Pers Ubiquit Comput (2006)
DOI 10.1007/s00779-006-0064-9

ORIGINAL ARTICLE

Eleanor Toye · Richard Sharp · Anil Madhavapeddy
David Scott · Eben Upton · Alan Blackwell

Interacting with mobile services: an evaluation of camera-phones and visual tags

Received: 3 June 2005 / Accepted: 20 September 2005
© Springer-Verlag London Limited 2006

Abstract We present a study of using camera-phones and visual-tags to access mobile services. Firstly, a *user-experience study* is described in which participants were both observed learning to interact with a prototype mobile service and interviewed about their experiences. Secondly, a *pointing-device task* is presented in which quantitative data was gathered regarding the speed and accuracy with which participants aimed and clicked on visual-tags using camera-phones. We found that participants' attitudes to visual-tag-based applications were broadly positive, although they had several important reservations about camera-phone technology more generally. Data from our pointing-device task demonstrated that novice users were able to aim and click on visual-tags quickly (well under 3 s per pointing-device trial on average) and accurately (almost all meeting our defined speed/accuracy tradeoff of 6% error-rate). Based on our findings, design lessons for camera-phone and visual-tag applications are presented.

1 Introduction

Using camera-phones to recognise visual tags is currently a hot topic, receiving significant media coverage throughout 2004 [4, 21, 25]. However, despite this press coverage and recent commercial tag-reader implementations [3, 18], there has so far been very little research into the *usability* of applications that exploit camera-phones and tags. Therefore it is unclear whether ordinary users are able to interact easily with camera-phone/tag-based applications and, further, what practical value (if any) ordinary users see in these types of application.

The study presented in this paper addresses these issues. More specifically we attempt to answer four key usability questions: (1) how quickly and accurately can novice users click on visual tags?; (2) is it clear to novice users how to use camera-phones and tags in the context of a realistic mobile service?; (3) what do users see as the advantages and disadvantages of services that employ camera-phones and tags? and (4) what are users' attitudes towards adopting this technology in their daily lives?

Our study is split into two parts. Firstly, we present a user-experience study in which we both observe participants learning to interact with a prototype mobile service and interview them about their experiences. This part of the study primarily addresses questions (2), (3) and (4) above. Secondly, we report quantitative results arising from our *pointing-device task*—an experiment designed to assess how quickly and accurately users can click on visual tags—providing an answer to question (1) above.

The remainder of the paper is structured as follows. We start by providing background information on camera-phone/visual-tag interaction generally (Sect. 2). Following this, we describe our implementation of the particular camera-phone/visual-tag interaction system that we employ in our user studies (Sect. 3). The user-experience study (Sect. 4) and the pointing-device task (Sect. 5) are then reported. Finally, we consider related work (Sect. 6) and present conclusions and design-lessons learnt from our experiments (Sect. 7).

2 Background

The idea of using tags to "bridge physical and virtual worlds" pre-dates camera-phones [28]. Several ubiquitous

E. Toye (✉) · R. Sharp
Intel Research Cambridge, Cambridge, UK
E-mail: eleanor.toye@gmail.com
E-mail: richard.sharp@intel.com

A. Madhavapeddy · D. Scott · E. Upton · A. Blackwell
Computer Laboratory, Cambridge University, 15 JJ Thomson Avenue, Cambridge CB3 0FD, UK
E-mail: anil@recoil.org
E-mail: dave@recoil.org
E-mail: ecu20@cam.ac.uk
E-mail: afb21@cam.ac.uk

computing researchers have explored this idea, proposing a number of motivating application scenarios. For example, as part of the CoolTown project [14], Kindberg et al. prototyped a museum in which users can scan exhibits with their PDAs in order to read more about them and bookmark this information for future reference. More recently, Microsoft's Aura project [26] successfully deployed a similar idea in the real world, using PDAs with barcode-readers to scan UPC/EAN barcodes on existing products. Other visual tag-based applications that have been proposed include an augmented calendar [22], and a scheme for assisting system administrators manage displayless rack-mounted machines [25].

The recent camera-phone boom has created a resurgence of interest in tag-based interaction techniques since, for the first time, large numbers of real-world users own a device capable of supporting the above applications and usage scenarios. The phone's embedded camera can photograph visual tags, software running on the phone can decode tags and wireless networking technologies, such as GPRS, 3G or Bluetooth, can be used to fetch data on the basis of tag sightings.

A number of commercially available products have already appeared which allow users to scan visual tags with their camera-phones and download mobile content [3, 17]. Applications that have been built using this technology include scanning a tag to retrieve bus timetables [18] and scanning a tag on a business card to record someone's contact details [21].

An important consideration is that the different applications and usage scenarios proposed for visual tags impose very different requirements on tag-reading technology. For example, in the bus timetable application [18], users typically click on a single tag in order obtain the waiting time for the next bus. In this case, the speed at which the tag must be decoded by the tag-reader is not technically challenging: after scanning the tag, users may be prepared to wait several seconds before timetable information appears on their phone's screen. In other applications users may require access to information much more quickly. For example, in the CoolTown museum [14], users may want to scan several exhibits in quick succession. Similarly, in the case of the system administration tool [25], users may quickly need to scan the tags of several rack-mounted machines in order to check their status. The speed at which users can scan tags with camera-phones thus determines the class of applications that the interaction technique can feasibly support.

3 Implementation

We implemented a software framework for camera-phone/tag-based interaction by extending our previous work on the *Mobile Service Toolkit*: a client/server architecture supporting situated services that interact with mobile phones [27]. Our implementation of the *Mobile Service Explorer*—the phone-based client which enables users to access situated services—runs on all Nokia Series-60 camera-phones.

On activating the Mobile Service Explorer the camera-phone's display immediately becomes a *viewfinder*: a live video feed continually capturing frames from the phone's embedded camera. Whenever a tag becomes visible in the viewfinder, it is highlighted with a red cross-hair and tracked in real-time (see Fig. 1). If multiple tags are visible on the screen at the same time, the tag nearest the centre of the viewfinder is highlighted. We adopt standard GUI terminology, saying that the highlighted tag has the focus. Users can click on a visual tag by pressing the select button on the phone's keypad when the tag has the focus.

A user interacts with a mobile service primarily by aiming and clicking on tags using their camera-phone. When a tag has the focus, a mobile service may display a small amount of information on the phone's screen in a manner analogous to augmented-reality overlays. Clicking on a tag may result in an action being performed by the service or a simple UI control appearing on the camera-phone's display. We currently allow mobile services to export numerical data-entry fields, list selection fields and multiple choice dialogues (e.g. "Continue? Yes/No") to a user's camera-phone. For example, clicking on a tag to join a queue in our prototype virtual queueing application (see Sect. 4.1) results in a numerical data-entry field appearing on the camera-phone's screen prompting the user to enter the number of people who want to join the queue.

When we implemented the Mobile Service Toolkit there were, to our knowledge, only three camera-phone-based tag reading packages generally available: SemaCode [3], SpotCode [18] and Rohs' tag recognizer [23]. We evaluated each of these platforms in terms of decoding latency, robustness to variable lighting conditions and robustness to variability in the phone's orientation with respect to visual tags. In each of these areas, which we hypothesised were important usability factors, the SpotCode Reader performed the best. For this reason, we chose to incorporate SpotCodes in our implementation.

Initially based on the TRIP tag format [7], SpotCodes are distinctive, circular tags with 2 data rings and 21 sectors (see Fig. 3). Each SpotCode encodes 42 bits of information, a data capacity more than sufficient for our purposes.

Since this paper focuses on evaluation rather than implementation, we do not discuss the technical details of the Spotcode format, nor the software architecture of the Mobile Service Toolkit here. Readers interested in these topics are referred to other publications [24, 27].

4 User experience study

Our user experience study was based around a Theme Park *virtual queueing system* in which participants used

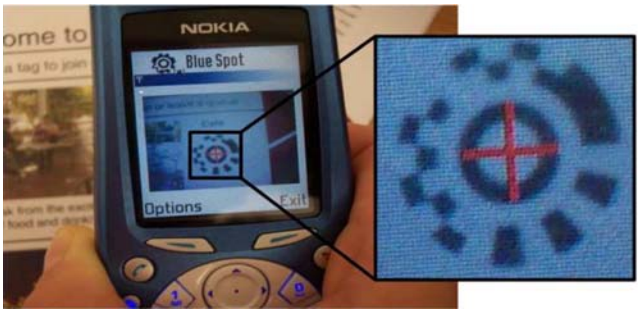

Fig. 1 Using the phone-based tag reader

camera-phones and visual tags both to obtain information about theme park attractions and to join and leave virtual queues. We chose this application as a technology-probe for two reasons:

1. We thought that the theme park scenario was likely to be familiar to our participants; indeed, when we interviewed our participants, we found that they had all visited theme parks. Therefore, despite the limitations imposed by presenting the application in a lab setting, we were able to gather useful data regarding participants' impressions, beliefs and opinions relevant to plausible uses of camera-phones and visual tags.
2. The virtual queueing system required participants to make heavy use of visual tags throughout the entire interaction period. Users had to aim their camera-phone at several tags in quick succession in order to compare the current queueing information for the various theme park attractions.

The approach of prototyping an application for a familiar environment in a lab-setting was inspired by Kindberg et al.'s UbiComp '04 paper [15]. Note that, since our virtual queueing system is purely a technology probe designed to elicit feedback from participants, we deliberately ignore the technical details involved in extending the prototype system to a real deployment.

We continue by describing the theme park virtual queueing application in more detail (Sect. 4.1). We then describe the participants who took part in the user-experience study (Sect. 4.2), the experimental procedure (Sect. 4.3) and, finally, our findings arising from the study (Sect. 4.4).

4.1 Application description: theme park virtual queueing system

Theme parks often have long queues for rides. Some theme parks have addressed this problem already by devising *virtual queueing systems* in which, rather than waiting in line, people are allocated ‘‘tickets’’ allowing them to get on rides at specified times. Tickets may be pieces of paper or may be virtual capabilities stored on an electronic device. At present, electronic virtual queueing systems deployed in theme parks involve specialised mobile devices or smartcards lent to visitors for the duration of their time in the park. In this section we describe our prototype virtual queueing system that exploits visual tags and users' own camera-phones.

We set up ‘‘TagLand Theme Park’’ in a large student computer laboratory in our building. TagLand consists of six individual posters sited in specific locations across the lab, each representing a different attraction. A large central poster contains a map showing the locations of the six attractions. A *visitor information leaflet* (handed to participants at the start of each study) contains a list of all six attractions, providing some information about what each one entails.

Each attraction has an associated visual tag, which is printed on the poster where the attraction is located, at the relevant position on the central map, and on the visitor information leaflet. For each of these six attractions, holding the camera-phone over the associated tag causes a text overlay to appear at the top of the camera-phone's viewfinder displaying relevant information (see Fig. 2). One of the rides is closed for maintenance and holding the camera-phone over its associated tag reveals this information; the other attractions display their current queueing times.

The posters representing attractions and the central map contain text inviting users to ‘‘click on a tag to join a queue’’, and explain that they can leave the queue at any time simply by ‘‘clicking the tag again’’. On clicking a tag, a text-entry field appears on the phone prompting the user to ‘‘enter the party size’’ (the number of people wishing to join the queue). After the party size is entered, using the keypad on the phone, the application records that the participant is now in the queue for a ride.

When a user focuses on a tag corresponding to an attraction for which they have a booking, the resulting text overlay informs them that they are in a virtual queue. The remaining queueing time in minutes and seconds is also displayed, ticking down to zero. Clicking on the tag in this state yields a confirmation dialogue asking the user if they want to leave the queue. Users can select ‘‘Yes’’ or ‘‘No’’ using their camera-phone's keypad.

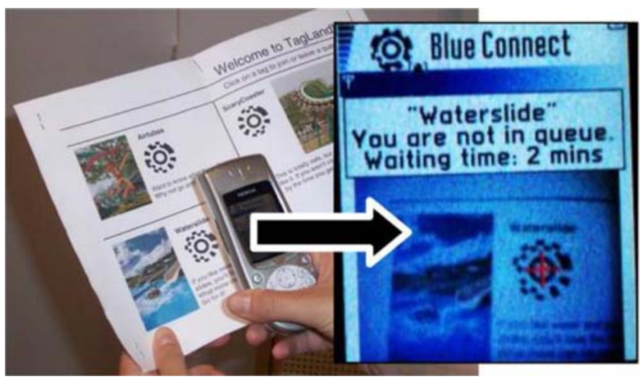

Fig. 2 Using the camera-phone in the Theme Park Application

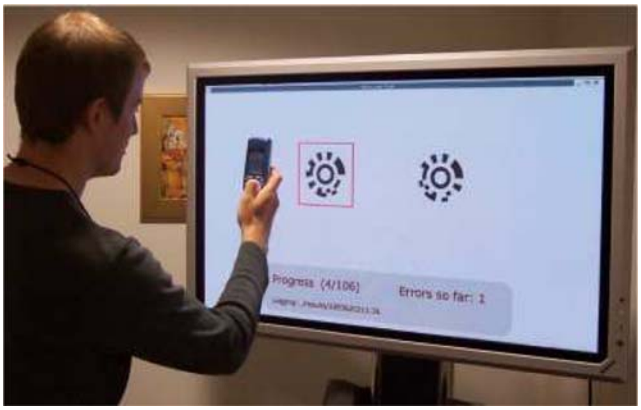

Fig. 3 Performing the pointing device task

Shortly before the user reaches the front of a queue, the phone beeps and displays a dialogue notifying them to "Go to ⟨*attraction name*⟩ in 30s". (Clearly in a real deployment more notice would be given.) When the front of a queue is reached a fanfare sound plays and an alert dialogue appears, confirming that they are now able to take the ride.

4.2 Participants

Eighteen people (nine men and nine women) aged between 24 and 43 years (mean = 29) served as participants. They were an educated and mostly professional group. None reported difficulties with vision or motor skills. None worked in the field of Computer Science or in the mobile phone industry. All were mobile phone users and three owned camera-phones. Participants were recruited either through flyers posted around the city and University of Cambridge, or by word of mouth and were each rewarded with a book token (value 20 UKP).

Our rationale for choosing this demographic was that it allowed us to identify any general usability difficulties which affected even healthy, younger adults. We intend to carry out further studies with broader samples of the population in future work.

4.3 Procedure

Each participant was given a Nokia 3650 camera-phone and a visitor information leaflet. They were instructed to explore the theme park and use the camera-phone to join and leave queues for the different rides and attractions. They were told that they could do this for as long as they liked and to meet the researcher back at the central map whenever they felt they had finished exploring.

Structured interviews were carried out at the end of each session. Initial questions in the interview schedule were entirely open-ended (e.g. "Was there anything about the Theme Park that you found confusing or difficult?"). Next, users were asked to make comparisons between using camera-phones/tags and other familiar technologies, in the context of virtual queueing systems. All the comparative questions were worded to avoid drawing attention to specific features of the technology (e.g. speed or convenience). Finally, more general questions were asked in order to explore participants' overall impressions of the technology and also their attitudes to adoption.

We wanted to know whether novice users would understand how to use the camera-phone and tags in the context of the theme park application. For this reason we deliberately kept the instructions given to participants to an absolute minimum. The purpose of the instructions was primarily to direct the participants' attention to the features of the application, without giving precise information about what to do.

4.4 Results and discussion

Our key result is that, within less than 15 min, each participant had succeeded in understanding and using every aspect of the queueing system without researcher intervention. This gives an unequivocally positive result for question (2) of Sect. 1: it is clear to novice users how to use the phone display in conjunction with visual tags in the context of a realistic mobile service.

Although all participants were able to interact successfully with the theme park application, two minor usability issues did arise. Two (of 18) participants commented that they had some trouble with the red cross-hair flickering on the viewfinder, so sometimes they were unsure whether they had successfully focused on a tag. Observation suggested that this was because they were holding the camera-phone just beyond the maximum distance at which the tag could be reliably decoded. We expect that this problem would disappear with further practice.

One participant also commented that "tags below waist height were difficult to focus on", but added that this could easily be remedied by: "putting low tags at an angle or moving them higher." No other ergonomic issues were reported.

The remainder of this section addresses our third and fourth usability questions, reporting users' perceptions of advantages and disadvantages of camera-phone/tag-based services, and their attitudes surrounding the widespread adoption of this technology.

4.4.1 Perceived advantages and disadvantages

Many participants saw advantages to using the camera-phone based queueing system. Eight made positive comments about having mobile access to current queue lengths and the ability to change their own queue status using tags. For instance:

> I liked the ability to get real-time updates by just looking at the tags to find out how far I was from the front of the queue.

> I liked the fact that it works exactly the same from various locations and from the map, leaflet and posters dotted around.

A particularly popular feature of the system was that tags were contained in the leaflet:

> Very portable. You've got the two things you need to make it all work (phone and leaflet) and you can do instant checks.

However, a number of participants were concerned about the use of camera-phones for this type of application. Concerns included camera-phones not working abroad and being stolen:

> it'd be a great place for mobile phone thieves if everyone was waving around fancy camera-phones.

One participant was also worried that a camera-phone might get broken on Theme Park rides:

> Phones and water-rides don't mix easily!

Another commented that they were not allowed to use camera-phones at work:

> I can't get a camera-phone because they aren't allowed in schools, where I often work...

Recall that we asked participants to make comparisons between using camera-phones/tags and other familiar technologies in the context of virtual queueing systems. Firstly, participants were asked to compare the Theme Park application with "entering their mobile phone number into a touchscreen" and then "receiving text messages telling them when to go to the ride". Thirteen participants felt that the touchscreen/text solution would be worse than our camera-phone/tag application (two believed it would be the same, one believed it would be better and one provided uninterpretable data for this question). Six of the participants who thought it would be worse said this because they felt text messages were much less convenient to access than our tag-initiated overlays (see Fig. 2). We were surprised at how negative participants were about the hypothetical text-message-based interaction technique. A characteristic comment from one participant was:

> With the texts, you have to wait for them to arrive, and sometimes they're late—could be minutes or days late; and they fill up your inbox, so you'd have to keep checking your messages and deleting. You have to hit more buttons to get your message than with the camera and tags system.

Secondly, participants were asked to compare the Theme Park application to a conventional ticket-based virtual queueing system. In response to this question, 16 participants said they would prefer the camera-phone system. The most common reason cited for this was that, in a ticket-based system, one has to carry round several tickets and might lose some of them—for example:

> You'd need a lot of tickets, you'd have too many things to carry and you might lose the tickets.

One participant, who had recently experienced a ticket-based virtual queueing system at Disneyland last year, said of his Disney queueing experience:

> ...[Disney's queueing system] is better than standing in line but I can't cancel my place on the ride and there'd be queues of people at the ticket machines—and I have to go to the machines, I can't just join queues from wherever I am.

To sum up, participants did see some unique benefits of using camera-phones to access mobile services. Although they also had some reservations, these focused almost entirely on the camera-phone technology itself (e.g. fragile, expensive, stealable) rather than the specifics of the interaction technique.

4.4.2 Attitudes to adoption

At the end of the interview schedule, we asked participants to make general comments about camera-phone/tag-based applications. Although most of the comments were positive, a few participants took this opportunity to express concerns regarding the widespread adoption of this technology. Two participants were worried about accessibility:

> How well would ... people with disabilities cope with this technology?
>
> The camera-tag system seems to require quite good hand-eye coordination, so it's not good for the elderly or anyone who is a bit shaky.

We recognise that further studies, with a broader sample of participants, are required to address accessibility issues.

One participant was worried that deployment of tag-based services may benefit camera-phone users to the detriment of everyone else:

> It could create a technology hierarchy where camera-phone owners get better services—I vaguely resent this idea. That's why I preferred the idea in the theme park of being lent a device: it seems fairer.

However, on the whole, attitudes to adoption were very positive. All of our participants (not including two who already had Bluetooth camera-phones) said that, if tag-based mobile services were widely available, they would choose to buy a phone that allowed them to use tags in preference to one that did not. Nearly half (7 out of 16 participants who did not already own camera-phones) said that they would be prepared to pay more for a camera-phone if tag-based mobile services were widely available.

Of the group tested, all already used mobile phones on a regular basis. It remains an open question whether other social groups, who may not already use mobile phones regularly, would be so open to adopting camera-phones and tags.

5 The pointing device task

The pointing device task is a controlled experiment that addresses the first of the usability questions presented in Sect. 1: how quickly and accurately can novice users click on visual tags?

Our key concern was to determine whether participants' clicking performance was adequate for even heavily tag-centric usage scenarios to be viable (e.g. the CoolTown Museum [14] and system administration application [25] that we discussed in Sect. 2). Given the nature of such applications, we chose to define acceptable performance as a clicking time of less than 3 s on average, with a clicking error rate of less than 10%.

Participants The participants for the pointing device task were the same as those who took part in the user experience study (see Sect. 4.2).

5.1 Experimental design

Our design of the pointing device task was heavily influenced by previous work in applying Fitts' law models to mice, tracker balls and graphics tablets [19, 20].

In each *trial*, a pair of equal-sized tags appears on a plasma screen, one highlighted with a red box around it (see Fig. 3). A participant uses their camera-phone to click first on the highlighted tag, then on the other tag.

The horizontal distance between tags and the diameter of the tags varies across different trials. Every combination of eight different tag diameters (ranging from 4 to 11 cm, at increments of 0.875 cm) and six different inter-tag distances (ranging from 18 to 63 cm, at increments of 7.5 cm) are presented, resulting in $6\times8=48$ trials.[1] Each of these 48 trials is performed twice: once for each *clicking order*—that is, whether at the start of a trial the highlighted tag is on the right or the left. Thus, the total trials performed in a single test run = $48\times2=96$. Trials with the same clicking order are presented consecutively, meaning that the clicking order only changes once between the two blocks of 48 trials. The order of the trials within these two blocks of 48 are randomly permuted for each participant.

The 96 trials are divided into sets of 15, each separated by 15-s rest periods, with a short set of six trials at the end. The choice of 15 trials per rest period was not determined by the experimental design; we simply found, during pilot testing, that this was the number of trials a participant could perform consecutively before they started to feel tired.

5.2 Procedure

Participants were asked to perform ten initial practice trials, followed by the 96 clicking trials. For each trial, participants were told to click first on the highlighted tag and then, as quickly as possible, on the other tag. To ensure a standard speed/accuracy tradeoff across participants we instructed them to aim for no more than one error in each set of 15 trials (corresponding to a maximum of six errors per session). To avoid ordering effects we ensured that the same number of participants started with the left/right clicking order as started with the right/left clicking order.

For each trial, a computer recorded the *clicking time*:[2] the time between the participant clicking on the first tag and clicking on the second tag rounded to the nearest 10 ms. (This resolution was sufficient for our purposes.) An error was registered if the participant pressed the select button on their camera-phone when a tag was not in focus, or if the wrong tag was clicked on. At the bottom of the screen a status bar showed the participant both the number of trials remaining and the number of trials in the current set of 15 that had so far resulted in error.

Participants were informed that there would be a forced break every 15 trials. They were also told that they could take breaks as often as they liked and for as long as they liked between trials (although in practice no-one actually did this). Every participant performed the pointing device task twice, separated by a break of at least half an hour. This gave us 192 trials per participant (and therefore a total of 3,456 trials across our sample of 18 participants). We have made this data available for download so that other researchers can verify our results and perform their own analyses [2].

5.3 Results and discussion

Our main result is that all participants' performance fell within the pre-defined acceptable limits for clicking

[1] We chose to study these ranges of diameters and distances because we felt they were typical of those that may be encountered in tag-based interactions with mobile services (e.g. consider the plausible tag sizes and inter-tag distances that may be designed into posters)

[2] Although recorded by computer, the times were measured by the camera-phone itself. This ensured that the latency and jitter of the Bluetooth connection (used by the phone to communicate with the computer) did not affect our data

speed and accuracy (see Sect. 5). Participants' mean clicking times ranged from 0.89 s to 2.63 s; number of errors per session ranged from 0 to 8 (out of a total of 96 trials per session). Table 1 summarises our data, showing means and standard deviations of clicking time and error rate (collapsed over all participants and broken down by session). Summary data is also shown graphically in Fig. 4.

This shows conclusively that even novice users can click on visual tags quickly and accurately using commercially available camera-phones and tag-reading software. It demonstrates that even heavily tag-centric interaction techniques are feasible without requiring manufacturers to add extra tag-reading hardware to their phones.

Although all participants met our acceptability criterion for clicking time in each session, participants clicked more quickly on average in the second session, indicating a significant[3] practice effect. We would expect to see continued improvement in clicking times as users become more familiar with camera-phones and tags. However, we re-iterate that even the first-time users' performance was, in all cases, already adequate for usability. (Note that, since our procedure specified a target error rate, it would be inappropriate to look for a practice effect on accuracy.)

Recall that we chose to study tag diameters between 4 and 11 cm because we felt that this was a plausible range of tag sizes for display on posters. We noticed a slight, but significant,[4] decrease in error rate as the tag-size increased. This suggests that, within this range of tag sizes, designers should aim to use larger tags where possible.

Figure 5 shows that all the tag sizes in the range we sampled are acceptable in terms of clicking time; for all tag diameters, mean clicking times were less than 3 s. There was no significant linear relationship[5] between clicking time and tag diameter.

Although the details are beyond the scope of this paper, we found that Fitts' law [20] correlations on the data were weak (taking tag diameter as a measure of Fitts' target width). This is perhaps unsurprising when one considers that a camera can be aimed by rotation about its vertical axis; Fitts' law models typically do not account for this method of target acquisition.

When generalising these results to other camera-phones running the SpotCode reader there are two factors to consider: the optical properties of the camera lens and the shape/ergonomics of the camera-phone. Fortunately, the optical properties of different camera-phone lenses are currently very similar so we expect our results to generalise well in this respect. For example, all Nokia camera-phones we studied had the same level of zoom (3.5 mm); other camera-phones we looked at had fixed zoom of between 2.8 and 4.5 mm.[6] To quantify how performance might vary with radically different levels of zoom, more studies would be required.

With regard to camera-phone shape and ergonomics, we note that most camera-phones have the same basic form factor as the Nokia 3650: a display and keypad on the front and a camera lens on the back, towards the top. We expect that our results would generalise well to other camera-phones with this basic form factor. Again, for radically different camera-phone designs further studies would be required.

Other tag formats/readers have very different usability properties to SpotCodes. For example, the SemaCode reader requires users to align rectangular tags within a target area appearing on the phone's viewfinder. Furthermore, there is no cross-hair to give real-time feedback while focusing on tags and, on selecting a tag, users incur a decode latency of about 3 s. We do not claim that our results generalise to SemaCode or, indeed, to any other tag reader or tag format. However, our experimental design provides a robust method for measuring the usability of tag formats and tag-reading software generally. Our work therefore forms a solid basis for future cross-platform evaluation.

6 Related work

We have already covered much of the related work in the field of visual-tag-based interaction techniques (see Sect. 2). Here we briefly outline work in other areas that is relevant to this paper.

There are many related interaction techniques which also use cameras as an input medium. Most of these systems involve fixed cameras trained on users [9]. This approach, pioneered by VideoPlace [16], has proved fruitful in areas such as gesture recognition [8] and location-aware computing [7]. In contrast, camera-phones are more suited to the approach often taken by the augmented reality community in which hand-held cameras are pointed at objects in the world. A variety of augmented reality systems based on optical tags have already been developed [5, 13]. These systems influenced our design of the theme park application to a large extent. In particular, the textual overlays that appear when the user focuses on a tag were inspired by the NaviCam [22] project.

[3] A one-way analysis of variance (ANOVA) established that the practice effect on clicking time was statistically significant (for $\alpha = 0.05$): $F(1,17) = 6.055$, $p = 0.025$

[4] Collapsing across both sessions, linear correlation between error rate and tag diameter yielded $r = -0.39$, with 99% confidence intervals for ρ of $(-0.554, -0.189)$. Note that this effect was also observed for session 1 and session 2 independently. For session 1, $r = -0.30$ with 99% confidence intervals for ρ of $(-0.48, -0.09)$; for session 2, $r = -0.37$ with 99% confidence intervals for ρ of $(-0.53, -0.17)$.

[5] Linear correlation between clicking time and tag diameter yielded $r = 0.05$, with 95% confidence intervals for ρ of $(-0.113, 0.212)$

[6] As a rule of thumb, a number smaller than 3.5 mm requires a user to hold their camera-phone closer to tags than our participants did; conversely, a number greater than 3.5 mm requires users to hold their phones further away from tags

Table 1 The tables show the mean (SD) clicking time per participant and the mean (SD) total number of errors per participant broken down for session 1 and session 2

	Mean (SD) clicking time (s)
Session 1	1.60 (0.65)
Session 2	1.51 (0.64)

	Mean (SD) total no. errors
Session 1	3.22 (2.18)
Session 2	3.33 (2.52)

One of the aims of using visual tags with camera-phones is to alleviate the constraints associated with small mobile displays and keypads. Many researchers have addressed this fundamental problem in different ways including panning [12], fish-eye views [10] and novel uses of audio [6]. We see visual tags as complementary to these approaches. Indeed, applications based on camera-phones and tags could also usefully employ (say) panning and zooming screen displays.

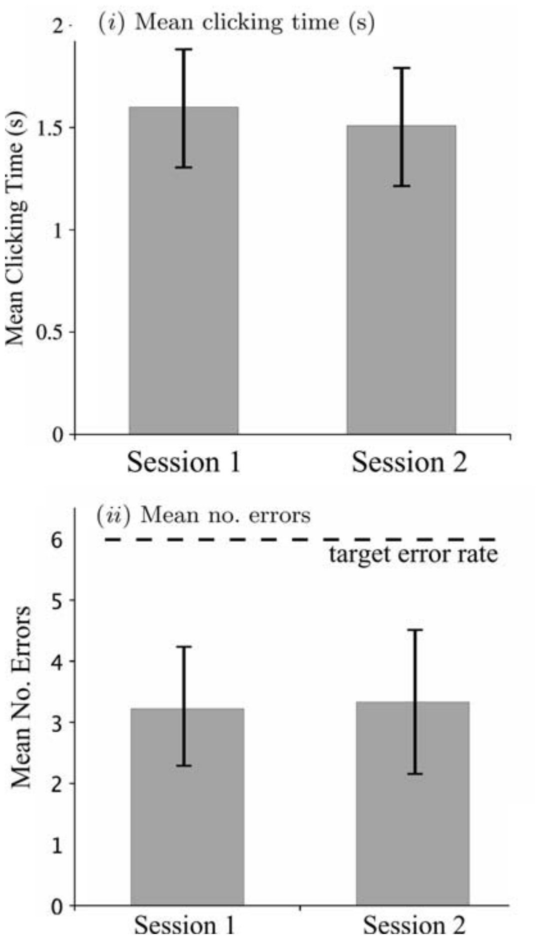

Fig. 4 The graphs show the mean clicking time and the mean number of errors per participant, for session 1 and session 2. The *error bars* show standard deviations

7 Conclusions, design lessons and future work

We have evaluated a commonly proposed interaction technique in which users point and click on visual tags using camera-phones. Our study focused specifically on four usability questions: (1) how quickly and accurately can novice users click on visual tags?; (2) is it clear to novice users how to use camera-phones and tags in the context of a realistic mobile service?; (3) what do users see as the advantages and disadvantages of services that employ camera-phones and tags? and (4) what are users' attitudes towards adopting this technology in their daily lives?

The first of these questions was addressed by the pointing device task (Sect. 5) which provided quantitative data demonstrating conclusively that novice users could click on visual tags quickly (well under 3 s per trial on average) and accurately (inside our specified acceptable error rate of 10%).

Although the pointing device task was performed using Nokia 3650 camera-phones, we expect that the results would generalise well to other currently available camera-phones running the SpotCode reader. Furthermore, the experimental design of our pointing device task provides a robust method for measuring the usability of tag formats and tag-reading software more generally. Our method thus forms a solid basis for future cross-platform evaluation.

Questions (2), (3) and (4) above were addressed by our user experience study, in which we observed our participants learning to interact with a prototype virtual queueing system and interviewed them about their experiences (Sect. 4). Our key result was that, in less than 15 min, each participant succeeded in understanding and using every aspect of the queueing system without researcher intervention. This gives an unequivocally positive result for question (2): it is clear to novice

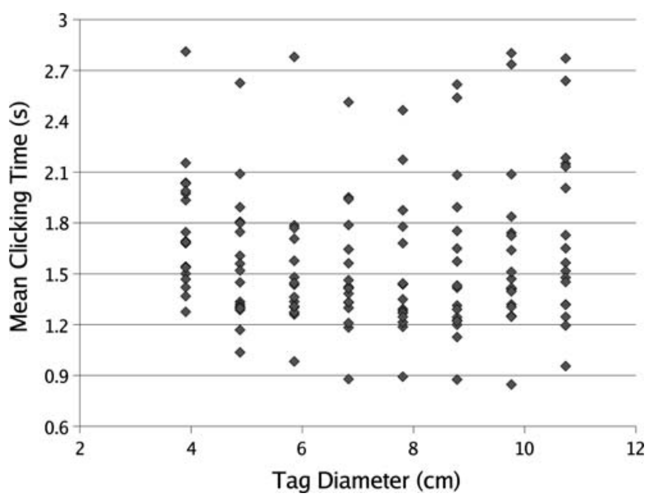

Fig. 5 Mean clicking times per participant by tag diameter (collapsed across both sessions)

users how to use camera-phones and tags in the context of a realistic mobile service.

With reference to question (3), opinions on the usefulness of the interaction technique were broadly positive, but with some important reservations. Mobile access to up-to-date information was generally considered a very attractive feature, with several participants also commenting that the visual tags themselves were "easy to use". Some participants were concerned about whether camera-phones were an appropriate device for accessing mobile services because of their currently limited user-base, their fragility and their potential for being easily lost or stolen. However, despite our participants' concerns, camera-phones continue to grow in popularity [11]. Tag-based services promise to add value for this growing number of camera-phone owners without imposing additional costs beyond those already incurred by camera-phone ownership.

In answer to question (4) we found that our participants' attitudes to adoption were generally positive. A few participants highlighted concerns about the social implications of providing public services which can only be accessed via individuals' camera-phones. However, all participants (who did not already own a camera-phone) considered that if this interaction technique became widely available, it would positively influence their attitude towards buying a camera-phone.

7.1 Design lessons

Based on our implementation experience and the findings from our user studies, we advise developers of camera-phones/visual tags to adhere to the following guidelines:

1. Keep data-entry on the phone's keypad to a minimum, relying on tag-based interaction instead whenever possible. (In the user-experience interviews, several participants reported that they preferred tag-based interaction to entering data via the camera-phone's keypad.)
2. Place text or icons that describe the function of a tag very close to the tag itself. Then, when a camera-phone is aimed at the tag, both the tag and its surrounding context will be visible in the camera-phone's viewfinder. This prevents users from having to constantly switch their gaze between two displays; all the information they need to identify a particular tag is on the phone's own display.
3. When small amounts of dynamic information are associated with visual tags, use overlays to display this information on the phone's screen (see Fig. 2). This allows the information to be seen as soon as the camera-phone is aimed at a tag, without even requiring users to click.

Beyond this specific application-design advice, our findings led to design lessons for camera-phone/tag applications more generally. Recall that, in Sect. 2, we observed that different tag-based applications impose very different requirements on tag-reading technology. For some applications (e.g. bus timetables) clicking speed may not be important. However, for other usage scenarios (e.g. the CoolTown Museum) clicking speed is more critical. The ease with which novice users were able to point and click on visual tags in the pointing device task and the success of the theme park application demonstrates that today's camera-phone technology is sufficient to support even heavily tag-centric interaction techniques. Given the findings presented in this paper we believe that all of the application scenarios proposed in Sect. 2 could feasibly be implemented on existing commodity camera-phones.

One caveat to this is that the wireless networking capabilities currently provided by mobile phone operators are often quite limited. Thus, even though our findings show that people can click on tags quickly using camera-phones, the latency incurred due to real-world networking technologies, such as GPRS for example, may prove unacceptable for some of the application scenarios proposed in Sect. 2. We expect that, over time, technical developments in mobile phone networks (e.g. 3G and 4G) will alleviate this problem. In the meantime we have demonstrated that Bluetooth can be used for low-latency short-range communication between camera-phones and tag-based situated services [27].

Despite growing adoption there are still some environments that, for reasons of security, prohibit the use of camera-phones (e.g. schools, gyms, industrial research labs). This was pointed out by one of our participants who often worked in schools. For environments in which cameras are unacceptable, other technologies such as RFID readers or even dedicated barcode scanners may provide a more suitable platform for tag-based interaction.

Our findings demonstrate that using camera-phones and visual tags to interact with mobile services offers a viable alternative to deploying shared public technology (e.g. the hypothetical theme park ticket machine that participants compared our camera-phone-based application to). By exploiting users' own camera-phones, the need to deploy shared public display/data-entry technology (and hence, the associated problems of vandalism and maintenance) can be avoided.

7.2 Future work

The work presented in this paper should be considered as a first exploratory step. When interpreting our results, one must bear in mind that our evaluation was based on prototype technology presented in a laboratory setting. However, given the constraints of a laboratory environment, we believe that our hybrid approach, combining quantitative measures along with observational and interview data, was successful in answering the usability questions we posed.

We see our contribution as the beginning of a longer process of iterative user-centred design of camera-phone/tag-based services. In the future we would like to explore further the implications of using different tag formats (e.g. QR codes [1]) and evaluate camera-phones and tags with a broader sample of the population. We also aim to use the findings presented in this paper to design, deploy and evaluate applications that exploit camera-phones and visual tags on a city-wide scale.

Acknowledgements The authors would like to thank Rob Ennals for his valuable comments and suggestions.

References

1. NTT DoCoMo Unveils New mova 506i i-mode Phone Series. Press release. Available from NTT DoCoMo press center
2. Raw data from pointing-device task. Available from http://chi.recoil.org/
3. Semacode website. http://semacode.org/
4. New York Times (7th October 2004). Connecting paper and online worlds by cellphone camera
5. Azuma RT (1997) A survey of augmented reality. Presence 6(4):355–385
6. Brewster SA (2002) Overcoming the lack of screen space on mobile computers. Pers Ubiquitous Comput 6(3):188–205
7. de Ipiña DL, Mendonça P, Hopper A (2002) Trip: a low-cost vision-based location system for ubiquitous computing. Pers Ubiquitous Comput 6(3):206–219
8. Dorfmueller-Ulhaas K, Schmalstieg D (2001) Finger tracking for interaction in augmented environments. In: Proceedings of the IEEE international symposium on augmented reality. IEEE Press, New York
9. Fails J, Olsen D (2003) A design tool for camera-based interaction. In: Proceedings of CHI. ACM Press, New York, pp 449–456
10. Gutwin C, Fedak C (2004) Interacting with big interfaces on small screens: a comparison of fisheye, zoom, and panning techniques. In: Proceedings of the 2004 conference on Graphics interface. Canadian Human-Computer Communications Society, pp 145–152
11. IDC Research. Moving pictures 2003: Worldwide camera phone survey, forecast, and analysis, 2003–2007. http://www.idcresearch.com/
12. Johnson J (1995) A comparison of user interfaces for panning on a touch-controlled display. In: Proceedings ACM CHI, pp 218–225
13. Kato H, Billinghurst M (1999) Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In: International workshop on augmented reality (IWAR), pp 85–94
14. Kindberg T, Barton J, Morgan J, Becker G, Caswell D, Debaty P, Gopal G, Frid M, Krishnan V, Morris H, Schettino J, Serra B, Spasojevic M. People, places, things: web presence for the real world. In: Proceedings of the 3rd IEEE workshop on mobile computing systems and applications (WMCSA'00). IEEE Press, New York
15. Kindberg T, Sellen A, Geelhoed E (2004) Security and trust in mobile interactions: a study of users perceptions and reasoning. In: Proceedings of UbiComp 2004, vol 3205 of LNCS. Springer, Berlin Heidelberg New York
16. Krueger MW, Gionfriddo T, Hinrichsen K (1985) Videoplace—an artificial reality. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM Press, New York
17. Bango Ltd. Bango spots link camera phones straight to mobile content. Press release, available from http://www.bango.net/
18. High Energy Magic Ltd. The spotcode framework. Available for download from: http://www.highenergymagic.com/
19. Scott MacKenzie I, Buxton W (1992) Extending fitts' law to two-dimensional tasks. In: CHI '92: Proceedings of the SIGCHI conference on human factors in computing systems. ACM Press, New York, pp 219–226
20. MacKenzie IS (1991) Fitts' law as a performance model in human-computer interaction. PhD Thesis, University of Toronto
21. Masnik M (2004) Mobile barcode scanning catching on in Japan. The feature (2nd July 2004). http://www.thefeature.com/article?articleid = 100700&ref = 5960372
22. Rekimoto J, Nagao K (1995) The world through the computer: computer augmented interaction with real world environments. In: ACM symposium on user interface software and technology. pp 29–36
23. Rohs M, Zweifel P (2005) A conceptual framework for camera phone-based interaction techniques. In: Proceedings of pervasive 2005, vol 3468 of LNCS. Springer, Berlin Heidelberg New York
24. Scott D, Sharp R, Madhavapeddy A, Upton E (2005) Using visual tags to bypass bluetooth device discovery. SIGMOBILE Mob Comput Commun Rev 9(1):41–53
25. Sharp R (2004) New ways to maximize cameraphone technology. DeviceForge (1st July 2004). http://www.deviceforge.com/articles/AT5785815397.html
26. Smith M, Davenport D, Hwa H (2003) Aura: a mobile platform for object and location annotation. In: Adjunct proceedings of Ubicomp 2003
27. Toye E, Sharp R, Madhavapeddy A, Scott D (2005) Using smart phones to access site-specific services. Pervasive Computing 4(2):60–66
28. Want R, Fishkin KP, Gujar A, Harrison BL (1999) Bridging physical and virtual worlds with electronic tags. In: CHI '99: Proceedings of the SIGCHI conference on human factors in computing systems. ACM Press, New York, pp 370–377