

# *Factors Contributing to Low Usage of Mobile Data Services: User Requirements, Service Discovery and Usability*

**Stavros Garzonis**  
**Eamonn O'Neill**

*Computer Science Department, University of Bath, Bath, BA2 4BU,  
UK*

*Email: {S.Garzonis, eamonn}@cs.bath.ac.uk*

*Fax: +44 (0)1 225 38 3493*

**There is evidence that most users do not exploit the capabilities offered by mobile data services. We propose that in addition to user requirements and usability, service discovery adds an extra dimension that may explain this lack of use. We conducted a 3-way evaluation study which combined expert evaluation, a diary study and cooperative evaluation to investigate the factors affecting low usage of data services. The task of selecting the appropriate service to solve everyday problems was demonstrated to be difficult using existing mobile services. The results indicated that service discovery plays a central role in determining user satisfaction and performance.**

## **Author Keywords**

Service discovery, mobile data services, usability evaluation.

## **1 Introduction**

Although mobile telephony has pervaded many people's lives, its widespread use is still restricted to two main services: voice calls and text messaging. The availability of mobile data services has been increasing in the last few years, but according to Strategy Analytics predictions [Strategy Analytics 2006], SMS will remain the dominant data application globally for 2006, accounting for 56% of end user spend on mobile data services. Second most popular services are downloads of ringtones, games and wallpapers or "Person-to-application" SMS. This leaves information retrieval services with very low usage. This paper reports a study that investigated the reasons for this lack of usage of mobile data services.

Many researchers have explored the usability issues of mobile devices, many of which arise from limited input and output capabilities. First, the small physical size of

mobile phones typically restricts the input to a small set of physically small buttons. This might have been adequate when phones were used only for voice calls but the increasing functionality of phones has encouraged alternative input methods. The original multi-tap method for texting through the 9 button keypad has been augmented by predictive text algorithms (e.g. [Silfverberg et al. 2000]), optimization of character placement (e.g. [Gong & Tarasewich 2005]) or number of keys (e.g. [MacKenzie 2002]) and most recently in PDAs and smartphones with stylus input (e.g. [Kristensson & Zhai 2005]) and miniature QWERTY keyboards (e.g. [Clarkson 2005]). Other approaches include gestures on the surface of the device [Pirohen et al. 2002], gestures with the device in three dimensional space [Patel et al. 2004] and device tilting for typing [Wigdor & Balakrishnan 2003]. All these methods have indeed improved input efficiency but have mixed results in other aspects of usability, such as learnability and satisfaction. Secondly, the physical limitations of mobile displays limit the amount of information that can be conveyed. Techniques that have been applied to overcome this are the use of sound [Brewster 2002], rendering and restructuring the content that would normally appear in larger displays [Lemlouma & Layaida 2005] or cloning the small display on nearby bigger displays [Miyahara et al. 2005].

However, these studies do not entirely explain why people would use the same, perhaps unusable, devices in the same distracting environments for text messaging or voice calls but not for data services. One possible explanation could be based on user requirements. If the demand for voice calls and text messages is strong enough to tolerate these problems, people will still use them. On the other hand, if frustration overcomes the benefits of using data services, people will refuse to use them. This threshold is related to the question that any user may ask herself: "Is it worth using?". In marketing terms this is referred as "perceived value" of a product or service. Although the perceived value of mobile data services is also influenced by factors such as financial costs and marketing techniques, these are outside the scope of this paper.

Identifying user requirements and designing usable interactions have long been acknowledged for their role in successful software design. Although these two factors still apply to designing successful mobile data services, there is one significant difference when compared to conventional software design for the desktop environment. On a desktop PC, a user is typically aware of most of the services that are provided by the PC. These often take the form of applications that the user has personally installed. In contrast, in the mobile environment, users may often be unaware of the services that are available through their handset or even how to find them. Services come with different labelling, structures, content and initiation processes. For example, to download games on the Orange UK network, one uses "Orange Multi Media services" [Orange 2006], while with Vodafone UK one uses "Vodafone Live!" [Vodafone 2006] – and in each of these cases, this is just the first step to accessing a game.

This lack of standardisation is similar to the vast and unregulated services that one can find on the Internet. However, interacting with a PC in order to discover such services is more tolerable. Users can find their desired service by applying a variety of techniques, such as browsing, using search engines or following indices of services. However, these explicit user actions are harder to perform through the limited interaction methods of the mobile device. Therefore, our hypothesis is that the process of service discovery on mobile devices will be a major factor contributing to low usage of data services. If this is true, awareness of the available services, their approximate content and the procedure for invoking them will play a central role in the level of usage of the services. We define this knowledge as the *Service Awareness* that any user has at any given moment. Although this awareness can be raised through formal and informal advertising (e.g. word of mouth), our concern is with how context-aware systems could address service awareness more dynamically.

Furthermore, with the growth of smartphones with greater hardware capabilities and software adaptability, services are more likely to become contextually available. Thus, a service may be available or unavailable depending upon the current context. Aspects of context that may determine service availability include personal, spatial and temporal dimensions. Different services will be available to different people in different places and at different times. There is an extensive literature around context aware computing (e.g. [Chen & Kotz 2000], [Dey 2001], [Barkhuus & Dey 2003]) that describes computing systems that can sense and adapt to the dynamically changing environment of the user. The development of context awareness and the contextual availability of services through mobile phones has the potential to render service awareness even more problematic. For example, one might have access to a local voting service upon entering a local forum or the ability to print or exchange documents in *ad hoc* manners depending on the availability of other devices. In these cases, changes of location, social and technical context affect the availability of a service. This would make it even more difficult for a user to maintain awareness of what services are available to him at a particular time in a particular place. To encourage usage of these services, users' awareness of contextually available services must be supported.

In the next section we present related work in the field of mobile evaluation. In the section following, we describe a study that we conducted by combining three evaluation approaches, aiming to investigate reasons for the lack of usage of mobile data services. We then present the results of our study, grouped around three themes: user requirements, service discovery and usability.

## **2 Evaluating Mobile Services**

We are specifically interested in the evaluation of mobile services and not of client applications. Our distinction between service and application draws on Microsoft's definition of a service being "a software-based functionality provided by network servers" [Microsoft 2006]. We view the application as the software running on the

handset and the service as the functionality provided between that application and the server. The focus of this study is not on application attributes such as font size or button functionality but on service requirements, discovery and usability.

In a survey conducted from 2000 through 2002, Kjeldskov and Graham [2003] report that only 41% of mobile-human interaction research involved any evaluation of system designs. These usability evaluations are usually focused on aspects such as devices' limitations [Brewster 2002], applications' user interfaces [Pirohen et al. 2002] or the architecture of prototype systems [Cheverest 2002]. They tend not to focus on the underlying structure of existing and wide-spread services.

Furthermore, there are no specific evaluation methods for mobile devices, with the exception perhaps of the Heuristic Walkthrough [Vetere et al. 2003]. This was an attempt to combine Heuristic Evaluation and Cognitive Walkthrough but it was found inadequate during its evaluation [Vetere et al. 2003]. In the absence of specialised evaluation methods, most researchers apply traditional HCI usability evaluation approaches.

Moreover, the literature is divided regarding the appropriate setting in which to conduct evaluations of mobile applications and services. While desktop environments are typically relatively stable (e.g. level of ambient noise, illumination, device availability), the context of mobile use is bound to be more unpredictable and more attention demanding. In the Kjeldskov & Graham [2003] survey, 71% of the mobile evaluations were performed in lab settings. Some researchers [e.g. Kjeldskov & Skov 2003] have attempted to replicate mobile conditions in the laboratory, where data collection and variable control is easier. On the other hand, others have argued that only field evaluations can identify issues relating to the true context of use where the user's attention may be divided among many things other than interacting with the device [Kjeldskov et al. 2004]. In an attempt to clarify the debate around lab vs. field evaluation for mobile phones, Kjeldskov et al. [2005] compared four different evaluation methods: cooperative field and lab evaluations, heuristic walkthrough and rapid reflection. The service that was evaluated was a location-aware mobile guide supporting the use of public transportation. They found that field and lab evaluations had considerable overlap in the 'critical' and 'serious' usability problems they picked up. Field evaluation only fell short in 'cosmetic' problems, which did not impede performance. In another study comparing 5 different lab-based evaluation techniques, Kjeldskov and Stage [2004] reported that seated participants found more 'cosmetic' usability problems than walking participants. In a review [Garzonis 2005] of the comparison between field and lab evaluations, we have argued that overall, typical lab conditions are better for device-centric usability studies.

In the study reported here, we used a range of evaluation methods and aimed to combine the advantages of both lab and field techniques.

### **3 The 3-Way Study**

The study was a combination of expert evaluation, diary study and field cooperative evaluation. Each of these three approaches addressed some or all of the factors we identified as contributing to the lack of usage of mobile data services: user requirements, usability and service discovery. In this section, we describe the aims and procedure of each part of the study.

#### **3.1 Expert Evaluation**

An expert evaluation was conducted with two goals in mind. First, it focused on mapping the current data services provided through the main mobile web portal of one of the leading global mobile network providers in order to inform task selection for the cooperative evaluation study and to identify the optimum path and completion time for each task. Secondly, the expert evaluation was used to identify specific usability problems.

The evaluator had knowledge of established evaluation methods and was familiar with the mobile device used. However, he was not experienced with the use of mobile Internet or the content of the particular set of data services under investigation. The evaluation was done on a Nokia 6600 handset while pen and paper were used for keeping notes. It lasted approximately 1 hour, with 2 further hours spent on follow-up analysis.

Adopting a cognitive walkthrough approach, the evaluator browsed through the services and developed realistic scenarios, which he then executed using the phone in order to discover the optimum solution. The steps to the optimum solution form the *optimum path* which was recorded in a form of a 'mindmap'. The scenarios reflected potentially everyday situations in order to engage users in the subsequent cooperative evaluation. They were designed to cover a range of 5 different data services: sports information, 2 types of directory information, travel information and the "Find & Seek" service. The latter is a location-aware service designed to provide maps and/or points of interests based on proximity to the user. The optimum paths of the tasks ranged from as shallow as 4 screens to as deep as 11 screens. They included specific parts of the services that were identified by the evaluator as liable to cause confusion. This provided the opportunity to validate the evaluator's findings against the results of the cooperative evaluation study. The easier and shallower tasks were included to counterbalance the study.

#### **3.2 Diary Study**

A 3-day diary study was conducted primarily to elicit information regarding user requirements for mobile services and secondly on service discovery. We designed our diary study drawing on related work by Rieman [1993]. The diary study gives an insight into peoples' ideas, problems and solutions in their own environment. It may be viewed as an informal equivalent of participatory design [Rieman 1993] and can inform future experimental designs.

There were 12 participants involved; 7 women and 5 men, aged from 24 to 36 (mean 27.7, SD 3.6). They all had strong computer science backgrounds with 10 of them having considerable experience in HCI (at MSc or PhD level). Rating their mobile phone usage from 1 to 5 (1:less than once a week, 5: many times per day), they produced a mean of 4.1 (SD 1.4). They were spread among 5 different major network providers and have been using a mobile phone for at least 4 years.

The procedure was explained to them in advance and they confirmed they had fully understood it. One question was emailed to the participants every day and they were asked to answer it during the day along with any questions they had received on previous days. Hence, on the first day, they had one question to consider. On the second day, they had two questions to consider, and on the third day, they had all three questions to consider. Therefore, the first question was most favoured in terms of time spent to answer and the third was least favoured. This was done for two reasons. First, the first question was the most open-ended and it would be compromised if it was given along with the rest of the questions. Similarly, the second question would be compromised by the third question. Secondly, the last question addressed our secondary goal of investigating what services users actually want. We therefore aimed for richer data on the first two questions, which addressed our primary goal of investigating users' preferences for service discovery techniques.

On the first day the participants were asked to use their imagination as to how their phone could help them in various situations throughout the day. Although we did not want to restrict the users' imagination, we advised them not to be totally unrealistic (e.g. "I want my phone to fly" or "to brew coffee"). Otherwise, the participants were free to 'invent' any service that could support their daily interactions and activities. This question aimed to capture users' needs regarding mobile applications. In particular, we were interested to see if and how often any ideas on context-aware services would come up before we introduced the notion to the participants with the second question.

The second question introduced the users to context-awareness by asking them what they would want their phones to be able to sense and what kinds of services could exploit that to their advantage. This question aimed to elicit information on what types of context aware services are most wanted in order to allow comparison with such existing services.

On the last day they were asked for their ideas about service discovery. The term "service initiation" was used in order to avoid confusion with the more technologically-oriented term "service discovery", as it is often used to mean devices-discovering-devices.

Each day of the diary study, the participants were asked to attempt to answer the question(s) early in the morning and then keep it/them in mind for the duration of the day. When they had an idea in relation to one of the questions, they were instructed to note it down along with information regarding their location, their main or parallel task(s) at the time, their social context and the time of the day. This was intended to

help us to infer the potential context of use of the proposed service. At the end of each day, they were instructed to aggregate the findings of the day and expand on their ideas if necessary. They were allowed to expand on previous ideas but they were asked to keep the original script intact. For example, one participant could add details of an idea he had the first day but on the second's day sheet. Thus, the original wording of the idea was preserved in question's 1 sheet.

### **3.3 Cooperative Evaluation**

#### **3.3.1 Background**

The third part of the study was a cooperative evaluation, which was based on work by Wright & Monk [1991]. They have developed the think-aloud method by adopting an approach where the participant is told to think of him or herself as a co-evaluator of the system. The participant is not only asked to "think aloud" but more specific questions are asked and the evaluator encourages the participant to provide commentary. Although this involves the risk of the evaluator biasing the participant in his views about the system, if performed carefully, cooperative evaluation can evoke rich user data. Wright & Monk [1989] argued that evaluators need to look for two types of evidence of usability problems: critical incidents and breakdowns. Critical incidents occur when the participant is deviating from the optimum path to the task solution. Breakdowns were introduced by [Winograd & Flores 1987] and can be identified "as any point at which the user's comments indicate that the system has become part of his or her subjective experience" [Wright & Monk 1991]. In other words, breakdowns occur when the participant is focusing on the technological medium instead of the task at hand.

Users' comments about the system and the identification of critical incidents aid breakdowns [Wright & Monk 1991]. Cooperative evaluation can provide useful data on the user's mental model. When these mismatches between the user's and designer's mental models of the system are taken into consideration, they can inform redesign of the system to better match users' perception of the service structure.

An alternative to cooperative evaluation is the use of retrospective verbal protocol. In this type of evaluation, the users do not interrupt the task execution for comments but are instead de-briefed after the task. One of the main benefits of this approach is that they can expand as much as they want, without compromising their performance. However, the post hoc commentary produced after task completion misses the more intuitive user reactions to the system. Also, users partially forget past interactions and reconstruct them, providing inaccurate information. We chose to apply cooperative evaluation followed by a post-test questionnaire due to time constraints and because performance was secondary to user comments on mobile data services.

The setting was in a busy café with the participant and the evaluator sitting at a table, having coffee. We chose to do a field evaluation as the lab is a particularly unnatural setting for use of mobile phone evaluation [Kiildskov et al. 2004]. Although mobile devices can be used in almost any setting, when interacting with them demands higher

cognitive load, people attempt to reduce other cognitive distractions such as walking [Kaikkonen et al. 2005]. Hence, we chose a setting where the participants were seated. Other distractions such as ambient noise and social presence were present, providing a level of ecological validity. At the same time, having the participants seated facilitated data collection and more closely resembled a lab evaluation, mitigating the Achilles heel of field evaluation in identifying cosmetic problems.

### 3.3.2 Procedure

As with the expert evaluation, the cooperative evaluation used the main mobile web portal of a leading network provider. The users' actions and interaction with the evaluator were recorded by video and audio and by the evaluator's taking notes. For each task that the users were asked to perform, an optimum solution path had been identified through the expert evaluation. In the cooperative evaluation, the users' actions were recorded and compared to the optimum solution path. A critical incident was identified whenever a user deviated from that path. We extended the analysis to look not just at the initial deviation but also at the users' subsequent actions.

There were twelve participants, 6 women and 6 men, aged from 23 to 36, average age 27.8 (SD 3.5). These were the same twelve participants who took part in the diary study apart from one substitution, as one participant had to be discarded due to a technical problem. This means they had the same strong computer science and HCI background, very frequent mobile phone usage but only 17% of them regularly used the phone for data services.

The phone used was a Nokia 6600 handset attached to a lightweight L-shaped wooden base with a web camera (PC Line) attached to the other end (Figure 1). The camera captured the mobile screen and the buttons just above the keypad (the keypad was not needed to complete the tasks). The camera had a built in microphone and was connected by a cable to a notebook computer resting on the table (Figure 2). A video camera was placed about 3 metres away from the table, capturing users' hand movements, body posture and facial expressions.



Figure 1: The phone with the camera attached



Figure 2: The cooperative evaluation setting

Participants who were not familiar with the phone were given a short tutorial on the buttons' functionality. To familiarize themselves with the device, they were asked to send a text message. Afterwards, they were presented with the instructions of the cooperative evaluation on the screen of the notebook computer. Each scenario was read to them before they attempted it and was visible on the screen throughout the process. This was to ensure that participants could revisit the question in case they lost focus or partially forgot their goal. There were 9 scenarios and each participant was randomly allocated to 3. The scenarios included finding walking directions (e.g. "Find the nearest Chinese restaurant" or "Find the nearest pharmacy"), and retrieving information regarding sports (e.g. "Find when is the next F1 race"), weather (e.g. "Find out how the weather is going to be tomorrow in Bristol") and entertainment (e.g. "Find a cinema around you that shows movie Star Wars III").

## **4. Results**

We analysed the results in three categories: user requirements, service discovery and interaction usability. Although these categories are strongly linked and sometimes have considerable overlap, they follow the logical course of actions to achieve a particular goal. First, the user forms the intention to reach that goal using her mobile device. Then she has to discover or recall the appropriate service and finally she interacts with that service to achieve the desired goal.

### ***4.1 What People Need***

As mentioned, one of the main goals of the diary study was to determine whether context aware services were required by users. In the first question of the diary study, 67% of the users came up with at least one idea for a context-aware service. If we exclude the participants who were "using their phone only for calls and texts", who produced no ideas throughout the study, this number comes up to 80%. In total, 57% of the ideas produced by users throughout the study were context-aware related.

As participants provided multiple ideas on different days, the data can be organised in two ways: the ideas that were proposed by most participants (popularity) and the ideas that came up most frequently regardless of the participant (frequency).

The most popular idea, produced by 33% of the participants, was a context-aware service related to bus or train notification systems. The common theme included user notification about the actual location of the vehicle and the expected time of its arrival. The second most popular ideas, produced respectively by 25% of the participants, were related to voice instructions, automatic mobile profile switching (e.g. from “silent” to “outdoors”) and trigger related events (i.e. notification of specific suggestions/actions only in the appropriate context). It is interesting to note that 3 out of the 4 most popular ideas were for context-aware services. Overall, 57% of the ideas were context-aware.

The most frequent idea however was not context aware. It represented 12% of all ideas and was related to voice or sound instructions and dictation. Second most frequent (6%) were ideas related to data synchronisation, profile switching and bus/train live information. Table 1, summarises the findings on ideas’ popularity and frequency.

Some further data regarding user requirements could be derived from user comments during the cooperative evaluation. Some of them were noted by the evaluator during the evaluation and the post-test questionnaire. However, the audio recordings were partially irretrievable due to hardware malfunction and therefore the comments do not have sound statistical value. For example, we know that at least 75% of the participants explicitly stated their dissatisfaction with the network connection speed. However, they do not represent the entire sample since some of the audio record was missing.

Idea	Frequency	Popularity
Voice Instructions	11.59%	25%
Bus/train Dynamic Information	5.8%	33.33%
Synchronisation	5.8%	16.67%
Profile Switching	5.8%	25%
Contextual Triggers	4.35%	25%

Table 1: Most frequent and popular ideas

#### ***4.2 What People Can Discover***

In this study, the term service discovery is limited to the action of choosing the appropriate data service among a set of services available to the network subscribers. When connected to the data service network, this set of services is visible as a grid of

icons on the phone display. Connection could be achieved via one of the soft buttons on the starting screen, or through the phone menu, under the label “Services”. The rest of the phone menu items (e.g. Camera, Messages) were considered as applications and not services. Although sometimes more than one service could lead to one of the desired goals, we define as *optimum* the service that requires the least steps to achieve the goal. All of these optimum services are on the top level of the menu and thus they can be accessed by just one click, or two clicks when not connected to the server. The time needed for an expert to access the optimum service ranges from 0 to 25 seconds, depending on the task.

The analysis of the cooperative evaluation indicated that overall, in 92% of the trials the participants interacted with the optimum service. Only 33% used the branded shortcut on the starting screen in their first attempt to connect to the server. The rest spent on average 6.4 (SD 6.1) minutes and went through 18 (SD 16.9) pages simply to connect to the server. The pages they viewed before arriving at the desired service were 9.4 (SD 10.9) on average, which is 500% more than the optimum. On average they chose 1.8 (SD 1.5) services per task or 1.7 (SD 1.2) without repetitions. They also spent 3.1 (SD 4.0) minutes to reach the service they considered as most appropriate. If the participant failed to complete the task, we counted as most appropriate the last service (s)he used. The time spent was as much as 1403% more than the optimum time needed. A service was counted twice if the participant gave up on that service only to return back to it after having selected other services. Table 2 summarises the results that indicate service discovery performance.

Measure to Service Discovery	Absolute Value (SD)	Increase from Optimum
Pages to Service Discovery	9.43 (10.93)	500.18%
Services selected per task	1.84 (1.49)	84%
Minutes to Service Discovery	3.13 (4.04)	1403%

Table 2: Measures to service discovery

Overall, 82% of the tasks were completed successfully and 95% of them were achieved through the optimum service. However, on average 1.7 (SD 1.4) services were chosen in the process. In 53% of the successful tests, the participants finished the task by making use only of the optimum service.

### **4.3 What People Can Use**

The results presented here cover user interaction during both service initiation and service usage. Times and number of pages visited were counted from the beginning of the interaction until the completion of each task. We argue that this count of pages visited is a more appropriate measure than just a deviation from the optimum path (i.e.

critical incident). User performance is more accurately measured by the number of the extra pages they visited.

The minimum number of pages the participant needed to visit in order to complete the task defined the optimum depth of the solution. The “Options” button of the handset would bring up a menu of potential actions and it was counted as one extra page even if the participant chose to exit the menu without selecting any of these actions. Selections that were cancelled before the new page was loaded were not counted.

The optimum depth of the solutions was on average 7.1 (SD 2.4). However, the average number of pages visited by the participants was 22.1 (SD 15.3), including repetitions and server error pages. This is as much as 213% more browsing than the optimum. Excluding the server error pages and the repetitions caused by these errors, the average number of pages visited was 21.1 (SD 14.9) and the browsing was 198% of optimum. The optimum path was followed in only 3.6% of the tasks and only by 2 participants. The time needed to complete the tasks was 499% more than the optimum. Table 2 summarises the results that indicate service usage performance.

Measure to task completion	Absolute value (SD)	Increase from Optimum
Pages to task completion	22.15 (15.33)	213%
Minutes to task completion	8.24 (5.45)	499%

**Table 3:** Measures to task completion

One of the most common confusions was the “Options” button (left soft button). 88% of the users misunderstood its function. Also, all of the participants who used the “Search” function within a service were expecting a text box but were faced with lists of options. These same usability issues were brought up during the expert evaluation. The two most persistent flaws were poor labelling and vague error messages.

## 5 Discussion

We have argued that the low usage of mobile data services is influenced by three major factors: user requirements, service discovery and usability. The results of the 3-way study suggest that all three factors did indeed have an influence.

First, in terms of user requirements, we noticed that the most popular and most frequent ideas for services are not currently available to consumers. More than 50% of the ideas produced were about context-aware services and only 10% of the ideas proposed services that are currently provided. None of the suggested services included the most widely advertised existing services, such as news, sports, ringtones and pictures. An interesting observation is that most of services proposed addressed everyday problems, such as efficient commuting, avoiding embarrassment in social situations and facilitating navigation.

Service discovery proved to be exceptionally difficult, even when this simply meant having to choose from a static set of 12 services. It is clear that this will be even more complicated if the public is introduced to contextually available services. The results of this study indicated that the struggle of choosing the appropriate service exceeded the struggle of completing the task. Participants visited 500% pages more than needed to find the optimum service when for the overall task the respective number was 213%. Similarly, they spent 1403% more time than was needed for service discovery and 499% more for task completion.

These results also indicate that participants faced major difficulties in interacting with the device. Although they were experienced mobile users, they found it particularly difficult to interact with the data services. This is also backed up by the frequency and sometimes intensity of user comments of dissatisfaction during the evaluation. One of the strongest findings was that most participants explicitly stated during the evaluation that they would not use it in real life. Unfortunately, the malfunction of the microphone stop us from having the exact number of these comments. Among other interesting comments were "I'd rather ask someone on the street" or "I'd wait until I got to my PC". There were even some who made very negative comments, such as "I hate it", "it's absolutely useless" or "can I swear on camera?".

There are also strong indications that the structure of most of the services does not match the users' mental models. The optimum path was followed in only 3.6% of the tasks and only by 2 participants. In most cases, task completion was hindered by deeply nested options with ambiguous labelling. For example, "Nearest Bookmakers" was under "More Sports", while "Essential Services" was one level down from "Tattoo & Piercing". In another example, labels such as "Pinpoint me" and "Find me now" strongly suggested that the service would display the user's location. However, upon selecting them, users were asked manually to input their location in order to see the relevant map. Even then, they had to go through 6 more steps (including purchase confirmation and a terms and conditions page) to get the map. The labelling problem was reported both from the expert evaluation and from user comments.

Furthermore, although this study did not aim to take into account the financial cost of accessing the services, interesting data resulted from the cooperative evaluation. Users intuitively and explicitly weighed their overall satisfaction about their experience against the money they would have to pay to access these data services. Although they were not personally being charged in the study, they were reluctant to accept charges during the interaction (e.g. 25p for a map) and expressed their concern with comments like "It's wasting my time and my money...", "will it charge me again?", even "... this would be costing me money, wouldn't it? So I'd be getting a bit pissed off at this stage".

One of the limitations of the study reported here is the small number of participants. The statistical power of the results is sometimes brought into question by a large standard deviation. Although this suggests relatively large differences in the participants' performance, it is also indicative of the differences in the strategies they

followed. Some users always preferred to try the “Search” option first, while others followed exploratory learning. We might assume that a population of more mixed background and experiences would apply an even greater variety of strategies. In this sense, the homogeneity of our sample becomes a strength of the study as it highlights the need for inclusive systems’ design.

Although concurrent verbal protocol is known to increase performance time, it does not account for an increase of as much as 500%. Besides, the number of extra pages that the users visited for each task is supporting the argument that they would also need much more time to complete the task.

Although all the equipment was thoroughly tested and functioned properly in the pilot study, an unpredictable fault made part of the audio recording irretrievable. This did not cause insuperable problems nor did it jeopardise the validity of the data reported here. We did lose some data on user comments which we are confident were of similar nature to the recorded data, since the evaluator was also taking notes.

Finally, the quality of service (QoS) could have further decreased user satisfaction. The error rate was one error every 13.5 minutes with 34 errors occurring in 7.8 hours of testing. On one occasion, one set of tasks had to be cancelled as the server did not respond 7 times in 14 minutes and the latency was very poor. Later we found out that this was due to exceptionally high number of requests to the server. QoS would have been better over a 3G network rather than the GPRS network that was used, but increased bandwidth does not always bring faster download times as service providers often take advantage of increased bandwidth to make the content richer.

## 6 Conclusion and Future Work

This study was motivated by the observation that mobile data services remain unused by the majority of mobile phone users. A combination of expert evaluation, diary study and cooperative evaluation was used to investigate the factors responsible for this lack of usage. Our prediction was that apart from the usual suspects of user requirements and usability, service discovery would also share this responsibility.

This prediction was indeed verified from the study results. Participants needed 1403% more time than optimum and 500% more pages to visit in order to initiate the service they considered as most appropriate for the task. These numbers are greater than the respective 500% extra time and 213% extra browsing that they needed to complete the tasks.

In our future work we aim to explore the concept of service awareness; an approach we consider will be particularly useful if services are to become contextually available. The mobile devices’ limitations on interaction and the dynamic availability of the services will hinder the process of service discovery. Context-aware systems could provide some assistance in raising service awareness but there are multiple interaction, social and technological issues that need to be considered.

## References

- Barkhuus, L. & Dey, A. [2003], Is Context-aware Computing Taking Control Away from the User? Three Levels of Interactivity Examined, in A. K. Dey, A. Schmidt & J. F. McCarthy (eds.), *UbiComp 2003: Ubiquitous Computing (Proceedings of the Fifth International Conference on Ubiquitous Computing)*, Vol. 2864 of *Lecture Notes in Computer Science*, Springer-Verlag, pp.159–66.
- Brewster, S. A. [2002], Overcoming the Lack of Screen Space on Mobile Computers, *Personal and Ubiquitous Computing* **6**(3), 188–205.
- Chen, G. & Kotz, D. [2000], A Survey of Context-aware Mobile Computing Research, Technical Report TR2000-381, Dartmouth College, USA.
- Cheverest, K., Mitchell, K. & Davies, N. [2002], Exploring Context-aware Information Push, *Personal and Ubiquitous Computing* **6**(4), 276–81.
- Clarkson, E., Clawson, J., Lyons, K. & Starner, T. [2005], An Empirical Study of Typing Rates on Mini-QWERTY Keyboards, in G. van der Veer & C. Gale (eds.), *CHI'05 Extended Abstracts of the Conference on Human Factors in Computing Systems*, ACM Press, pp.1288–91.
- Dey, A. K. [2001], Understanding and Using Context, *Personal and Ubiquitous Computing* **5**(1), 4–7.
- Garzonis, S. [2005], Usability Evaluation of Context-aware Mobile Systems: A Review, Paper presented at 3rd UK-UbiNet Workshop. [http://www.bath.ac.uk/compsci/hci/UK-Ubinet%20Files/Garzonis/ukubinet05\\_S.Garzonis.pdf](http://www.bath.ac.uk/compsci/hci/UK-Ubinet%20Files/Garzonis/ukubinet05_S.Garzonis.pdf), last accessed 2006-05-09.
- Gong, J.&Tarasewich, P. [2005], Alphabetically Constrained Keypad Designs for Text Entry on Mobile Devices, in G. van der Veer & C. Gale (eds.), *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI'05)*, ACM Press, pp.211–20.
- Kaikkonen, A., Kekalainen, A., Cankar, M., Kallio, T. & Kankainen, A. [2005], Usability Testing of Mobile Applications: A Comparison between Laboratory and Field Testing, *Journal of Usability Studies* **1**(1), 4–17.
- 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 & Information Technology* **24**(1), 51–65.
- Kjeldskov, J. & Graham, C. [2003], A Review of Mobile HCI Research Methods, in L. Chittaro (ed.), *Human-Computer Interaction with Mobile Devices and Services: Proceedings of the 5th International Symposium on Mobile Human-Computer Interaction (Mobile HCI 2003)*, Vol. 2795 of *Lecture Notes in Computer Science*, Springer-Verlag, pp.317–35.

- Kjeldskov, J. & Skov, M. B. [2003], Creating a Realistic Laboratory Setting: A Comparative Study of Three Think-Aloud Usability Evaluations of a Mobile System, in M. Rauterberg, M. Menozzi & J. Weeson (eds.), *Human-Computer Interaction — INTERACT '03: Proceedings of the Ninth IFIP Conference on Human-Computer Interaction*, IOS Press, pp.663–70.
- Kjeldskov, J. & Stage, J. [2004], New Techniques for Usability Evaluation of Mobile Systems, *International Journal of Human-Computer Studies* **60**(5-6), 599–620.
- Kjeldskov, J., Skov, M. B., Als, B. S. & Høegh, 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 S. Brewster & M. Dunlop (eds.), *Human-Computer Interaction — Mobile HCI 2004: Proceedings of the 5th International Symposium on Mobile Human-Computer Interaction*, Vol. 3160 of *Lecture Notes in Computer Science*, Springer-Verlag, pp.61–73.
- Kristensson, P.-O. & Zhai, S. [2005], Relaxing Stylus Typing Precision by Geometric Pattern Matching, in R. St. Amant, J. Riedl & A. Jameson (eds.), *Proceedings of the 10th ACM International Conference on Intelligent User Interface (IUI 2005)*, ACM Press, pp.151–8.
- Lemlouma, T. & Layaida, N. [2005], Content Interaction and Formatting for Mobile Devices, in A. Wiley & P. R. King (eds.), *Proceedings of the 2005 ACM Symposium on Document Engineering (DocEng'05)*, ACM Press, pp.98–100.
- MacKenzie, S. [2002], Mobile text entry using three keys, in O. W. Bertelsen, S. Bødker & K. Kuuti (eds.), *Proceedings of NordiCHI 2002*, ACM Press, pp.27–34.
- Microsoft [2002], Glossary of Networking Terms for Visio IT Professionals, <http://www.microsoft.com/technet/prodtechnol/visio/visio2002/plan/glossary.msp>, last accessed 2006-05-09.
- Miyahara, K., Inoue, H., Tsunesada, Y. & Sugimoto, M. [2005], Intuitive Manipulation Techniques for Projected Displays of Mobile Devices, in G. van der Veer & C. Gale (eds.), *CHI'05 Extended Abstracts of the Conference on Human Factors in Computing Systems*, ACM Press, pp.1657–60.
- Patel, S. N., Pierce, J. S. & Abowd, G. D. [2004], A Gesture-based Authentication Scheme for Untrusted Public Terminals, in S. K. Feiner & J. A. Landay (eds.), *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology (UIST'04)*, *CHI Letters* **6**(2), ACM Press.
- Pirhonen, A., Brewster, S. & Holguin, C. [2002], Gestural and Audio Metaphors as a Means of Control for Mobile Devices, in D. Wixon (ed.), *Proceedings of SIGCHI Conference on Human Factors in Computing Systems: Changing ourWorld, Changing Ourselves (CHI'02)*, *CHI Letters* **4**(1), ACM Press, pp.291–8.

Rieman, J. [1993], The Diary Study: A Workplace-oriented Research Tool to Guide Laboratory Efforts, in S. Ashlund, K. Mullet, A. Henderson, E. Hollnagel & T. White (eds.), *Proceedings of INTERCHI'93*, ACM Press/IOS Press, pp.321–6.

Silfverberg, M., MacKenzie, I. S. & Korhonen, P. [2000], Predicting Text Entry Speed on Mobile Phones, in T. Turner & G. Szwillus (eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'00)*, *CHI Letters* 2(1), ACM Press, pp.9–16.

Taylor, P. [2006], Mobile Consumer Applications Outlook 2006, Report 2751, Strategy Analytics.

<http://www.strategyanalytics.net/default.aspx?mod=ReportAbstractViewer&a0=2751>, registration required.

Vetere, F., Howard, S., Pedell, S. & Balbo, S. [2003], Walking through Mobile Use: Novel Heuristics and their Application, in S. Viller & P. Wyeth (eds.), *Proceedings of Australian Conference on Computer–Human Interaction (OzCHI 2003)*, Information Environments Program, University of Queensland, pp.24–32.

Wigdor, D. & Balakrishnan, R. [2003], TiltText: Using Tilt for Text Input to Mobile Phones, in *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology, UIST'03*, *CHI Letters* 5(2), ACM Press, pp.81–90.

Winograd, T. & Flores, F. [1986], *Understanding Computers and Cognition: A New Foundation for Design*, Ablex. From 1988, an Addison–Wesley publication.

Wright, P. C. & Monk, A. [1989], Evaluation for Design, in A. Sutcliffe & L. Macaulay (eds.), *People and Computers V (Proceedings of HCI'89)*, Cambridge University Press, pp.345–58.

Wright, P. C. & Monk, A. F. [1991], A Cost-effective Evaluation Method for Use by Designers, *International Journal of Man–Machine Studies* 35(6), 891–912.