costs** | Time | + | + | ++ | - |
| | Resources | - | + | ++ | - |

applications. In addition, to measure ecological validity in summative approaches, field testing is necessary. The participation of end-users with specific domain and context knowledge is more important than in desktop evaluation. In this way, the table can be used to get a quick overview of the appropriateness of these specific evaluation techniques.

## FUTURE TRENDS

Two important future trends in evaluation of mobile, context-aware systems include employing more and diverse game-based techniques, user experience sampling in context, and mixed reality techniques. The increasing use of game-based evaluation techniques are easily explained by their advantages over other techniques (see Box 3) such as adaptability, flexibility, and accuracy in measurements, particularly for use for professionals such as police officers and rescue workers (Te Brake et al., 2006).

By employing creative solutions, limitations of traditional evaluation techniques can be reduced. For example, giving subjective judgments in rating scales out of the use context is difficult due to recall problems. By employing context-aware questionnaires that are triggered by specific system events or specific context factors (such as location), these recall problems are circumvented. An additional benefit is that the factors that triggered the questionnaire can be simultaneously logged, providing a deeper insight into the use context (Kort & De Poot, 2005). Another novel development is evaluation in "augmented" mixed reality settings. The participant is wearing a head-mounted display while performing a task in a natural setting. This display shows both the real environment and a layer on which extra

information can be presented. By employing this setup, the power of simulation can be employed in natural evaluation settings.

## 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 & Graham, 2003).

In this chapter, it is argued that evaluation of mobile context-aware systems for professionals benefits from a systematic 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 a case study in designing support for mobile police officers. This resulted in specific guidelines for evaluation of mobile, adaptive systems for four specific techniques (focus group, Wizard of Oz, game-based, and field evaluation).

The approach helped to select a concise and coherent set of appropriate evaluation methods and techniques and to tune these to the appropriate application domain as was demonstrated by the case study. In addition, the framework contributed to the field of evaluation research by stating specific guidelines that reach beyond the current application domain and are suited for use in other domains as well. Evaluation of adaptive mobile systems in the professional domain 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 task and use context. In this way, both short-term usable services as well as long-term innovative support concepts for police officers are realized.

## ACKNOWLEDGMENT

The PAUI project is part of MultimediaN and funded by the Dutch Ministry of Economic Affairs.

## REFERENCES

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 6, pp. 21-28), *Industrial ergonomics, HCI, and applied cognitive psychology.* Aldershot: Ashgate.

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.

Dix, A., Finlay, J., Abowd, G., & Beale, R. (1998). *Human-computer interaction* (2nd ed.). New York: Prentice Hall.

Duh, H. B. L., Tan, G. C. B., & Chen, V. H. (2006). Usability evaluation for mobile device: A comparison of laboratory and field tests. In *Proceedings of Mobile HCI'06* (pp. 181-186). Helsinki, Finland.

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

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.

Gorlenko, L., & Merrick, R. (2003). No wires attached: Usability challenges in the connected mobile world. *IBM Systems Journal, 42*(4), 639-651.

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

Jiang, X., Chen, N. Y., Hong, J. I., Wang, K., Takayama, L., & Landay, J. A. (2004). SIREN: Context-aware computing for firefighting. In *Proceedings of Pervasive 2004*, Vienna.

Kellar, M., Inkpen, K., Dearman, D., Hawkey, K., Ha, V., MacInnes, J., Mackay, B., Nunes, M., Parker, K., Reilly, D., Rodgers, M., & Whalen, T. (2004). *Evaluation of mobile collaboration: Learning from our mistakes* (Tech. Rep. CS-2004-13). Canada: Dalhousie University, Faculty of Computer Science.

Kort, J., & De Poot, H. (2005). Usage analysis: Combining logging and qualitative methods. In *Proceedings of CHI '05 extended abstracts on human factors in computing systems,* (pp. 2121-2122).

Kjeldskov, J., & Graham, C. (2003). A review of mobile HCI research methods. In L. Chittaro (Ed.), *Proceedings of the 5th International Mobile HCI 2003 Conference* (pp. 317-335). Germany: Springer-Verlag.

Kjeldskov, J., Graham, C., Pedell, S., Vetere, F., Howard, S., Balbo, S., & Davies, J. (2005). Evaluating the usability of a mobile guide: The influence of location, participants and resources. *Behaviour and Information Technology, 24*(1), 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. In *Proceedings of Mobile HCI 2004*, Glasgow, Scotland.

Kolasinski, E. M. (1996). *Prediction of simulator sickness in a virtual environment.* Unpublished doctoral dissertation, University of Central Florida, Orlando.

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

Lindenberg, J., Nagata, S. F., & Neerincx, M. A. (2003). Personal assistant for online services: Addressing human factors. In D. Harris, V. Duffy, M. Smith, & C. Stephanides (Eds.), *Human-centred computing: Cognitive, social and ergonomic aspects* (pp. 497-501). London: LEA.

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*(5), 12-18.

Marsh, S., & Meech, J. (2000). Trust in design. In *Proceedings of the 2000 Conference on Computer Human Interaction* (pp. 45-46). The Netherlands: The Hague.

Nagata, S.F. (2006). *User assistance for multitasking with interruptions on a mobile device.* Unpublished doctoral dissertation.

Neerincx, M. A., & Lindenberg, J. (2005). Situated cognitive engineering for complex task environments. In *Proceedings of 7th International Conference on Naturalistic Decision Making.* The Netherlands, Amsterdam.

Neerincx, M. A., & Lindenberg, J. (in press). Situated cognitive engineering for complex task environments. In J. M. C. Schraagen, L. Militello, T. Ormerod, & R. Lipshitz (Eds.), *Natural decision making & macrocognition.* Aldershot, UK: Ashgate Publishing Limited.

Neerincx, M. A., Ruijsendaal, M., & Wolff, M. (2001). Usability engineering guide for integrated operation Support in space station payloads. *International Journal of Cognitive Ergonomics, 5*(3), 187-198.

Paramythis, A., Totter, A., & Stephanidis, C. (2001). A modular approach to the evaluation of adaptive user interfaces. In *Proceedings of Workshop at the 8th International Conference on User Modeling, UM 2001* (pp. 9-24).

Pica, D., & Sørensen, C. (2004). On mobile technology in context: Exploring police work. *Journal of Computing and Information Technology*, *12*(4), 287-295.

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

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 de Walle & M. Turoff (Eds.), *Proceedings of the 3rd International ISCRAM Conference.* Newark, NJ.

Van Eijk, R., De Koning, N., Steen, M., & Reitsma, E. (2006). *Developing a we-centric service for mobile police officers to support spontaneous communication*. (Tech. Rep. Freeband FRUX D1.8a). Available at https://doc.freeband.nl/dscgi/ds.py/Get/File-67631 (14-01-2007).

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

Weibelzahl, S. (2005). Problems and pitfalls in the evaluation of adaptive systems. In S. Chen & G. Magoulas (Eds.), *Adaptable and adaptive hypermedia systems* (pp. 285-299). Hershey, PA: IRM Press.

Zhang, D., & Adipat, B. (2005). Challenges, methodologies, and issues in the usability testing of mobile applications. *International Journal of Human-Computer Interaction, 18*(3)*,* 293-308.

Zipf, A., & Jöst, M. (2005). Implementing adaptive mobile GI services based on ontologies—Examples for pedestrian navigation support. In *CEUS-Computers, Environment and Urban Systems—An International Journal* (Special Issue on LBS and UbiGIS).

## KEY TERMS

**Adaptive User Interface:** A computer interface that adapts the interaction guided by a model of a specific set of factors

**Context Aware Technology:** Technology that models a set of factors about the use context to adapt the interaction with the user

**Empirical Evaluation:** An evaluation method in which results are derived by observation or experiment instead of theory

**Ethnography:** An evaluation method in the analysis stage during which end-users are observed and interviewed in their work environment

**Game-Based Evaluation:** An evaluation method in which a simulated environment is used adapted from existing computer games

**Requirements Analysis:** An analysis method during the analysis stage in which the user needs and requirements for design solutions are specified

**User-Centered Design:** A design methodology in which end-users needs and requirements guide the design choices

**User Experience:** The set of cognitive, affective, and social responses that are induced by the use of a product or service

**Wizard of Oz Evaluation:** An evaluation method in which a functioning system is simulated by a person controlling a non-functioning prototype