

Research Article

A Comparison of Field-Based and Lab-Based Experiments to Evaluate User Experience of Personalised Mobile Devices

Xu Sun¹ and Andrew May²

¹ *Science and Engineering, University of Nottingham Ningbo, 199 Taikang East Road, Ningbo 315100, China*

² *Loughborough Design School, Loughborough University, Ashby Road, Loughborough, Leicestershire LE11 3TU, UK*

Correspondence should be addressed to Xu Sun; xu.sun@nottingham.edu.cn

Received 30 July 2012; Accepted 12 January 2013

Academic Editor: Caroline G. L. Cao

Copyright © 2013 X. Sun and A. May. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There is a growing debate in the literature regarding the tradeoffs between lab and field evaluation of mobile devices. This paper presents a comparison of field-based and lab-based experiments to evaluate user experience of personalised mobile devices at large sports events. A lab experiment is recommended when the testing focus is on the user interface and application-oriented usability related issues. However, the results suggest that a field experiment is more suitable for investigating a wider range of factors affecting the overall acceptability of the designed mobile service. Such factors include the system function and effects of actual usage contexts aspects. Where open and relaxed communication is important (e.g., where participant groups are naturally reticent to communicate), this is more readily promoted by the use of a field study.

1. Introduction

Usability analysis of systems involving stationary computers has grown to be an established discipline within human-computer interaction. Established concepts, methodologies, and approaches in HCI are being challenged by the increasing focus on mobile applications. Real-world ethnographic studies have received relatively little attention within the HCI literature, and little specific effort has been spent on delivering solid design methodologies for mobile applications [1]. Researchers and practitioners have been encouraged to investigate further the criteria, methods, and data collection techniques for usability evaluation of mobile applications [2]. Lab-based experiments and field-based experiments are the methods most discussed in relation to evaluating a mobile application [2–4].

There has been considerable debate over whether interactions with mobile systems should be investigated in the field or in the more traditional laboratory environment. There seems to be an implicit assumption that the usability of a mobile application can only be properly evaluated in the field, for example, Kjeldskov and Stage [5]. Some argue that it is important that mobile applications are tested in

realistic settings, since testing in a conventional usability lab is unlikely to find all problems that would occur in real mobile usage (e.g., [2, 6, 7]). For example, Christensen et al. [6] presented a study of how ethnographic fieldwork can be used to study children's mobility patterns via mobile phones. Authors consider that field studies make it possible to carry out analysis that can broaden and deepen understanding of peoples' everyday life. However, some authors have highlighted how ethnographic field experiments are time consuming, complicate data collection, reduce experimental control, or are unacceptably intrusive [1, 3, 5].

Laboratory experiments are generally not burdened with the problems that arise in field experiments as the conditions for the experiment can be controlled, and it is possible to employ facilities for collection of high-quality data [5, 8]. However, Esbjörnsson et al. [9] and many other authors have argued that traditional laboratory experiments do not adequately simulate the context where mobile devices are used and also lack the desired ecological validity. This may lead to less valid data, where there is a potential disconnect between stated preferences, intentions, and actual experiences [4].

There are alternatives to field studies when assessing the impact of mobile devices. Adding contextual richness to

laboratory settings through scenarios and context simulation can contribute to the realism of the experiment while maintaining the benefits of a controlled setting [3, 5, 10]. The extent to which simulated scenarios represent a real-life situation is a critical determinant of the validity of the usability experiment [11]. In addition, for mobile devices, two basic contextual factors which need to be considered are mobility and divided attention. To replicate real-world mobility within a lab setting, test participants have been asked to use a treadmill or walk on a specifically defined track in a lab setting (e.g., [12]). To replicate divided attention, a range of measures have been used in the past. For example, to assess the impact of information provision on drivers, a range of simulators have been used, from low-fidelity personal computer-based simulations [13] to high-fidelity simulators with large projection screens involving real dashboards [14]. These simulators recreate divided attention, as well as enabling task performance measurement.

This paper reports a combination of a field and lab-based evaluation in order to assess the impact of personalisation on the user experience at sports events. This mixed approach enabled a comparison of evaluation methodology, and comments on their relative effectiveness for mobile users.

2. Background

2.1. Mobile Personalisation. Personalization techniques can be classified into three different categories [15]: rule-based filtering systems, content-filtering systems, and collaborative filtering systems. Some recent techniques used in collaborative filtering are based on data mining in order to infer personalisation rules or build personalisation models from large data sets.

The main aspects of personalization which are relevant within this paper are *what* is personalized and *how* this is achieved. This paper focuses on content personalization [16], that is, the tailoring of information within a particular node within the human-device navigation space. This form of personalization is based on the key assumption that the optimal content for an individual is dependent on contextual factors relating to the individual, the situation they are in, and the activities they are undertaking—these factors can be used as triggers for the adaptation of content for the individual [17] in order to enhance their user experience.

The personalization framework in this research contains four modules, which (1) cooperate to perform the functions of classification of information, (2) collect relevant contextual factors, and (3) personalize content accordingly (Figure 1).

The overall context of this research (which enabled the field versus laboratory comparison described here) was the investigation of the benefits of personalization of mobile-based content. However, the research project also investigated the benefits/drawbacks of either the user or the system performing personalization of content. Some research favours user-initiated personalization and its focus on the natural intelligence of the user, while others found that system-initiated approaches were more effective for dynamic contexts [18].

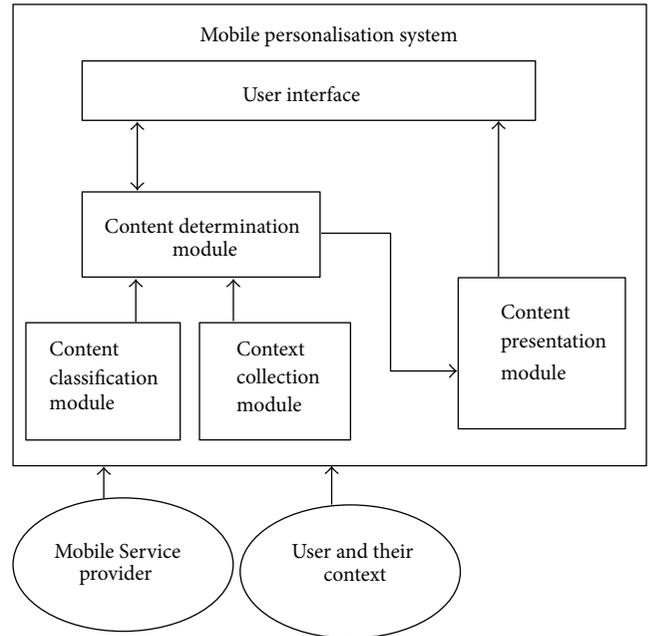


FIGURE 1: Modules forming a mobile personalization application.

2.2. User Experience. “User experience” is a broader concept than usability. As user experience affects the success of a product, studies of user experience should therefore be considered as an important part of the product development process [19]. The importance of user experience stems from mobile devices being personal objects used by individuals with particular social and cultural norms, within an external context defined by their environment.

There is much interest in user experience from design, business, philosophy, anthropology, cognitive science, social science, and other disciplines. Among these, there have been some initial efforts to create theories of user experience. Rasmussen [20] argues that as society becomes more dynamic and integrated with technology, there is a need for a greater multidisciplinary approach in tackling human factors problems. Arhippainen and Tähti [21], in evaluating mobile application prototypes, describe five categories of influences on the user experience, evoked through interaction with an application. These are user factors, social factors, cultural factors, context of use, and product (i.e., application) related factors. They also list specific attributes for each category, such as the age, emotional state of the user, habits and norms as cultural factors, the pressure of success and failure as social factors, time and place as context of use factors, and usability and size as product factors. Similarly, Hassenzahl and Tractinsky [22] define user experience as “a consequence of a user’s internal state (predispositions, expectations, needs, motivation, mood, etc.), the characteristics of the designed system (e.g., complexity, purpose, usability, functionality, etc.) and the context (or the environment) within which the interaction occurs (e.g., organizational/social setting, meaningfulness of the activity, voluntariness of use, etc.)”

Despite the emerging importance of user experience, there are several barriers to using this concept as a key design

objective, and, hence, how they may be employed within either a field or lab-based evaluation process. There is not yet a common definition of user experience because it is associated with a broad range of both “fuzzy” and dynamic concepts, for example, emotion, affect, experience, hedonic, and aesthetics [24]. There is also currently a lack of consensus regarding the constructs that comprise user experience, and how they may be measured [25].

This research follows the approach taken by Arhipainen and Tähti [21] and Hassenzahl and Tractinsky [22] in considering user experience to comprise multiple components, namely: user, social, usage context, cultural, and product. Consequently, multicomponent user experience was measured using 15 agree-disagree scales, addressing the five components. This approach therefore considers *user experience* to be a formative construct that is measured in terms of its components [26].

3. Field-Based Evaluation of Content Personalisation

3.1. Aims. The first empirical study was a field experiment, the aims of which were to evaluate the impact of using a mobile device, and personalising that device, for a spectator within a real sports environment. Field studies typically sacrifice some experimental control in order to maximise the ecological validity of the experiment.

3.2. Method

3.2.1. Setup. The study took place in a sports stadium during a competition involving local football clubs. This competition comprised fast-moving sporting action and a large gathering of spectators, most of whom were unfamiliar with each other. The user experience was therefore typical of that encountered during a large sports event. Information was broadcast to spectators over a public address system and shown on a large display screen in one corner of the stadium.

3.2.2. Experimental Scenarios. The successful use of scenarios takes into account the diversity of contexts encountered by spectators; these were derived from previous studies [27, 28]. Four scenarios were developed including (1) checking the schedule of forthcoming matches and finding one of particular interest; (2) obtaining information on a particular player of interest; (3) reviewing the progress of the current match (dynamic information access); (4) joining a “community” and participating in community-based activities in the stadium.

3.2.3. Prototypes. Prior to the experiments described in this paper, a total of seven field studies were undertaken with spectators at large sports events [27, 28]. These studies found that a large number of contextual factors influenced the design of service/information provision to a spectator, but that three had the greatest potential impact on the user experience. These were the *sporting preferences* of the spectator, their *physical location* in the stadium, and the *event progress*.

Following the previous studies, one paper prototype and two mobile prototypes (a personalised prototype and a nonpersonalised prototype) were developed that provided content to support the experimental scenarios above.

Both mobile prototypes were identical in terms of their functionality and visual design. The personalised prototype had the option to manually configure the presentation of information according to the key contextual factors using the content filtering personalisation technique (description of the technical development is out of scope of this paper). With this prototype, users were asked to set their preferences relating to the sports types and athletes taking part from an extended tree menu structure (Figure 2). As a result, tailored event-based information (e.g., information on athletes) and event schedules were presented to the spectator (Figure 3). In contrast, the nonpersonalised mobile prototype did not require the user to set the personalisation attributes and as a result presented information and services applicable for a more general audience (Figure 4).

The personalised prototype also enabled users to assign themselves to virtual communities with common interests within the stadium using collaborative filtering technique, based on their stated interests. This was via online chat and media sharing within groups defined by their personal preferences. The nonpersonalised mobile prototype enabled the same chat and media sharing, but within a larger group not differentiated according to personal interests.

In addition, a paper leaflet was prepared that was based on the information that a spectator would traditionally get during a real event from posters and programs. It provided information on match schedules and players’ profiles.

3.2.4. Participants. Eighteen participants were recruited by an external agency. Their ages ranged from 18 to 45 years, mean 28.5, with an equal gender split. A range of occupations were represented, including sales, journalist, engineer, teacher, secretary, and accountant; eight participants were university students. A recruitment criterion was that all participants should have undertaken personalisation of their ringtone, screen background, or shortcut keys at least once a week. In addition, all participants had watched a large sports event in an open stadium within the last six months.

3.2.5. User Experience Measurement. The key dependent variable was the user experience that resulted from using the prototypes. User experience was measured in terms of the multidimensional components described in Section 2. User experience was therefore rated by participants in a multicomponent assessment that was theoretically grounded and empirically derived (see the appendix).

3.2.6. Procedure. A pilot study was used to check the timings, refine the data collection methods, and resolve any ambiguities with the instructions and data collection tools. This was carried out in a stadium.

At the beginning of the study, participants were given instruction on how to use the mobile prototype. Participants then undertook the scenario-based tasks, using either the

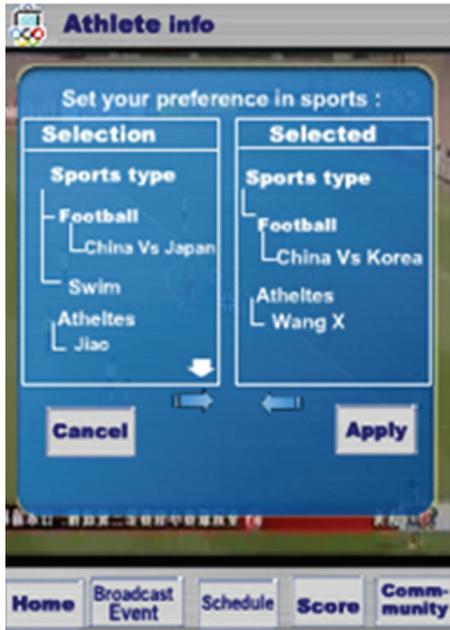


FIGURE 2: Setting personalisation preferences.

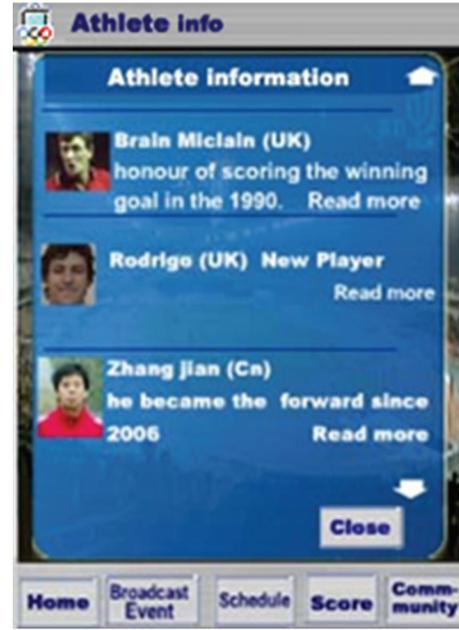


FIGURE 4: Obtaining *general* information on all athletes.



FIGURE 3: Obtaining *personalised* information on particular athletes.

paper-based programme or one of the mobile prototypes. They then completed the 15-item questionnaire that differentiated the five components of user experience.

Finally, they were interviewed in a semistructured format to discuss the experiment design in the field study. The study lasted around 60 minutes for each user.

3.2.7. Data Collection. A multiplicity of data collection methods was used within the study to enable a limited triangulation of subjective rating, verbal report, and observational

data. A video camera was used to record their interactions with the mobile prototypes. Users were encouraged (but not required) to “think aloud” during the trial. A user assessment in relation to overall user experience was captured using six-point agree/disagree rating scales. Direct observation was also used, and informal interaction with the researcher was encouraged during the trial. A posttrial structured interview was video recorded.

3.2.8. Analysis of Data. For quantitative data, Friedman nonparametric tests for three related samples were calculated for the main within-subjects factors. Multiple paired comparisons were undertaken using the technique described in Siegel and Castellan [23, page 180], to take into account the increased likelihood of a type I error with multiple comparisons. The qualitative data (interview transcripts and concurrent verbal reports) and observational data were analyzed using an affinity diagram technique [29] to collate and categorise this data.

3.3. Field Study Results. The field experiments generated large amounts of rich and grounded data in relatively short time. For all tasks, the personalised mobile device consistently generated the highest user experience rating (see Table 1). The nonpersonalised device was consistently worse but still an improvement over the control condition (paper-based programme). The control condition (paper-based programme) was consistently rated poor—a limitation also noted by Nilsson et al. [30] in their field observations of sports events.

To make sure that the user interface itself (rather than the personalization approach) was not majorly influencing the experiment outcome, the user-initiated interface was evaluated by calculating the percentage of tasks completed by participants and analysing user comments. The experiment

TABLE 1: The overall user experience assessment for each task.

Task	Friedman (3 related samples)	Multiple paired comparisons [23, page 180], ($\alpha = .05$)
Checking match schedules	$\chi^2(2) = 31.3$ $P < .001$	$N = 18, R_{\text{Personalized}} - R_{\text{Nonpersonalized}} = 21.5, >Z = 14.36^*$ $N = 18, R_{\text{Personalized}} - R_{\text{paper}} = 29.5, >Z = 14.36^*$ $N = 18, R_{\text{Nonpersonalized}} - R_{\text{paper}} = 8, <Z = 14.36$
Obtaining player information	$\chi^2(2) = 34$ $P < .001$	$N = 18, R_{\text{Personalized}} - R_{\text{Nonpersonalized}} = 17, >Z = 14.36^*$ $N = 18, R_{\text{Personalized}} - R_{\text{paper}} = 34, >Z = 14.36^*$ $N = 18, R_{\text{Nonpersonalized}} - R_{\text{paper}} = 17, >Z = 14.36^*$
Reviewing match progress	$\chi^2(2) = 30.6$ $P < .001$	$N = 18, R_{\text{Personalized}} - R_{\text{Nonpersonalized}} = 16.5, >Z = 14.36^*$ $N = 18, R_{\text{Personalized}} - R_{\text{paper}} = 33, >Z = 14.36^*$ $N = 18, R_{\text{Nonpersonalized}} - R_{\text{paper}} = 16.5, >Z = 14.36^*$
Building a community	$\chi^2(2) = 32.1$ $P < .001$	$N = 18, R_{\text{Personalized}} - R_{\text{Nonpersonalized}} = 17, >Z = 14.36^*$ $N = 18, R_{\text{Personalized}} - R_{\text{paper}} = 34, >Z = 14.36^*$ $N = 18, R_{\text{Nonpersonalized}} - R_{\text{paper}} = 17, >Z = 14.36^*$

* Indicates a significant difference.

recorded approximately 18 hours of video capturing the 18 subjects' interaction steps while completing each task. In summary, 95.5% of the tasks were completed successfully. For the 4.5% of unfinished tasks, 35 usability problems with the personalized mobile prototype were reported.

Participants in the field experiment stressed problems of mobile “use” rather than simply application “usability,” and typically those problems were expressed in the language of the situation. For example, users were concerned about spending too much time personalizing the application during the event (detracting from the event itself) and the font on the interface being too small to read in an open stadium under bright sunlight. The field study also identified issues of validity and precision of the data presented by the application. For example, users were concerned about the reliability of information provided by the prototypes after they found that some player information presented on the mobile device did not match with the real events.

To the participants, the field experiment environment felt fairly informal, and the users talked freely about the use of the application and their feelings. Users expressed how the field experiment allowed them to feel relaxed and able to communicate with the researcher as they undertook the evaluation scenarios. Rather than focusing on interface issues, users generally expressed broader views and were able to give a wide range of evaluation-related information during the experiment, such as expressing contextually related requirements. A particular example of the kind of data generated during the field study was the need for the mobile content (and its delivery) to be highly integrated into the temporal flow of the sporting action.

Using a field experiment approach, it may be possible to obtain a higher level of “realism.” However, this evaluation method is not easy to undertake. Experiments in the field are influenced by external factors, such as the weather, and moreover, it is more difficult to actually collect data from participants. Users were impacted by events happening in the field, such as noise and other disturbances. For example, some participants actually forgot that they were taking part in a

research study—until prompted, they focused their attention on the competition happening in the stadium and were substantially distracted from the field experiment. Flexibility and pragmatism are needed to still collect useful data, while not detracting unduly from the experience generated within the field setting.

4. Lab-Based Evaluation of Personalisation Approach

4.1. Aims. Whereas the previous section has described a field-based study, this section describes a very similar lab study, this time centred around a multievent athletics meeting. The specific objectives of this study were to use a more controlled experimental setup to compare the user experience for a spectator at a large sports event under three conditions: (1) using paper-based (not mobile) content; (2) using a mobile prototype where personalisation parameters were set by the user; (3) using a similar prototype where parameters were set automatically. Similar procedures and data collection methods enable a comparison of findings with the field-based study described in Section 3.

4.2. Method

4.2.1. Setup. This evaluation took place in a usability laboratory in the UK. The usability lab was set up to resemble a part of the sports stadium. To mimic the divided attention that would result from a spectator watching a sporting event, sports footage, including auditory output, was projected onto the front wall of the laboratory. A crowd scene was replicated on each of the two side walls. In contrast to typical mobile applications such as tourist guides [31], a spectator is usually seated and relatively static at a sporting event. Therefore, mobility per se did not need to be incorporated into the experimental environment. A video camera was used to record users' interaction with the prototypes (Figure 5).



FIGURE 5: User study in a usability lab.

4.2.2. Experimental Scenarios. Five scenarios were developed, based on the same key spectator activities as used in the previous field-based experiment. Four of the experimental tasks were the same as those employed in the field experiment. This lab-based study also employed a fifth task which required the participant to select a suitable viewing angle for a mobile broadcast on the device. This enabled the participants to follow the sporting action from wherever they were in the stadium.

4.2.3. Prototypes. Two personalised mobile prototypes were developed using content filtering and collaborative filtering personalisation technique. They shared the same look and feel as those used in the field study described in Section 3.

One prototype enabled *user-initiated* personalisation of content, using an extended menu structure as before (see Figure 2). As a result, this prototype presented event information such as athlete details and event schedules based on the users' settings.

In contrast, the prototype with *system-initiated* personalisation did not require that participants set personalisation parameters. Personalised content was then presented automatically to the participant using the content and collaborative filtering personalisation technique.

As for the field experiment, a control condition was included that was representative of a nondigital event programme that a spectator would typically have. This was a paper leaflet that provided information on competition times and athlete information.

4.2.4. Participants. A different cohort of eighteen participants took part in the study, again split equally male-female, aged between 18 and 38, and with various occupations. As in the previous field-based study, all participants had regular experience of personalising mobile devices and had attended a large sports event within the last six months.

4.2.5. Procedure and Analysis of Data. A pilot study was used to maximise the realism of the simulation and ensure that the data collection methods that were used during the field trial were effective within a laboratory setting. Two key changes were made: rearrangement of speakers to broadcast audience noise; the side projection of a video of the audience was replaced with large-scale posters. This enhanced the perceived social atmosphere within the laboratory without

unnecessarily distracting the participant from the main projected view (the sporting action).

The procedure followed that used for the field experiment described in Section 3. Each task was completed in turn, with the three personalisation conditions counterbalanced across participants within each task. The design of the lab experiment was also discussed with participants at the end of the study. The study lasted approximately one hour for each participant.

As for field study, the data comprised quantitative rating scale data, concurrent verbal reports and posttrial interview data, and observational data. This was analysed as described previously.

4.3. Lab Study Results. Across all tasks, there were clear advantages to having personalised content delivered over a mobile device. The control condition (representing that present at current stadium environments) was significantly worse in all scenarios (see Table 2). Table 2 also indicates that in terms of the impact on user experience, neither user nor system-initiated personalisation emerged as a *single* best approach across the range of tasks studied.

The experiment recorded approximately 18 hours of video during the lab study. In summary, 92% of the tasks were completed successfully. For the 8% of unfinished tasks, 42 usability problems with the personalized mobile prototype were reported. In general, participants considered the user interface of both personalized mobile prototypes easy to use.

The laboratory study was relatively easy to undertake and collect data from. During the lab study, participants quickly revealed considerable information about how a spectator uses the prototypes. The lab environment also offered more control over the conditions for the experiment. In comparison to the field study, participants were more focused on the experiment being undertaken, and were not influenced by external factors, such as weather, noise, or others disturbances from the sporting environment.

There were some drawbacks to the lab experiment, including only a limited representation of the real world, and greater uncertainty over the degree of generalization of results outside laboratory settings. This laboratory study tried to "bring" the large sporting event into the experiment by carefully setting up the lab to resemble a stadium, designing scenarios based on previous studies of context within large sports events, and involving users who were familiar with the usage context. It also addressed the issue of users' divided attention by requiring subjects to watch a sport event video which was projected on the front wall of the lab room while performing the scenario-based tasks with the mobile prototypes. As a result, participants were able to identify some context related problems (e.g., the font was too small to read in an open stadium) during the lab experiment. In addition, participants expressed their concerns with using the personalized prototypes from contextual and social perspectives, including concerns over spending too much time personalizing the device during the event, and therefore actually missing some of the sporting action.

TABLE 2: The overall user experience assessment for each task.

Task	Friedman (3 related samples)	Multiple paired comparisons [23, page 180] ($\alpha = .05$)
Follow sporting action	$N = 18,$ $\chi^2(2) = 27.5,$ $P < .001$	$N = 18, R_{\text{system-initiated}} - R_{\text{use-initiated}} = 15, >Z = 14.36^*$ $N = 18, R_{\text{system-initiated}} - R_{\text{paper}} = 49, >Z = 14.36^*$ $N = 18, R_{\text{user-initiated}} - R_{\text{paper}} = 34, >Z = 14.36^*$
Obtaining athlete information	$N = 18,$ $\chi^2(2) = 29.4,$ $P < .001$	$N = 18, R_{\text{system-initiated}} - R_{\text{use-initiated}} = 2, <Z = 14.36$ $N = 18, R_{\text{system-initiated}} - R_{\text{paper}} = 16, >Z = 14.36^*$ $N = 18, R_{\text{user-initiated}} - R_{\text{paper}} = 18, >Z = 14.36^*$
Reviewing athletics event results	$N = 18,$ $\chi^2(2) = 35.5,$ $P < .001$	$N = 18, R_{\text{system-initiated}} - R_{\text{use-initiated}} = 17.5, >Z = 14.36^*$ $N = 18, R_{\text{system-initiated}} - R_{\text{paper}} = 35, >Z = 14.36^*$ $N = 18, R_{\text{user-initiated}} - R_{\text{paper}} = 17.5, >Z = 14.36^*$
Building a community	$N = 18,$ $\chi^2(2) = 29.6,$ $P < .001$	$N = 18, R_{\text{system-initiated}} - R_{\text{use-initiated}} = 6, <Z = 14.36$ $N = 18, R_{\text{system-initiated}} - R_{\text{paper}} = 24, >Z = 14.36^*$ $N = 18, R_{\text{user-initiated}} - R_{\text{paper}} = 30, >Z = 14.36^*$
Checking event schedules	$N = 18,$ $\chi^2(2) = 34.5,$ $P < .001$	$N = 18, R_{\text{system-initiated}} - R_{\text{use-initiated}} = 16.5, >Z = 14.36^*$ $N = 18, R_{\text{system-initiated}} - R_{\text{paper}} = 34.5, >Z = 14.36^*$ $N = 18, R_{\text{user-initiated}} - R_{\text{paper}} = 18, >Z = 14.36^*$

* Indicates a significant result.

The lab experiment did not allow users to feel relaxed during the experimental procedure. Participants acted more politely during the study, and they pointed out that they were uncomfortable about expressing negative feelings about the applications. In one example, when interviewing a participant about aspects of his user experience, he generally stated that it was fine. However, when presented with the Emotion Cards (a group of cartoon faces that were used to help promote discussion of affective aspects of interaction), he tended to pick up one emotion card and talked a lot about concerns over the time and effort required to manually personalize the application, without feeling that he was being overly critical.

5. Discussion and Conclusion

In recent years, there has been much debate on whether mobile applications should be evaluated in the field or in a traditional lab environment, issues including users' behaviour [11]; identification of usability problems [3]; the experiment settings [5, 8, 32]; the communication with participants [33]. This research enabled a comparison between field and laboratory experiments, based on similar users, mobile applications, and task-based scenarios. In particular, a similar user-initiated prototype was used in each, even though the field experiment took place at a football competition, and the lab experiment recreated an athletics meeting.

The number of usability problems identified during both the field and lab experiments was similar (over all participants, when using the user-initiated prototype, there were 35 usability problems identified in the field setting, and 42 found in the lab setting). These findings are consistent with those of Kjeldskov et al. [3]. They specifically compared lab and field-based usability results and found that the difference in effectiveness of these two approaches was nonsignificant

in identifying most usability problems. Also, some context related problems, such as the font being too small to read in an open stadium, were identified in both experiment settings. However, some key differences in the effectiveness of the field and laboratory approaches were found; the lab experiment identified problems related to the detail of the interface design, for example, the colours and icons on the interface; the field experiment identified issues of validity and precision of the data presented by the application when using the application in a stadium. The field experiment also stressed the problems of mobile "use" rather than simply application usability, and typically these problems were expressed in the language of the situation [11].

An analysis of positive versus negative behaviours [11] was undertaken. This data included verbal reports and rating scale data according to the user experience definitions. Accepting the limitations of a direct comparison, participants reacted more negatively in the laboratory setting when completing similar tasks (using the similar user-initiated personalisation approach). In the field, individuals were influenced by the atmosphere surrounding the sports event, and this resulted in an enhanced user experience. In addition, they focused more attention on the actual usage of personalisation on the mobile device, instead of issues to do with the interface. The lab setting was less engaging than the field setting; participants were more likely to be critical, and in general they took longer to perform certain tasks by focusing (and commenting) on interface issues such as fonts and colours used.

The field experiment was more difficult to conduct than the lab experiment, a point noted by many authors, including Kjeldskov et al. [3] and Baillie [8]. Confounding factors were present, for example, variations in the weather and noise from other spectators. In addition, although it was desirable that participants engaged in the sporting action, spectators' foci of attention could not be controlled and was difficult to predict.

Some participants were “distracted” from the experimental tasks, and this did not occur during the laboratory setting. The greater control possible with a laboratory study (as discussed by a range of authors, including [2, 5, 8]) was clearly demonstrated during these studies.

Where there is an interest in qualitative data, good communication between the researcher and participants is vital. The field experiment provided a more open and relaxed atmosphere for discourse. Users more freely discussed their use of the mobile applications, their underlying beliefs, and attitudes that arose during the study. The field experiment helped the communication tensions with the participants as they felt they were not being directly examined. As well as generally promoting the generation of qualitative data, the field experiments encouraged the expression of broader, as well as more contextually relevant views. An example is the identification of contextually dependent requirements, which occurred much less frequently during the lab-based studies.

Some suggestions for user impact assessment with mobile devices can be made based on the findings of this study. A lab experiment is recommended when the focus is on the user interface and device-oriented usability issues. In such cases, a well-designed lab study should provide the validity required, while being easier, quicker, and cheaper to conduct. However, the results suggest that a field experiment is better suited for investigating a wider range of factors affecting the overall acceptability of mobile services, including system functions and impact of usage contexts. Where open and relaxed communication is important (e.g., where participant groups are naturally reticent to communicate), this is more readily promoted by the use of a field study.

The natural tension between a deductive and inductive research design was also apparent. This research in general was essentially deductive, since it set out to explain causal relationships between variables, operationalized concepts, controlled variables, and used structured and repeatable methods to collect data. However, the field study in particular also comprised an inductive element, as there was a desire to understand the research context and the meanings attached to events in order to help design the lab study. Van Elzakker et al. [2] underline how it is often desirable to combine approaches within the same study. The undertaking of a field experiment followed by a lab experiment was an attempt at multiplicity of methods from an essentially deductive viewpoint. This recognised that the natural research process is often that of moving from a process of *understanding* to one of *testing*, whilst attempting to avoid the unsatisfactory middle ground of user evaluations that are divorced from any underlying research objectives.

Appendix

Likert Items Used to Assess User Experience

In all cases, participant responses were based on a six-point scale ranging from 1 (strongly disagree) to 6 (strongly agree).

User Aspect:

- (1) I feel happy using [A/B/C] during the event.

- (2) My expectations regarding my spectator experience in the stadium are met using [A/B/C].
- (3) My needs as a spectator are taken into account using [A/B/C].

Social Aspect:

- (4) Using [A/B/C] helps me feel I am communicating, and sharing information with others in the stadium.
- (5) The [A/B/C] helps me create enjoyable experiences within the stadium.
- (6) The [A/B/C] helps me share my experiences with others within the stadium.

Usage Context Aspect:

- (7) The [A/B/C] provides me with help in the stadium while watching the sporting action.
- (8) The [A/B/C] provides me with information about other spectators in the stadium.
- (9) The [A/B/C] helps provide me with a good physical and social environment in the stadium.

Culture Aspect:

- (10) The [A/B/C] helps me feel part of a group.
- (11) The [A/B/C] helps me promote my group image.
- (12) The [A/B/C] helps me interact with my group.

Product Aspect:

- (13) The [A/B/C] is useful at the event.
- (14) The [A/B/C] is easy to learn how to use.
- (15) The [A/B/C] is easy to use.

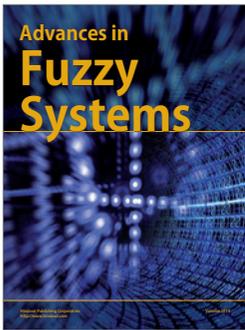
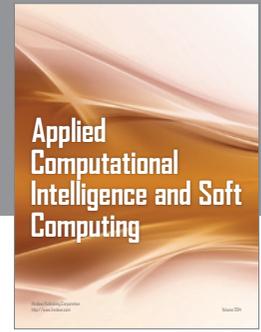
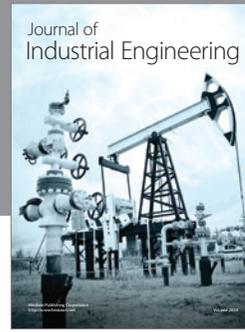
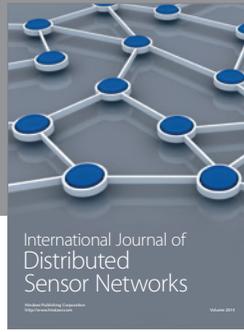
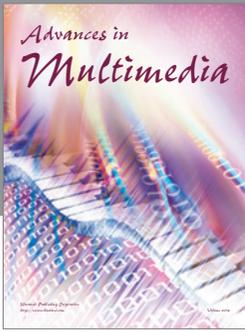
Acknowledgments

This work has been carried out as part of the Philips Research Programme on Lifestyle. The authors would like to thank all the participants who were generous with their time during this study.

References

- [1] D. Raptis, N. Tselios, and N. Avouris, “Context-based design of mobile applications for museums: a survey of existing practices,” in *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services*, 2005.
- [2] C. P. J. M. Van Elzakker, I. Delikostidis, and P. J. M. Van Oosterom, “Field-based usability evaluation methodology for mobile geo-applications,” *Cartographic Journal*, vol. 45, no. 2, pp. 139–149, 2008.
- [3] J. Kjeldskov, M. B. Skov, B. S. Als, and R. T. Høegh, “Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the Field,” in *Proceedings of the 6th International Mobile Conference (HCI '04)*, Springer, 2004.

- [4] X. Sun, D. Golightly, J. Cranwellý, B. Bedwell, and S. Sharples, "Participant experiences of mobile device-based diary studies," *International Journal of Mobile Human Computer Interaction*. In press.
- [5] J. Kjeldskov and J. Stage, "New techniques for usability evaluation of mobile systems," *International Journal of Human Computer Studies*, vol. 60, no. 5-6, pp. 599-620, 2004.
- [6] P. Christensen, M. Romero, M. Thomas, A. S. Nielsen, and H. Harder, "Children, mobility, and space: using GPS and mobile phone technologies in ethnographic research," *Journal of Mixed Methods Research*, vol. 5, no. 3, pp. 227-246, 2011.
- [7] E. G. Coleman, "Ethnographic approaches to digital media," *Annual Review of Anthropology*, vol. 39, pp. 487-505, 2010.
- [8] L. Baillie, "Future Telecommunication: exploring actual use," in *Proceedings of the International Conference on Human-Computer Interaction*, IOS press, 2003.
- [9] M. Esbjörnsson, B. Brown, O. Juhlin, D. Normark, M. Östergren, and E. Laurier, "Watching the cars go round and round: designing for active spectating," in *Proceedings of the Conference on Human Factors in Computing Systems (CHI '06)*, pp. 1221-1224, New York, NY, USA, April 2006.
- [10] X. Sun, "User requirements of personalized mobile applications at large sporting events," in *Proceedings of the IADIS Multi Conference on Computer Science and Information Systems (MCCSIS '10)*, Freiburg, Germany, 2010.
- [11] H. B. L. Duh, G. C. B. Tan, and V. H. H. Chen, "Usability evaluation for mobile device: a comparison of laboratory and field tests," in *Proceedings of the 8th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '06)*, pp. 181-186, September 2006.
- [12] A. Pirhonen, S. Brewster, and C. Holguin, "Gestural and audio metaphors as a means of control for mobile devices," in *Proceedings of the Conference on Human Factors in Computing Systems (CHI '02)*, pp. 291-298, New York, NY, USA, April 2002.
- [13] R. Graham and C. Carter, "Comparison of speech input and manual control of in-car devices while on-the-move," in *Proceedings of the 2nd Workshop on Human Computer Interaction with Mobile Devices (HCI '1999)*, Edinburgh, UK, 1999.
- [14] J. Lai, K. Cheng, P. Green, and O. Tsimhoni, "On the road and on the web? Comprehension of synthetic and human speech while driving," in *Proceedings of the Conference on Human Factors in Computing Systems (CHI '01)*, pp. 206-212, New York, NY, USA, April 2001.
- [15] B. Mobasher, "Data mining for personalization," in *The Adaptive Web: Methods and Strategies of Web Personalization*, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds., pp. 1-46, Springer, Berlin, Germany, 2007.
- [16] D. Wu, I. Im, M. Tremaine, K. Instone, and M. Turoff, "A framework for classifying personalization schemes used on e-commerce websites," in *Proceedings of the 36th Hawaii International Conference on Systems Sciences (HICSS '03)*, p. 222b, Maui, Hawaii, USA, 2003.
- [17] L. Norros, E. Kaasinen, J. Plomp, and P. Rama, *Human-Technology Interaction Research and Design. VTT Roadmap*, VTT Research Notes 2220, Espoo, Finland, 2003.
- [18] E. Frias-Martinez, S. Y. Chen, and X. Liu, "Evaluation of a personalized digital library based on cognitive styles: adaptivity versus adaptability," *International Journal of Information Management*, vol. 29, no. 1, pp. 48-56, 2009.
- [19] J. Dewey, *Art as Experience*, Perigee, New York, NY, USA, 1980.
- [20] J. Rasmussen, "Human factors in a dynamic information society: where are we heading?" *Ergonomics*, vol. 43, no. 7, pp. 869-879, 2000.
- [21] L. Arhipainen and M. Tähti, "Empirical Evaluation of User Experience in Two Adaptive Mobile Application Prototypes," in *Proceedings of the 2nd International Conference on Mobile and Ubiquitous Multimedia*, Luleå, Sweden, 2003.
- [22] M. Hassenzahl and N. Tractinsky, "User experience—a research agenda," *Behaviour and Information Technology*, vol. 25, no. 2, pp. 91-97, 2006.
- [23] S. Siegel and N. J. Castellan, *Nonparametric Statistics For the Behavioral Sciences*, McGraw-Hill, New York, NY, USA, 1988.
- [24] E. Law, V. Roto, A. P. O. S. Vermeeren, J. Kort, and M. Hassenzahl, "Towards a shared definition of user experience," in *Proceedings of the 28th Annual Conference on Human Factors in Computing Systems (CHI '08)*, pp. 2395-2398, Florence, Italy, April 2008.
- [25] E. L. C. Law and P. van Schaik, "Modelling user experience—an agenda for research and practice," *Interacting with Computers*, vol. 22, no. 5, pp. 313-322, 2010.
- [26] C. H. Lin, P. J. Sher, and H. Y. Shih, "Past progress and future directions in conceptualizing customer perceived value," *International Journal of Service Industry Management*, vol. 16, no. 4, pp. 318-336, 2005.
- [27] X. Sun and A. May, "Mobile personalisation at large sports events—user experience and mobile device personalisation," in *Human-Computer Interaction 2007*, vol. 11, Springer, Berlin, Germany, 2007.
- [28] X. Sun and A. May, "The role of spatial contextual factors in mobile personalization at large sports events," *Personal and Ubiquitous Computing*, vol. 13, no. 4, pp. 293-302, 2009.
- [29] J. Hackos and J. Redish, *User and Task Analysis For Interface Design*, John Wiley & Sons, New York, NY, USA, 1998.
- [30] A. Nilsson, U. Nuldén, and D. Olsson, "Spectator information support: exploring the context of distributed events," in *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work*, 2004.
- [31] B. Oertel, K. Steinmuller, and M. Kuom, "Mobile multimedia services for tourism," in *Information and Communication Technologies in Tourism 2002*, K. W. Wober, A. J. Frew, and M. Hitz, Eds., pp. 265-274, Springer, New York, NY, USA, 2002.
- [32] D. D. Salvucci, "Predicting the effects of in-car interfaces on driver behavior using a cognitive architecture," in *Proceedings of the Conference on Human Factors in Computing Systems (CHI '01)*, pp. 120-127, New York, NY, USA, April 2001.
- [33] A. Kaikkonen, A. Kekäläinen, M. Cankar, T. Kallio, and A. Kankainen, "Usability testing of mobile applications: a comparison between laboratory and field testing," *Journal of Usability Studies*, vol. 1, pp. 4-16, 2005.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

