RWTH AACHEN UNIVERSITY

Evaluation Strategies for Pervasive Games

Diploma Thesis at the
Media Computing Group
Prof. Dr. Jan Borchers
Computer Science Department
RWTH Aachen University

by
André Kuntze

Thesis advisor:
Prof. Dr. Jan Borchers

Second examiner:
Prof. Dr. Reinhard Oppermann

Registration date: March 12th, 2007
Submission date: July 31st, 2007

Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Schriften entnommen sind, sind als solche kenntlich gemacht.

Aachen, den 31. Juli 2007

Contents

List of Figures

List of Tables

Abstract

The development of HCI systems generally follows the iterative process of applying design, implementation and analyzing to ensure that the system fits the user's needs. Methods to support the process are rare for ubiquitous computing applications.

This thesis focuses on evaluating pervasive games. Therefore analyzing the effectiveness and appropriateness of traditional human computer interaction methods to evaluate design, interactions, and user experience in the context of pervasive games are the topic of this thesis.

Specifically here, the focus will be on supporting a human-centered iterative design process with formative evaluations of pervasive games, which are intended to shape and improve designs.

Pervasive games are played in the real world, so traditionally the evaluation takes place in the field. Thus we will first focus on the examination of methods including product-interactive focus groups and analysis of interviews as well ass video recordings using grounded theory.

But evaluation in the field is not always appropriate. Therefore afterwards we simulate the experience of playing in a city while being in a laboratory. This includes developing a game simulator for REXplorer, a permanently installed pervasive game which helps tourists to learn about the city of Regensburg, Germany.

Putting these methods to practice by evaluating and improving REXplorer, we enable the main contribution of this thesis—a comparison between field evaluations and laboratory experiments for pervasive games.

Überblick

Die Entwicklung von Mensch-Maschine-Systemen folgt standardmäßig einem iterativen Prozess bestehend aus Design, Implementierung und Analyse um sicher zu stellen, dass das System den Anforderungen des Benutzers gerecht wird. Methoden dieser Art fehlen zu einem großen Teil für Anwendungen im Bereich der ubiquitären Computertechnik.

Diese Diplomarbeit beschäftigt sich mit der Evaluierung mobiler Spiele. Dazu werden die Effektivität und Angemessenheit traditioneller Methoden der Mensch-Maschine-Systeme überprüft und getestet, inwieweit sie zur Evaluierung des Designs, der Interaktion und der Eindrücke des Benutzers von mobilen Spielen geeignet sind.

Im Speziellen wird der Fokus auf die Unterstützung des benutzerzentierten Designprozesses mit stufenweiser Evaluierung gelegt, welcher dazu dient das Design von mobilen Spielen zu verfeinern und zu verbessern.

Mobile Spiele werden in der realen Welt gespielt. Daher findet ihre Evaluierung generell im Feld statt. Daher werden wir zuerst den Fokus auf die Bewertung traditioneller Methoden legen. Dies beinhaltet product-interactive focus groups und die Analyse von Interviews, so wie Videoaufnahmen basierend auf gegenstandsverankerter Theorienbildung.

Allerdings ist die Evaluierung im Feld nicht immer angemessen. Daher beschreiben wir anschließend, wie die Erfahrungen des Spielens in der Stadt im Labor simuliert werden können. Dies beinhaltet die Entwicklung eines Spielesimulator für REXplorer, einem permanenten Stadtspiel, welches in Regensburg dazu dient, Touristen die Stadt näher zu bringen.

Anhand der praktischen Anwendung der Techniken auf REXplorer und der damit verbundenen Evaluierung und Verbesserung des Systems zeigen wir den Hauptbeitrag dieser Diplomarbeit - den Vergleich zwischen der Evaluierung im Feld und Experimenten im Labor für mobile Spiele.

Acknowledgements

First i would like to thank my thesis advisor, Prof. Dr. Jan Borchers, for giving me this interesting topic and his great lectures, which encouraged me to take this path.

This thesis deals with a emerging topic and i imagined it to be hard to find a second examiner. So i would like to thank Prof. Dr. Reinhard Oppermann for quickly accepting my thesis topic.

Special thanks go to Tico Ballagas for giving me valuable feedback in every phase of my thesis. He was a big help for me.

Concerning my user tests in Regensburg Tico Ballagas and Steffen P. Walz made this time very productive and enjoyable. Especially Steffen's feedback on my interview guideline and questionnaire was highly welcomed and very useful, besides getting all the players for our field studies.

I also want to thank Thorsten Karrer for sharing his great template that saved me a lot of time. I would not have been able to put the thesis in this form without it.

Finally i dedicate this work to my parents, Mechtild and Richard Kuntze, and my girlfriend, Daniela Krüger, for their ongoing support and love during my whole studying. I would never have been able to reach my goals without them.

Thank You!

Conventions

Throughout this thesis we use the following conventions.

Text conventions

Definitions of technical terms or short excursus are set off in coloured boxes.

EXCURSUS:
Excursus are detailed discussions of a particular point in a book, usually in an appendix, or digressions in a written text.

Definition:
Excursus

The whole thesis is written in American English.

Download links are set off in coloured boxes.

File: myFile[a]

[a] http://media.informatik.rwth-aachen.de/~ACCOUNT/thesis/folder/file_number.file

Chapter 1

Introduction

"All work and no play makes Jack a dull boy."

—Jack Torrance, The Shining

Ubiquitous computing

Today there is a shift from the desktop computer we know to ubiquitous computing. Mark Weiser [1999] states: "The most profound technologies are those that disappear". This means that the computer fades into the background and supports the user and his task. The results are new applications like pervasive games.

1.1 Pervasive games

Pervasive games change the way people play computer games

Traditional computer games are played at the desktop without additional participants. Pervasive games make people move from the desktop back to the outside world, meet people and, with edutainment oriented games, learn something. There is a great potential to support the life long learning process in a fun way.

Evaluating pervasive games will be the main topic of this thesis but before we can dive into the subject we have to define more precisely what we mean by pervasive games.

1.1.1 Defining pervasive games

Pervasive games are ubiquitous computing applications. There are several definitions of pervasive games. Some examples are [Montola, 2005], [Nieuwdorp, 2005] and [Hinske et al., 2007]. To understand this term we first have to define *games* and *pervasive*.

Games

As an overview the common understanding of *games* is sufficient. An explicit and extensive definition is presented in [Hinske et al., 2007].

Magic circle

Games define a so called *magic circle*. This term was first described in [Huizinga, 1971] and applied to digital games in [Salen and Zimmerman, 2003]. Nieuwdorp offers a meaningful description: "When we look at theoretical literature surrounding games, there is one concept that is often used to describe the soap bubble that is the game world in relation to the environment outside it."

Breaking the magic circle

This concept restricts traditional games to playing in certain spaces at certain times by certain players. Pervasive games are different as they break the magic circle by *expanding* in social, spatial and temporal aspects [Montola, 2005].

Expansions

In detail the three expansions can be specified as:

- *Spatial* Spatial expansion indicates that the socially constructed location of the game is unclear or unlimited.
- *Temporal* Pervasive games expand temporally from the explicit play sessions; the socially constructed game session is interlaced and mixed with ordinary life.
- *Social* In the unexpected places and times where the expanded games are played, unexpected people make a difference regarding the gameplay.

A detailed elaboration of terms and concepts can be found in [Montola, 2006]. We choose the following definition of

pervasive games by Montola because it fits our own considerations well:

PERVASIVE GAMES CF. [MONTOLA, 2005]:
A pervasive game is a game that has one or more salient features that expand the contractual magic circle of play socially, spatially or temporally.

Definition: *Pervasive games cf. [Montola, 2005]*

Below we emphasize this definition by offering examples of current pervasive games and analyzing them according to the three expansions mentioned above.

Most of the following games are still research projects. The field is still emerging but there are already many interesting examples, which are categorized in [Magerkurth et al., 2005]. For us the games, which were evaluated in practice, are of particular interest and so the following subset was chosen.

Choosing already evaluated games

1.1.2 Example: Epidemic Menace

Epidemic Menace

We start with Epidemic Menace - a game developed within the IPerG (Integrated Project on Pervasive Gaming: IPerG[1]). "In Epidemic Menace players become medical experts and need to save mankind from threatening virus mutations. A villain scientist, craving for power, creates a lethal virus mutation and contaminates campus Birlinghoven. From there the viruses shall spread and infect all humans. To master this threat, expert teams - the players - are appointed. They have the task to destroy the viruses before they manage to escape the campus and to uncover how this could have happened." [Lindt et al., 2007].

As a cross media game Epidemic Menace is "played across different devices and media channels and that employ a wide variety of gaming devices and media channels in the game play, including state-of-the-art mobile and stationary computing devices as well as more traditional communication and information channels such as television broadcast or print media" [Lindt and Adams, 2005].

[1] http://iperg.sics.se/

This involves several different gaming devices and game spaces which contribute to the experience. The game was tested in August 2005. Figure 1.1 shows Epidemic Menace in action.

Figure 1.1: The crossmedia game Epidemic Menace during a play session (cf. [Lindt et al., 2007])

In [Montola, 2006] Epidemic Menace was already described according to its expansions. The author concludes:

Spatial expansion of Epidemic Menace

Spatial expansion
Epidemic Menace is played by teams having both online players getting tactical overview of the game area, and by onstreet players doing the fieldwork in the campus area. Although the game area is limited to the campus, players find new areas during the game from inside the campus buildings, and also get to explore some scenographed physical areas. Epidemic Menace is also adaptronic, as the virus reproduction and behavior depend on the weather outside. On a beautiful day the players need to spend more time outside.

Temporal expansion of Epidemic Menace

Temporal expansion
Varies. In some experimental versions the game goes on around the clock. Thus, the players playing less actively suffer in the game. The first game was played during day time only with all players present.

Social expansion
Social expansion of Epidemic Menace

The most of the game time the only social expansion is created by the players tracking the viruses in campus area during office hours wearing the shirts of fictional medical agency. Thus game related persons can be distinguished from the employees on the campus.

1.1.3 Example: Feeding Yoshi

Feeding Yoshi

The second example for a pervasive game is Feeding Yoshi which was developed in a partnership between researchers in Glasgow University's Equator group, the University of Nottingham, and the University of Lincoln.

"Feeding Yoshi is a mobile multiplayer game that is played over a relatively long period - the game we report on here lasted a week. [...] The aim of Feeding Yoshi is for each team of players to collect as many points as possible, by feeding Yoshis the fruits they desire." [Bell, 2006].

The game takes place in a city of choice. Players get a PDA as equipment. Secured networks are used as Yoshis and unsecured as plantations. At a Yoshi the user can pick up seeds which he can dibble at a plantation. This will generate fruits, which can be picked up and feeded to Yoshis for gaining points. When users are near to each other they can exchange foods and seeds.

Feeding Yoshi is based on seamful design

The optimal situation for a game like Feeding Yoshi would be an area-wide open network infrastructure, which is not available. The designers did not try hide the absence of open networks but used that as a part of the game itself. Utilizing the network infrastructure in such a way is an example of *seamful design* [Barkhuus, 2005].

Feeding Yoshi expands the magic circle in the following areas:

Spatial expansion
Spatial expansion of Feeding Yoshi

The main trial of Feeding Yoshi took place in Glasgow, Derby and Nottingham. To play the game people had to move through the city and find networks (i.e. Yoshis and

plantations).

Temporal expansion of Feeding Yoshi

Temporal expansion
The evaluation of the game was scheduled for a whole week but it could also be played longer.

Social expansion of Feeding Yoshi

Social expansion
On the one hand, Feeding Yoshi was not designed to involve bystanders. On the other hand, the participants may attract bystanders when they run around in the city and communicate with Yoshis, plantations and other players. The evaluation was conducted with people who knew each other, so there should not have been confusions.

1.1.4 Example: Can You See Me Now?

Can You See Me Now?

"CYSMN (Can You See Me Now?) is a game of catch - but with a twist. Online players are chased through a virtual model of a city by *runners* or street players, who have to traverse the actual city streets in order to capture the online players." [Benford et al., 2006].

This means the game mixes reality and online play. The online players run through a virtual city and are captured by runners, professional staff from Blast Theory, who run around in the real world city.

The position of the runners is tracked with wireless network connections plus GPS receivers and appears online. Equiped with a handheld device the runners can see the online players. The connection between online and offline players is emphasized with an audio connection.

The expansions of the magic circle were already analyzed in [Montola, 2006]:

Spatial expansion of CYSMN

Spatial expansion
The streets used for the game are open for business, so the runners need to maneuver in the traffic. Gaming area is limited, both in physical and in virtual space.

Temporal expansion
Negligient. A game session normally takes about two hours.

Temporal expansion of CYSMN

Social expansion
The socially expanded feature of Can You See Me Now? is the runners interacting with bypassers on the streets.

Social expansion of CYSMN

1.1.5 Example: REXplorer

The last game we introduce is REXplorer. It was the motivation for this thesis and therefore the following description will be more detailed than the ones above.

REXplorer as motiviation for this thesis

"REXplorer is a mobile, pervasive spell-casting game designed for tourists of Regensburg, Germany. The game uses location sensing to create player encounters with spirits (historical figures) that are associated with historical buildings in an urban setting" [Ballagas et al., 2007].

A decent overview of the game can be found in [Walz and Ballagas, 2007]: "In the game, historical and mythological spirits are stationed at touristic points of interest throughout the mostly Gothic and Romanesque city core of Regensburg. Players rent a special *paranormal activity detector* - a device composed of a mobile phone and a GPS receiver in a custom designed shell - at Regensburg's tourist information. Players interact with the location-based and site-specific spirits by performing a gesture, i.e. by waving the wand-like detector through the air in a specified fashion, thus *casting a spell*. Situated gestures allow players to evoke and communicate with spirits to receive and resolve quests. With their detector, players can also take pictures, or shoot videos, which appear on each player's individually generated souvenir, a weblog. The weblog also maps a player's route, describes spirits a player has encountered, and lists books and deepening URLs for each character and site". Figure 1.2 shows the game in action.

Feature overview of the location-based game REXplorer

Figure 1.2: REXplorer is designed to be played in groups of two or three

Application of pervasive persuasive design tactics

REXplorer was developed in a collaboration between the RWTH Aachen and the ETH Zürich. Applying *pervasive persuasive design tactics* [Walz and Ballagas, 2007] the game's rhetoric was specified to Regensburg. In detail the applied tactics were *formal pervasive persuasive design tactics*

- Stageability: Architecture of the game board
- Travelability of the game board
- Navigability and spatial sequentiality

- Spatial layout-progression and graphing
- Detector functionalities
- Reward structure
- Persuading the player to replay

and *dramaturgical pervasive persuasive design tactics*

- Bridging the worlds for a premise
- Persuasive detector character design
- Persuasive character design
- Narrative persuasion through emotional and spatial bonding

As a very short description we can say that for the design of REXplorer the specific properties of Regensburg, like the historical background and the city structure, was taken into account. The result is a game that specifically fits Regensburg. One outcome of the design process was a unique device, which can be seen in Figure 1.3.

Figure 1.3: The final design of the REXplorer device

The device is handed out, besides a paper map, and is used to communicate with *spirits* in the city. In [Ballagas et al., 2007] a design rationale is given:

- Designing for Narrative Consistency
- Balancing Competitiveness and Leisure
- Balancing Cooperative Experience vs. Outdoor Play
- Designing for a *Heads Up* Experience

Some of these points resulted from user tests and their analysis which are described in chapter 3 and 4.

When this diploma thesis started REXplorer was in the beta phase. An iterative design process was chosen for developing REXplorer and so this was a chance to apply the evaluation strategies we present here in practice. The game was released on the 29th of June. It breaks the magic circle in the following sense:

Spatial expansion of REXplorer

Spatial expansion
REXplorer is played all over the city of Regensburg. The players have to move from one hotzone to the next to fulfill their quests.

Temporal expansion of REXplorer

Temporal expansion
The temporal expansion could be neglected because the game sessions only take about one and a half hour. In our considerations we query if the restricted time is a problem for our users. At the end of the game a blog is created, which can be viewed online after the game. It includes the user's path and taken photos.

Social expansion of REXplorer

Social expansion
Bystanders are not a directly a part of the game. REXplorer takes place in a public space and so communicating with the gestures and the character's voices attract attention. The game was also designed to be playable in small groups.

The unique features of pervasive games described above can lead to problems when evaluating these games. The definition of evaluation and the particular problems with it will be the topic of the next section.

1.2 Evaluation

Evaluation of computer interfaces is a well established field. So the question rises: Why is it necessary to rethink the evaluation of pervasive games? This section will emphasize the question and offer answers to it.

1.2.1 Defining evaluation

Every standard development process for interactive software should include a phase for testing if the system fulfills the user requirements. [DIX, 2003] defines this as evaluation:

EVALUATION:
Evaluation tests the usability and functionality of an interactive system.

Definition: *Evaluation*

Dix states: "Evaluation has three main goals: to assess the extent and accessibility of the system's functionality, to assess users' experience of the interaction, and to identify any specific problems with the system."

Three goals of evaluation

Depending on the development process the evaluation can take place at different times. For example a standard way to develop software in earlier days was the waterfall model [Preece, 2002]. The software was developed by programmers for their peers. If the user requirements were collected at all it took place at the beginning of the process and then the whole system was developed [Shneiderman, 1998]. When releasing the software the user does a summative evaluation to judge the whole system [DIX, 2003].

Summative evaluation

The problem with that model is that the requirements are never right in the first place and applying the waterfall model can lead to unusable software. While it is cheap to discard ideas at the beginning of the development process it costs a lot to correct potential errors in already released software.

Iterative design process

So an iterative design process was established in the HCI field [Nielsen, 1993]. Here the design process is made up of four main phases plus an iteration loop.

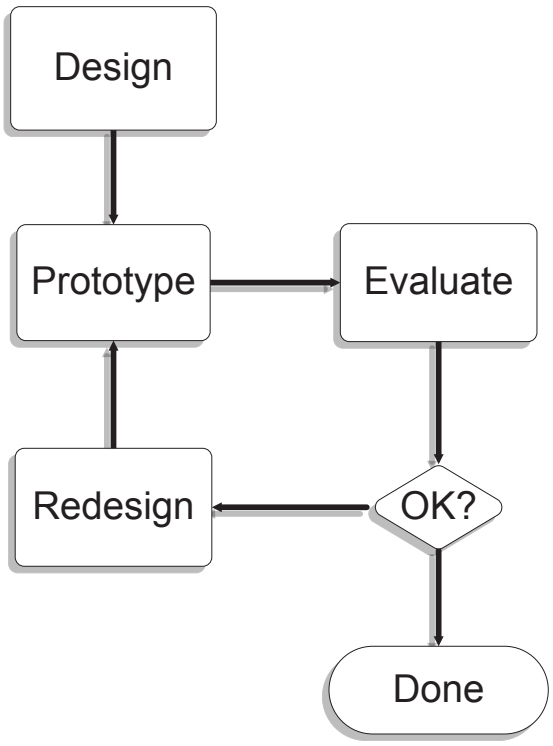

Figure 1.4: The iterative design process adapted from [DIX, 2003]

The different phases are described in short below:

Requirements engineering

Requirements engineering

Requirements engineering is not a phase but the start of the iterative design process. Each development process should begin with collecting the user's needs and center development around them. The problem is that these will change during development and so it is not enough to only get feedback at this early stage.

Design

Design

In this phase a design solution is compiled to fulfill the requirements. The transformation is a difficult but important step. There are several aids that can help to improve

it like design guidelines in [App, 1992], design patterns in [Borchers, 2001], or heuristics in [Shneiderman, 1998] and [Norman, 2002].

Prototype
Prototype

Prototypes are realized design ideas, that can have different fidelities to fit the status of the development. Beginning the development early prototypes based on pen and paper are created and tested. During the design process prototypes are further refined and converge to the final product.

Evaluate
Evaluate

This phase is most important for this thesis. Prototypes are tested with or without users to find errors and flaws. There are several ways to conduct evaluation which will be explained in detail later.

Redesign
Redesign

The results of the evaluation are used to modify the product's design. This is used to develop new prototypes, which are traversed in the design process again.

Implementation and Deployment
Implementation and Deployment

When the design of the system is concrete enough the system gets implemented. The result is not ready for release but a high fidelity prototype, which should be evaluated and improved to produce the final product.

Formative evaluation

To fit the iterative design process evaluation is formative instead of summative. The difference is that the evaluation takes place several times during the design process and not only once at its end.

Formative evaluation is not only useful for traditional systems but also for pervasive games. So this diploma thesis deals with that concept. To evaluate a system many testing methods can be used. Below some of these methods are explained in short and a basis for decision-making is given.

Evaluation may take place in the field or in the laboratory. It can be done with experts from the field or include users. The collected data can be quantitative or qualitative.

Characteristics of evaluation methods

[DIX, 2003] explains eight factors to distinguish evaluation

techniques:

- the stage in the cycle at the evaluation is carried out
- the style of evaluation
- the level of subjectivity or objectivity of the techniques
- the type of measures provided
- the information provided
- the immediacy of the response
- the level of interference implied
- the resources required

He differs between analytic (cognitive walkthrough, heuristic evaluation, review based and model based), experimental and query (experiment, interviews and questionnaires) and observational (think aloud, protocol analysis and post-task walkthrough) methods.

The evaluation method meets the status of the prototypes if the right balance between the fidelity of the testing method and the fidelity of the prototype is given. Dix also states that it is important to bring the user in as early as possible: "However, useful as these techniques are for filtering and refining the design, they are not a replacement for actual usability testing with the people for whom the system is intended: the users."

1.2.2 Traditional methods

Traditional evaluation methods

It is hard to choose the right evaluation method because there are plenty of options and it can be difficult to decide in which granularity to test the system. For example having a full fledged test system when only a paper prototype is tested which will be thrown away anyways is unnecessary and does not amortize.

1.2.3 Ubiquitous Computing

Ubiquitous computing makes it even harder to apply evaluation as additional complexity is added. In [Carter et al., To appear.] a list of special attributes for these systems is given:

- Sensing and Actuation
- Scale
- Many Tasks
- Many People
- Many Devices
- Many Places

Ubiquitous computing adds complexity to the evaluation

Carter concludes: "The sensing and scale issues of ubicomp make studying these systems more challenging than traditional desktop applications."

So traditional methods may not be appropriate because they do not scale with the requirements of ubiquitous computing. The third chapter will show how a combination of traditional methods can be used to improve the scale situation.

To get rid of these problems new methods for evaluation should be developed. The problem is that evaluation research in that field is not very active. In [Kjeldskov and Stage, 2004b] a literature review from

- Proceedings of the International Conference on Mobile HCI: 1998, 1999, 2001
- Proceedings of the ACM Conference on Computer-Human Interaction: 1996-2002
- ACM Transactions on Computer-Human Interaction (TOCHI): 1996-2002

was executed. The result shows a dilemma. On the one hand, the usage of traditional methods with ubiquitous computing is restricted and, on the other hand, no new methods are developed. That makes evaluating ubiquitous computing applications particular hard.

1.2.4 Pervasive games

One part of these games has its origin in ubiquitous computing and suffers from the same added complexity. The other part is derived from games. That means we do not only have to find out if people were able to interact with the system but if they actually liked playing the game. This includes points that contradict our natural understanding of usability.

For example one part of usability should be not to waste the user's time. The approach would be to let him fulfill his task as quick as possible. In a game that may be different. If the player likes the game they actually might want to spend a lot of time with it.

Another point would be that usability tries to make interacting with the system as easy as possible. But if games are not challenging, they might be boring after a few minutes of play.

This leads us to playability, "the instantiation of the general concept of usability when applied to videogames" [Fabricatore et al., 2002].

Considering the definition of pervasive games in [Montola, 2005], we also have to find out how the expansions influence the playability of the game.

So the evaluation of pervasive games has to deal with many new aspects of interaction, which result in the following requirements:

- Testing if users enjoy playing the game

- Testing fulfillment of the user's requirements in respect to expansions in social, spatial and temporal areas

- Testing if the game is usable in respect to aspects unique to ubiquitous computing

Altogether we can derive that games must not only be evaluated regarding to traditional usability questions, but integrate the player's experience and enjoyment.

Even if we could get rid of all the problems described above and choose the collected data carefully there might be huge amounts of data. So another step is to analyze the data and get the results one needs from it.

1.2.5 Analyzing data

Analyzing data

Two types of data can be collected—quantitative and qualitative.

Analyzing Quantitative data

If quantitative data was gathered it can be analyzed using statistical methods. The data in this thesis is qualitative and so we will not discuss these methods here. The interested reader can find more information online: Electronic Statistics Textbook [2]

Analyzing Qualitative data

For qualitative data there are different approaches. Two examples are *rapid ethnography* and an *affinity analysis*.

The first example can be found in [Millen, 2000]: "Rapid data collection is of limited use to the corporate ethnographer if the data analysis still proceeds at the painstakingly slow rate that is typical of most qualitative research." Millen notes that two general approaches can help: computer-assisted analysis and collaborative data analysis. His method of choice for collaborative analysis is concept mapping.

[2] http://www.statsoft.com/textbook/stathome.html

The second example, and our technique of choice, follows the idea of grounded theory: Creating an affinity diagram [Beyer, 1997]. The analysis can be done collaboratively and used to sort huge amounts of data. A detailed description of the approach in practice can be found in chapter 3.

1.3 Thesis structure and contribution

Every new technology development includes the danger of creating an usable system. This means developers have to make sure that pervasive games like every HCI system fit the user requirements.

Presenting different methods to test the system and applying them to REXplorer, to show them in practical use, is the first part of our contribution. The second is showing a alternative way for conducting user tests for pervasive games in the laboratory. We summarize with comparing the laboratory study to a field test. For that we structured the thesis in the following way:

Chapter 2 presents related work to this thesis. The main focus lies on the description of evaluation methods other pervasive gaming projects applied. As these games are also ubiquitous computing applications, an overview about evaluation methods for ubiquitous computing is included, too.

We will judge the existing methods according to their capabilities of supporting the evaluation of the expansions in spatial, temporal and social aspects.

In chapter 3 we choose a combination of methods for field evaluation and apply them to REXplorer. The method of choice for data analyzing is an affinity diagram. The findings from the user test and a conclusion, about how effective the evaluation and the analysis was will, close the chapter.

Chapter 4 focuses on the comparison of evaluation in the field and in the lab. First a second user test in the field is

conducted. The results are captured with questionnaires. After that we present the iterative development of a game simulator for the lab and show a cost effective experimental setup.

Finally, chapter 5 summaries the thesis and gives ideas for future development.

With the theoretical background explained, we will judge approaches other teams applied to evaluate their systems in the next chapter.

Chapter 2

Related work

"I like work: it fascinates me. I can sit and look at it for hours."

—Jerome K. Jerome, Three Men in a Boat

In the first chapter, we stated that research in the evaluation sector for pervasive games is just starting to emerge. So in this chapter we present available work which is related to the topic and argue how it could be used to successfully evaluate pervasive games.

As we already gave an overview of traditional methods in chapter 1, we will only focus on evaluation literature from the ubiquitous computing and pervasive game sector.

If a paper contains the description of a concrete evaluation method it will be judged according to the potential of evaluating the expansions we explained in chapter 1. In addition, we add a technical expansion to describe the evaluation of controllers and the interaction with it.

2.1 Evaluation of ubiquitous computing

Pervasive games are ubiquitous computing applications and so it makes sense to take a look at the available meth-

ods from that field first.

2.1.1 General literature

Kjeldskov observed closely related topics than the ones we considered. So his findings will be explained on detail below. In [Kjeldskov and Stage, 2004b] an overview about available research papers, regarding evaluation of ubiquitous computing applications is given. In the same paper Kjeldskov contributes to the topic with the comparison of six techniques for the evaluation of mobile devices:

Six techniques for the evaluation of mobile devices

1. Sitting on a chair at a table
2. Walking on a treadmill at constant speed
3. Walking on a treadmill at varying speed
4. Walking at constant speed on a course that is constantly changing
5. Walking at varying speed on a course that is constantly changing
6. Dividing the user's attention between conscious actions and use of the mobile system

As a reference walking in a pedestrian street and executing a typical use situation is added. Regarding to these techniques one experiment including the need for navigation in physical space, using technique 1-5, and one requiring the devision of attention, utilizing technique 6, were conducted.

Kjdelskov derived some surprising results from these setups. The most interesting ones are summarized below.

The sitting technique

The sitting technique revealed that participants who just sat at a table found the most flaws. Even if most of the additional usability errors were cosmetic this finding was surprising. Kjeldskov gives the reduced workload, which resulted from not moving around, as reason. Besides that no

significant difference was found between the different evaluation techniques.

data collection in the field

For *data collection in the field* Kjeldskov has chosen a less intrusive approach with taking notes and video recording. He describes problems as users tend to put their hand between the camera and the device while interacting, which made it hard to get the wanted data.

A solution could be a promising looking portable configuration for high quality data collection, which is introduced in [Kjeldskov et al., 2004a]. The paper will be discussed below.

An expansion to the recordings could be a product-interactive focus group interview [Lee and Bhatkhande, 2004]. In our own experiments we had a good experience with conducting such an interview after the test. This will be explained in detail in chapter 3. While this technique requires more effort it will bring more detailed results plus insights from the users themselves.

involving social context

Another interesting point for our own experiments were problems with *involving social context*. Regarding pervasive games this means that the social expansion is hard to test. Chapter 4 will explain this issue in detail.

Kjeldskov gives further insight about the evaluation of mobile devices in the laboratory. In [Kjeldskov et al., 2004a] he discusses if it is necessary to conduct evaluation sessions in the field.

Mobileward

For comparison he executes an experiment with *Mobileward*, a context-aware electronic patient record prototype, using it in a laboratory setup and a field study. Following we summarize the conclusions from the experiment.

little added value of taking the evaluation into a field condition

Interestingly Kjeldskov concludes that there is only *little added value of taking the evaluation into a field condition*. Nearly all of the errors could be found in both setups and *both the lab and the field revealed context-aware related problems*.

Another finding is that the *lack of control undermined the extendibility of the field condition*. On the one hand, he describes

that the participants did not use some of the system's features in the field condition. This happened because people could not be urged to use the functions like in a laboratory study. The result is that these features were left untested. On the other hand, Kjeldskov points out that the laboratory condition may produce false positives.

As a consequence the number of errors are hard to compare because some of the errors may not come up or might be irrelevant.

Collection high quality data

During the test data was collected with a unique setup. The setup contained a clip-on camera, that recorded the user's interactions with the device and *facilitated high-quality data collection of mobile use*. Figure refkjeldskov shows the setup.

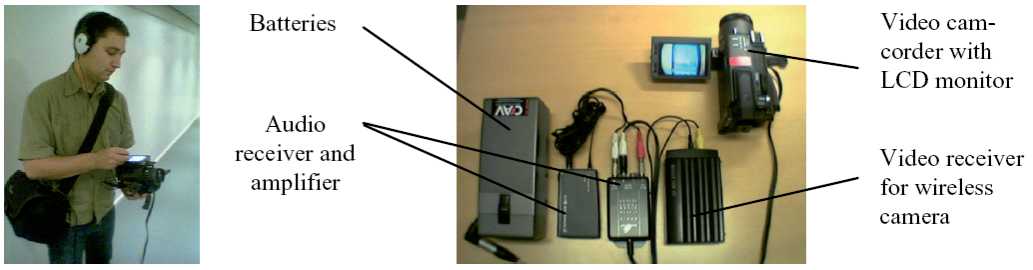

Figure 2.1: Data capturing for high quality data (cf. [Kjeldskov et al., 2004a])

A similar approach was unnecessary for REXplorer's evaluation since the display only contains few information and the clipped-on camera would have restricted the easy execution of the gestures.

Research directions

Another active researcher providing studies about how to evaluate ubiquitous computing applications is Carter. In [Carter et al., To appear.] they contribute with an extensive overview of five central challenges for *ecologically valid design* and suggest research strategies, methods and tools to address these. The five challenges are:

Conversations with materials
This point means that it should be able to bring tangible things into discussions. The development effort would include to discover new material that can be used.

Prototyping for evaluation
Tools for easy prototyping are needed to discuss ideas.

Supporting in-the-world evaluation
Ubiquitous computing should be evaluated in the real world. Therefore it is necessary to be able to build robust prototypes, minimize deployment costs and minimize per-participant costs.

Support for machine learning and sensor-based interaction
Aspects of machine learning are needed to support the design process and create richer applications, that can adjust themselve to a certain degree.

Data sparsity
Developers need tools for easier data collection. One example is Momento which will be discussed in detail in the next section.

These points give valuable insights about what kind of development would be useful for the evaluation of ubiquitous computing applications.

2.1.2 Momento

Momento

In [Carter et al., 2007] a system based on a set of tools and communication via SMS/MMS, HTTP or a special Context Toolkit. The purpose is to offer a system that can be configured without writing source code so that anyone with basic computer skills can use it. The system architecture can be seen in Figure 2.2:

The application is open source and can be found on Momento's Homepage[1]

Spatial expansion coverage
As Momento runs on several mobile devices the spatial expansion is not restricted and depends only on the application.

[1] http://www.m0ment0.com

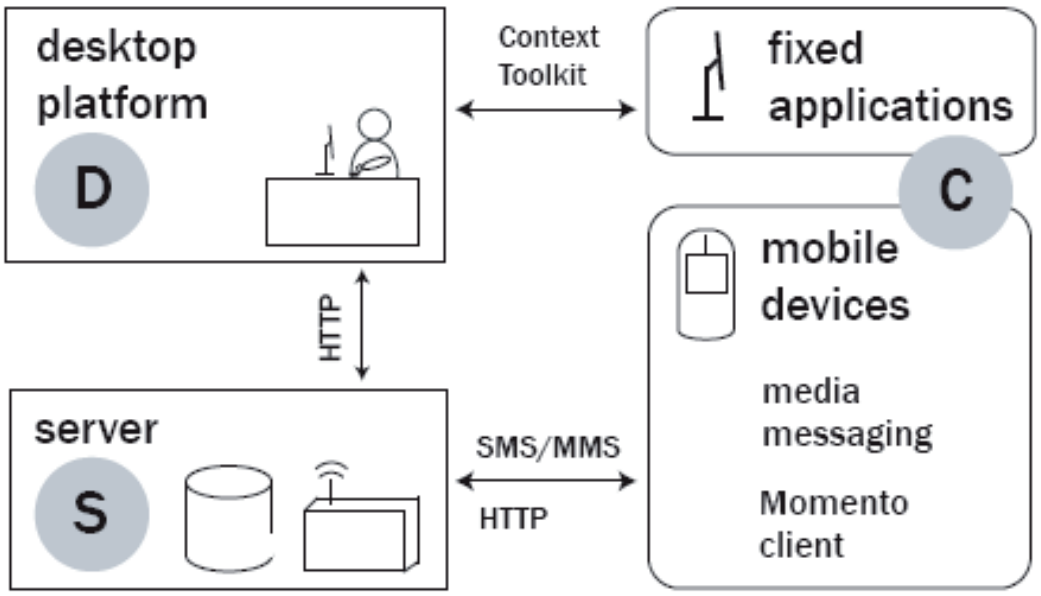

Figure 2.2: Momento's system architecture (D = desktop platform, S = server, C = clients)

Temporal expansion coverage
The system is not restricted temporal. Tests in the paper take up to 51 days.

Social expansion coverage
Momento itself does not support testing the social expansion. With the general structure of the system it should be easy to extend the software for further testing in that direction.

Technical expansion coverage
The software does not have a special support for testing devices. Containing of software which runs on most mobile phones it should be easy to extend the software and collect device data.

Conclusion
Altogether Momento is a promising system. The support for configuring the system without writing source code lowers the barrier for useful testing. But on the one hand, as SMS services are widly available, communicating with text and multimedia messages makes it easy to communi-

cate with a server without caring about the connection. On the other hand, only interacting with these messages can make it hard to collect enriched data. For example there is no support for recording the user's interaction with the system or similar.

Therefore we preferred a more traditional approach with video recording for our own considerations in chapter 3.

2.1.3 Ubiwise

Ubiwise

Ubiwise is a simulator for ubiquitous computing applications and presented in [Barton and Vijayaraghavan, 2003]. Featuring a modern 3D graphics engine, QUAKE III Arena makes it possible to present two views: user's first-person view of the physical environment and a view which shows the device and objects the user can interact with. This way it only needs some programming to create new devices and interactions.

The system in action can be seen in Figure 2.3.

As Ubiwise creates a virtual system the real world expansions can be neglected. Therefore the following overview focus on the virtual coverage.

Spatial expansion coverage
Taking a look on some QUAKE III Arena maps the virtual spatial expansion could be covered well. But as the creation of the maps can take time and effort it is questionable if the cost for developing large environments would amortize.

Temporal expansion coverage
The system itself only needs a running server so play time could be easily extended. Otherwise the participants would have to stay at the monitor the whole time so the virtual interaction is not possible.

Social expansion coverage
The system supports more than one user to interact at the same time. As QUAKE III Arena features a multiplayer

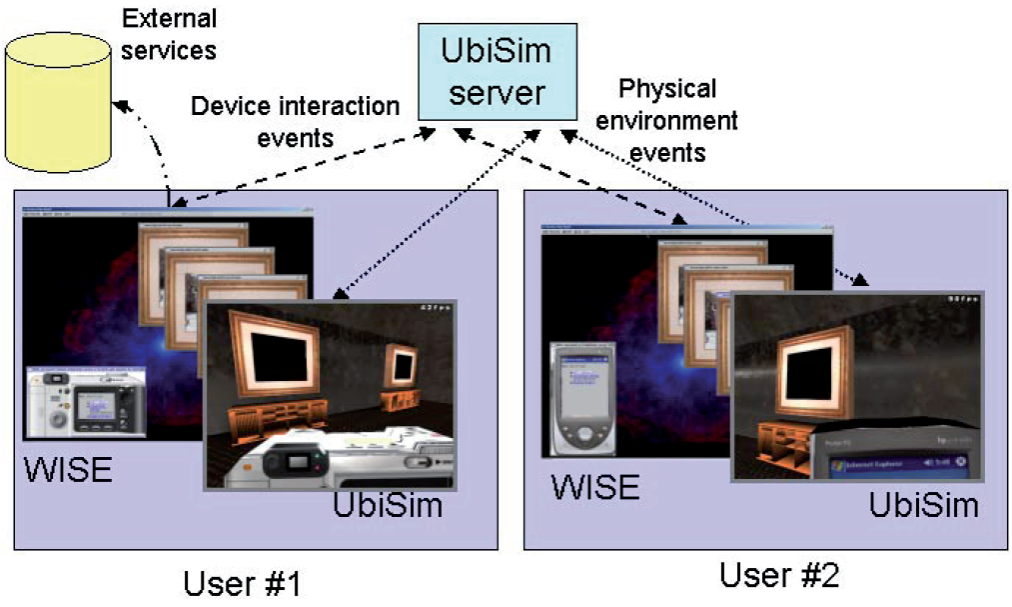

Figure 2.3: UbiWise

mode the number of users can be extended. But then the interaction between the virtual characters is limited.

Technical expansion coverage
It is possible to create any imaginable device but the interaction with the device is limited. A controller like the one we use with REXplorer would be hard to test.

Conclusion
While it is not the same to interact with the "real" physical device and the desktop PC showing it, the need to create a high fidelity prototype to do middle fidelity testing is eliminated. Therefore the simulator can be useful to support the iterative design process. According to pervasive games the approach has its limits.

If a game like REXplorer should be tested overall the whole city would need to be simulated. But for middle fidelity prototypes it could be enough to only simulate some part of the city in a simplistic way. In chapter 4 ee will try to simulate the environment even simpler only including few static pictures.

2.1.4 Tatus

Tatus

The Tatus system [O'Neill et al., 2005] is similar to Ubiwise in the sense that they both use game engines to simulate a 3D environment. Other than Ubiwise, Tatus seperates between a system under test and the 3D engine. This has the advantage that the engine code has not to be changed to setup an usability study.

A second difference is that the Tatus system does not simulate the input device virtually but connects a physical device to the system under test. A high level overview of the system can be seen in Figure 2.4.

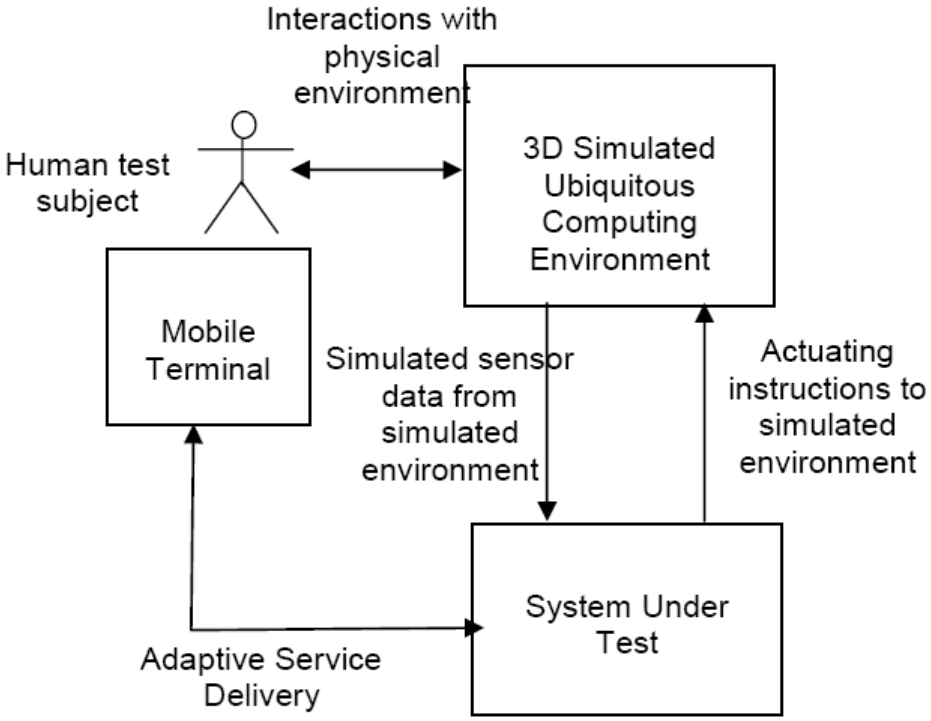

Figure 2.4: Tatus system architecture; With extra software the physical device can be used and is not only supported virtually

Judging this approach we have to consider the virtual and the real side of the system.

Spatial expansion coverage
Like the Ubiwise system the virtual spatial expansion is es-

sentially unlimited but there is a tradeoff between amortization of the development cost and the size of the map. The real side can be neglected as the system cannot be moved.

Temporal expansion coverage
This expansion is identical to the Ubiwise section. The server showing the 3D environment can run a long time but the user has to stay at the system which limits the expansion.

Social expansion coverage
Working with the 3D system there is no interaction in the real world. But it is impossible to include Bots in the system which can be scripted. As a script cannot substitute a real person's behavior the social expansion is limited but exists.

Technical expansion coverage
The system requires a physical device so it is not as easy to create new version as it would be with the Ubiwise system. The advantage is that changes in the interaction with the system do not mean editing game code.

Conclusion
Tatus suffers from some flaws similar to Ubiwise like the restricted possibility to create great maps and be cheap at the same time.

On the one hand, it offers physical interaction with the system via a device. On the other hand, for such situations normally a high fidelity prototype is created. That means that Tatus can only be brought into evaluation at the end of the development and it is questionable if it offers more data with less effort than a field study.

2.1.5 CrossWeaver

CrossWeaver

Even if REXplorer was already in the beta phase at the time this review was conducted we add some low fidelity methods for the sake of completeness. In [Sinha and Landay, 2003] a tool for rapid prototyping, conducting tests and analyzing the test session is given. To cover that it uses three different modes for *design, test and analysis*.

CrossWeaver was planned for non-programmers. In the design mode it offers a paper prototype like interface but it offers additional support for the definition of interaction, like shown in Figure refcrossweaver.

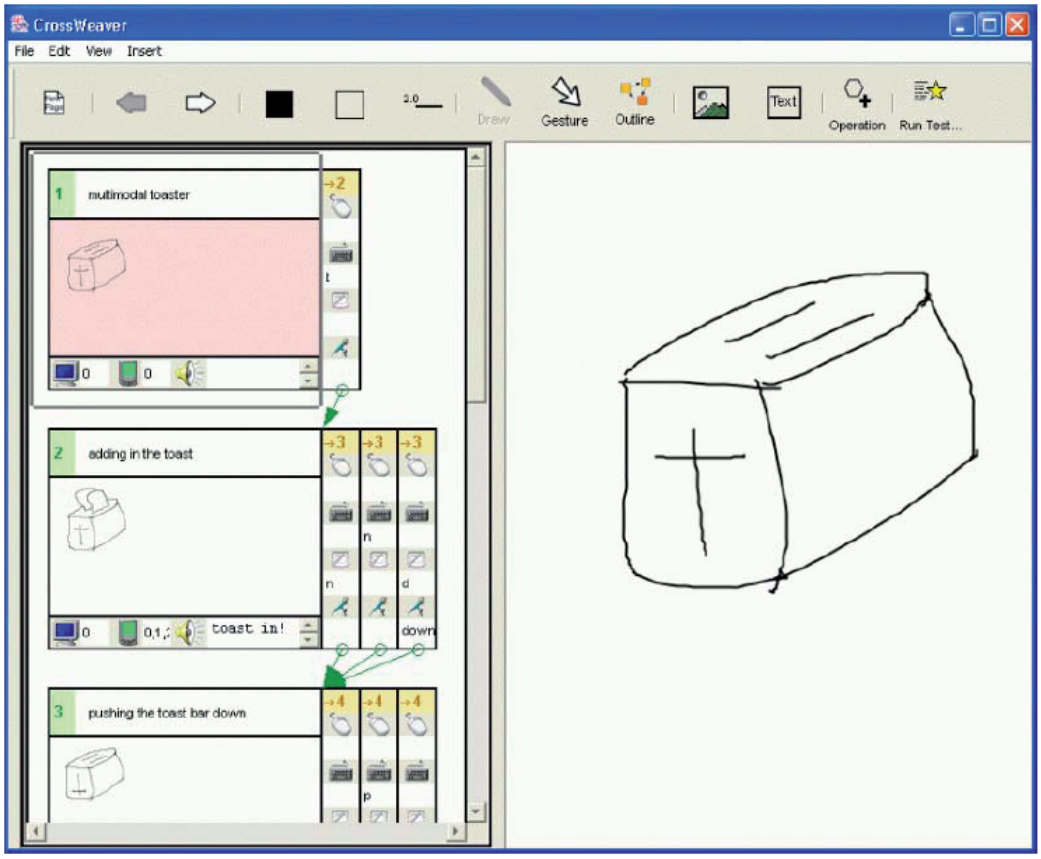

Figure 2.5: CrossWeaver in design mode; Interactions are defined on the left side

Following the defined interaction the screens of the storyboards are displayed on the interacting device. The system can be analyzed like that:

Spatial expansion coverage
Spatial expansions is neglected.

Temporal expansion coverage
There is no temporal expansion coverage.

Social expansion coverage
The social expansion is unattended.

Technical expansion coverage
CrossWeaver focus lies on the support of rapid prototypes so this area is covered.

Conclusion
CrossWeaver looks in particular promising for early prototypes. It is questionable if the provided interaction scales to pervasive games as there are no animations and state machines provided.

2.1.6 Paper prototyping

Paper prototyping

A short overview about the possibility of applying paper prototyping to ubiquitous computing applications can be found in [Liu and Khooshabeh, 2003].

Comparing a paper prototype with an interactive prototype utilizing the *Kitchen-Net* system reveals that the interactive prototype captured more results than the paper prototype and is easier to conduct. While the paper prototype was in use there always needed to be some people to simulate the feedback from the system.

Conclusion
Paper prototyping can be used to test early interaction ideas with the devices used in a pervasive game. This can be useful because paper prototypes are cheap and if the concept has potential flaws the are just thrown away. Besides that no expansion is supported.

2.1.7 A Hybrid Test and Simulation Environment

A Hybrid Test and Simulation Environment

The following system is not really important for pervasive games in particular but it is added to show other aspects of evaluation. One part of ubiquitous computing requires that devices work flawlessly with each other and that networks are available as communication mediums. [Morla and Davies, 2004] approaches this problem with a *hyprid test and simulation environment*.

Their goal *isn't improving the application but the test environment and how we can use it as part of the test and evaluation process for this and other, similar applications*. Therefore Morla et al. integrate applications into a test environment. The application only implements a web service interface to communicate with the rest of the system. Figure ? shows the standard setup.

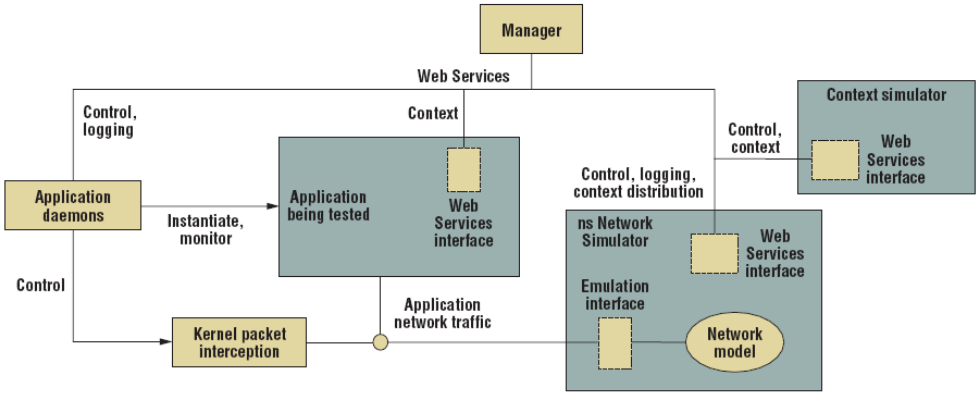

Figure 2.6: Setup of the evaluation system; The application code does not need to be changed after the web service interface was implemented

Spatial expansion coverage
As the focus lies on testing network-related issues the spatial component is neglected.

Temporal expansion coverage
The simulation has the potential to run for a long time while the log-data is collected automatically. Therefore the temporal expansion is supported.

Social expansion coverage
Users are simulated with scripts. So potential social problems are not covered.

Technical expansion coverage
The interaction between human and system is neglected but testing the interoperability of devices is supported.

Conclusion
Even if this approach is not important for this particular

diploma thesis it shows a different aspect of evaluation. It is promising because of its capability to discover network - related and interoperability issues. The lack of of methods to discover usability problems with system interaction makes it inadequate for evaluating pervasive games but it could be useful as extension for other tools.

2.1.8 Flying Emulator

Flying Emulator

Another problem ubiquitous applications have to deal with is not only the lack of networks but also the often changing networks. When a device is moved physically it may have to connect to a new one and use available services. A simulator for such situations called *Flying Emulator* can be found in [Satoh].

Physical devices are designed as mobile agents, compositions of software and data that make it possible to move autonomously from one system to another. The execution can be continued on the new system. This is displayed in the following figure:

Spatial expansion coverage
Through the mobile agents an spatially expansion is supported.

Temporal expansion coverage
Letting the system run over time does not offer new insights. So there is no special support for the temporal dimension.

Social expansion coverage
The system runs without users.

Technical expansion coverage
As there are no controllers the system does not cover the expansion.

Conclusion
Altogether as REXplorer is not based on different networks the prototype is not applicable to our approach. But then it

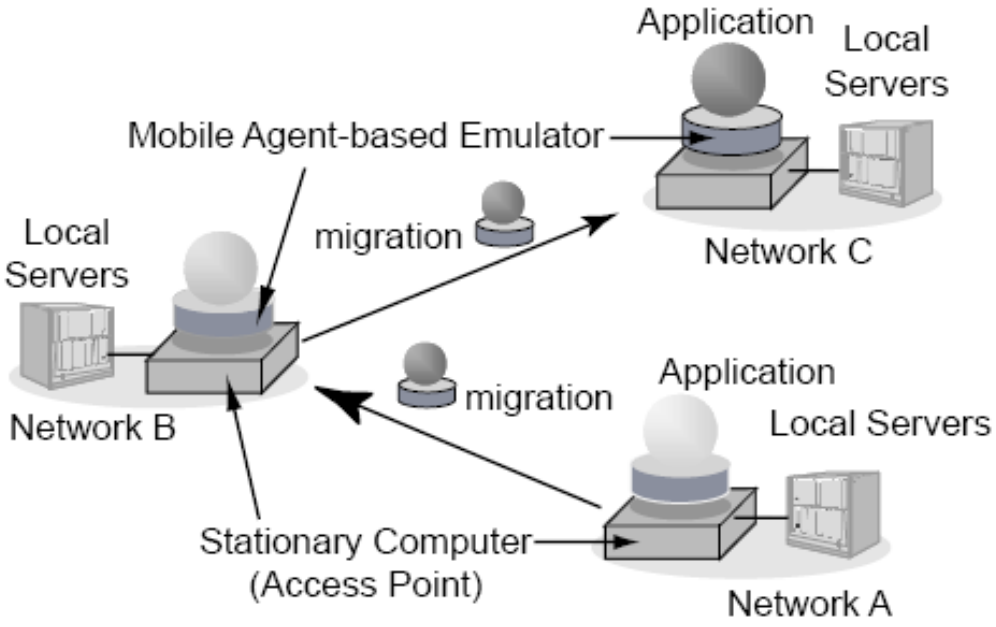

Figure 2.7: Setup for emulating logical mobility

could be useful for other games like Feeding Yoshi, which depend on changing networks.

2.2 Evaluation of pervasive games

The last section showed general according evaluation of ubiquitous computing applications. All of them had their short comings according to the expansions required to test for pervasive games. Therefore in this section we take a look at other projects and pervasive games to see how they evaluated the system.

2.2.1 General literature

Traditional methods used withing the IPerG

The Integrated Project on Pervasive Gaming (IPerG: IPerG[2]) compiled a detailed overview of traditional designing and evaluation techniques [Benford and Capra, 2006]. They show how they can be applied to pervasive gaming. Available design techniques are:

- participatory design
- scenario based design
- ethnographic field studies of current games
- cultural probes
- game design patterns
- game space and artifacts
- player game presence
- public performance as a research method
- ethical aspects

These are complemented with the evaluation techniques below:

- cognitive walkthrough
- questionnaires
- ethnography of trials with prototypes
- laboratory experiments
- critical review

The analyzed methods are used and put to practice within the IPerG. However we believe that pervasive games can be evaluated with traditional methods but it is necessary to combine the different methods to get useful results.

[2] http://iperg.sics.se/

2.2.2 Epidemic Menace

Evaluation of Epidemic Menace

The crossmedia game Epidemic Menace was already introduced in chapter 1. Here we want to explain how the game was tested. The goal was to evaluate the game concept and story, the game play across media, and the role of the devices [Lindt et al., 2007].

The setup for evaluation was composed of the combination of player feedback, questionnaires and observations by four observers according to different interactions the user could perceive:

- player-environment
- player-devices
- player-to-player
- player-gamemaster

Spatial expansion coverage
As observers can essentially follow the users everywhere the spatial expansion is mostly covered. This can be inappropriate if the players use the games in their daily life like in Feeding Yoshi. Questionnaires and player feedback can be collected everywhere.

Temporal expansion coverage
This is connected to the spatial experience: If players stay within a given range for the game time it should be possible to observe them for days. The user test of Epidemic Menace took two days and was evaluated this way.

Social expansion coverage
The observers had guidelines to cover points of the social expansion like taking notes about player-environment and player-to-player. A well designed questionnaire can also give information about how the players interacted with the world without the magic circle.

Technical expansion coverage
This is covered in the observation point player-devices.

Conclusion
The Epidemic Menace team could draw some interesting results from their evaluation. We believe that only taking notes has some flaws like already interpreting the event during the play session. This problem was attacked by collecting user feedback and utilizing questionnaires.

While video recordings, for example, may bring more detailed results, it could be problematic to analyze video sessions of two days. So note taking may be the better way in that case.

2.2.3 Can You See Me Now?

Evaluation of Can You See Me Now?

In [Anastasi et al.] the evaluation of Can You See Me Now? (CYSMN), which was already introduced in chapter 1, is presented. Collecting data was based on several sources of information. Starting with collecting offline feedback from players via email and their website, the also did debriefing meetings with the team after the session, ethnographic observations based on video and field notes and instrumented system logs plus text messages.

With this data they derived issues regarding gameplay and orchestration of the game. The following judgment will show if this complete approach can cover the test of all expansions for pervasive games.

Spatial expansion coverage
The ethnographic approach utilizes video and note taking. That makes it possible to follow players during an evaluation session. Therefore the spatial expansion is covered.

Temporal expansion coverage
In a longer session it would not be possible to track users with video recordings the whole time. But then using automatic tracking methods like log file and text message analysis, as well as gathering offline feedback, enables the approach to expand the observation and so the temporal expansion is covered, too.

Social expansion coverage

Video recording, note taking and offline feedback cover this expansion.

Technical expansion coverage
This can be covered by automatic tracking as well as recording, field notes and questionnaire. So it is covered.

Conclusion
The applied observation methods were very complete. If an appropriate analysis method is available for all the aspects the system should offer a lot of high quality data. The only questionable part is which kind of fidelity can be tested with it. If the system is not in a high fidelity status the effort may not pay off.

2.2.4 Bill

Evaluation of Bill

In [Chalmers et al., 2005] the pervasive game Bill, which is similar to Feeding Yoshi introduce in chapter 1: Players collect virtual *coins* from outside the wireless network, and then runs back into network range to *upload* the coins to gain points.

QCCI

The idea for evaluation is to automatically "create a a coherent and synchronized visualization or *replay* of the game." For that information from several sources like system logs and videos from several cameras are combined with a tool called *QCCI*.

This tool is not only able to enhance the collected video data with timestamps and the actual location, but also with GPS signals and audio notes on the fly. The timestamp signal can be communicated over the network so all observers use the same.

The result of the data collection makes a unique replay possible that can be played back and forth for further analysis. Videos can be imported and played side by side with the replay.

This nice idea can be used to cover the expansions in the following way:

Spatial expansion coverage
As we discovered when testing REXplorer GPS can suffer from inaccuracy if it used in environments with high buildings and small streets. Nonetheless utilizing logging through the device and collecting video data can help to test about every spatial expansion as the collecting devices are easy to carry around.

Temporal expansion coverage
While video recording can theoretically be done an unlimited time it can be hard to record the data while participants are living their normal life. System logs can be collected without interweaving with normal life.

Social expansion coverage
The system logs cannot be helpful to find flaws in interaction with bystanders, other people, ... But then video recordings can be used to play that back.

Technical expansion coverage
Using video and a possible extension of QCCI could give needed information about the technical expansion as the user's interaction can be replayed later.

Conclusion
The replay feature of QCCI is unique and very useful. The video recordings are enhanced in a useful way and the underlying system should be easily expendable. For example a benefit for evaluating REXplorer could be to log how many times a user tried to make a specific gesture.

2.3 Physiological measurements

Physiological measurements

A new way to evaluate collaborative entertainment technology systems with physiological measurements is described in [Mandryk and Inkpen, 2004]. The idea is to collect different body responses and compare them with subjective responses, different situations, and subjective reports by the user.

Traditional evaluation methods only test productivity and

performance based on cognitive science while this approach enables to test enjoyment and interaction.

Spatial expansion coverage
The bottleneck for the spatial expansion coverage is the size of the equipment needed for measurements. If the tools are not robust and small enough it may permit moving around and traveling with it.

Temporal expansion coverage
The data gathering can be used for continuous collection.

Social expansion coverage
Experiments explained in the paper included measurements of reactions while playing against the computer and a friend. The results were different and so it is possible to measure the expansion. Including the video recording situations can be identified in which the expansion takes place.

Technical expansion coverage
Experiments could be conducted to measure how the physical reaction to several devices differ.

Conclusion
The physical measurements offer new ideas for evaluation. The results are not restricted to traditional aspects of usability but can also make it possible to evaluate emotional reactions that can help to draw conclusions about playability. But then the equipment may not be robust and light enough to carry around. So the testing of games like REXplorer, which is played in a whole city, may suffer from the restricted movement.

Chapter 3

Evaluation in the field

"Imagine if every Thursday your shoes exploded if you tied them the usual way. This happens to us all the time with computers, and nobody thinks of complaining."

—Jef Raskin

The last chapter gave an overview about how other groups evaluated pervasive games. The dominant way seems to be using field studies which makes sense because as an emergent field only few information is available about how players behave in the wild. Field studies capture that knowledge by observing participants in their natural environment.

Starting this thesis REXplorer was available as high fidelity prototype which needed testing. A field study should reveal existing problems and new insights about pervasive games. This situation made it possible to choose, apply, and show capture methods in practice.

3.1 Choosing an evaluation method

This section will explain how we combined traditional evaluation methods to test and analyze REXplorer.

3.1.1 Requirements

Requirements

Field studies take place in the environment the system will be used in after release. So their advantage is that they offer realistic results. Besides that it requires time and effort to conduct such a test.

Especially location based pervasive games like REXplorer imply that the user study must be conducted in a foreign city which brings up additional problems. A subset would be:

- Cost for traveling
- finding participants
- bringing all the needed equipment

In this sense field studies are mostly appropriate for high fidelity prototypes. The system should be tested this way if further investigation in the laboratory cannot bring new insights. Chapter 4 will discuss this topic in detail.

If a field study is conducted the results should amortize the effort. Therefore an adequately chosen field study method should offer:

- effort in balance with the state of the prototype
- easy capturing
- easy recreation of data
- getting desired information

Below the points are discussed in detail.

Effort in balance with the state of the prototype This point simply describes that the granularity of the test and the fidelity of the prototype should match. On the one hand, for example, it makes no sense to develop a detailed test environment if only a paper prototype should be tested which

will probably be thrown away anyways. On the other hand, there is a point when cognitive walkthroughs bring no new insights and users should be integrated.

Easy capturing
Pervasive games are often played outside. Therefore a capture method is required that can easily be carried around.

Easy recreation of data
The REXplorer play session took about an hour plus another 30 minutes for the interview. If this data is collected through taking notes much important information may be missed. The note taker has to write down the scene and focus on the note. So he will miss further interactions.

Another problem would be that these notes are not detailed transcriptions but will already be interpreted because of the lack of time. So the results may be biased and an easy way to replay the results later should be used.

Getting desired information
During an user study a lot of information can be collected. Huge amounts of data may accumulate and it can be hard to find the interesting parts. Therefore we need a method to focus on subsets of the collection. The next section describes how we decided which method to use.

3.1.2 Decision

Decision

Analyzing the requirements it was easy to conclude that one method alone cannot bring up the information we needed. First we wanted the results to be reproduceable. Therefore we decided to record a product-interactive focus group discussion (PIFG) [Lee and Bhatkhande, 2004]. PIFGs mean that the participants get a certain set of tasks which they have to fulfill and afterwards a group discussion including a moderator and a specific topic is conducted.

In our case the set of tasks was playing the game which we also recorded to do a protocol analysis [DIX, 2003] of the play sessions. The specific topic we discussed were the ex-

pansions of the game according to the magic circle. Therefore and for keeping a general structure an interview guideline was handed out to each interviewer; it can be found in Appendix A.

The guideline covers not only the expansions of the magic circle but also a technical view as we wanted to know about the interaction with the device. The temporal questions were not targeted at the traditional expansion. While Montola focuses on an expanding time parameter REXplorer suffers from limited sessions and we wanted to know about how people react on this.

The idea was that the participants will produce the most information if they discuss with each other. Therefore we always tried to let two groups play in parallel and conducted the interview with all players at the same time. The demand was to let people discuss and only bring up questions if the interview got stuck or answers were unclear.

During the game normally locations are indicated utilizing GPS. As this was not ready for the user test we used a Wizard of Oz technique to simulate it. That means we used a limited functionality prototype and provided the missing functionality via human intervention [DIX, 2003]. In our case a Nokia 770 was used to send location data to the device. So we had groups of two, one to record the session and one to simulate GPS.

The starting point of the game after release will be the Salzstadl. This was not usable at the time of the test.

The test was conducted with the following procedure:

1. Showing an introduction movie to emphasize the setting
2. Walking with the participants to the starting point Salzstadl. It was prohibited to talk about the game on the way.
3. Following the users and recording the interaction with the device

4. After an hour: Stopping the session and walking back to the place we started

The play sessions and the interview delivered about one and a half hour of recorded video per group. With eight groups with mostly two people per group this made approximately 20 hours of video. This required a structured method to get the results we were looking for.

So we defined six research questions to collect the data:

- What influences/affects users enjoyment of the experience?
- How do users navigate unfamiliar space? How does the system support navigation (or not support navigation)?
- How do users feel about the experience of interacting with the system in context?
- How do users perceive the content of the game?
- Do players follow the activities intended by the game's design?
- What is the stageability of the location - how does the current location affect the player experience?

The questions intentionally were left very general to not limit the results coming up. Regarding to the questions we watched the movies from the tests in groups of two and transcribed what people said about these topics. The notes should include the original text of the participants to be less interpretive. We also added some extra text including group and session type for later reference.

For the analysis we searched for a method that makes it possible to use the statements described above to derive high level findings about the game and the users' problems.

Therefore we needed a bottom-up approach like an affinity analysis described in [Beyer, 1997] which is based on

grounded theory. This methods helps to extract common patterns from the collected data which show issues regarding the interaction with and the key quality requirements of the system.

The methodology behind an affinity analysis is to create a diagram which spatial order reflects the logical order of the statements. Therefore all user statements were printed on Post-It notes which results in approximately 600 notes in our case. The first step is to put them all on a wall side by side. Afterwards in a group of two we moved notes that deal with similar topics near each other.

The resulting diagram can be seen in 3.1.

Figure 3.1: Complete affinity diagram, containing approximately 600 notes, after clusters were build

The different clusters in the affinity diagram can be used to derive design ideas. It is important to note that these findings are not proven statements. The interpretative character of the ordering requires further testing of these statements. That means the collected information can be used to derive hypothesis for experiments but cannot be treated like design proofs.

The next section will show our findings from the analysis

in detail.

3.2 Findings

Findings

The affinity analysis gave some surprising results. These are represented in the following subsections with the translated quotes. The original quotes in German can be found as footnotes.

3.2.1 Fact vs. fiction

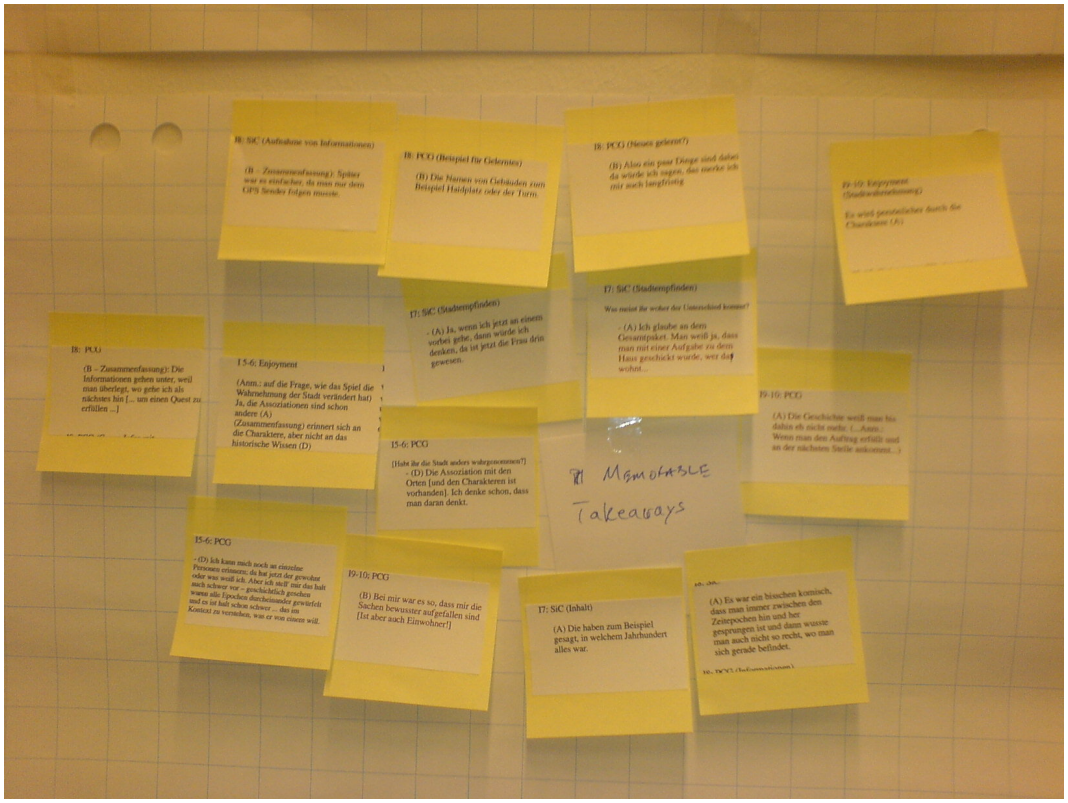

Figure 3.2: Statements about "memorable takeaways"

Fact vs. fiction

Before REXplorer was designed one question was how much historical content should be in the game. On the one hand, the system should be engaging and enjoyable. On the other, hand REXplorer is an edutainment system that

should offer historical facts which are not known to be very entertaining.

The design decision to prefer engagement over facts was made. In our test we noticed that our players liked that. For example Nancy said: "... the characters are absolutely nice"[1] or Calvin stated: "I loved the atmosphere. The other way would have been that special tourism style again"[2].

Besides telling their own story our characters also reported historical facts. Interestingly our participants did not memorize the historical dates but associated characters and their relation to places. Irene told us: "I think i can remember the associations between the places and the characters"[3] or "I can still remember some of the characters"[4]'Ich kann mich noch an einzelne Personen erinnern".

Some participants still complained that the information was too superficial. As a result an info gesture was added that brings up deeper information about places. The gesture was designed as an I so it could be remembered easily.

3.2.2 Quests

Quests

Playing REXplorer participants do not have to follow a default route but can decide themselves where to go and which quests to accept. From our diagram we could derive that our players liked that. Ebony put it that way: "It is great that i can decide myself how to order the game... i can stop it whenever i want"[5]. Danny imagined that he could do daily things between playing like "buying cigarettes or stop walking"[6]. Adriane even bought an ice cream and kept on playing without any problems.

[1] "... die Figuren sind ja total nett, wie sie das erzählen"

[2] "Das war von der Atmosphäre her das Schönste... das andere war eher so Tourismusinformationsstil"

[3] "Die Assoziation mit den Orten [und den Charakteren]. Ich denke schon, dass man daran denkt"

[4] '

[5] "Es ist toll, dass man sich den Ablauf selber einteilen kann... ich kann es beliebig unterbrechen"

[6] "Zigarettenkaufen oder mal stehen bleiben"

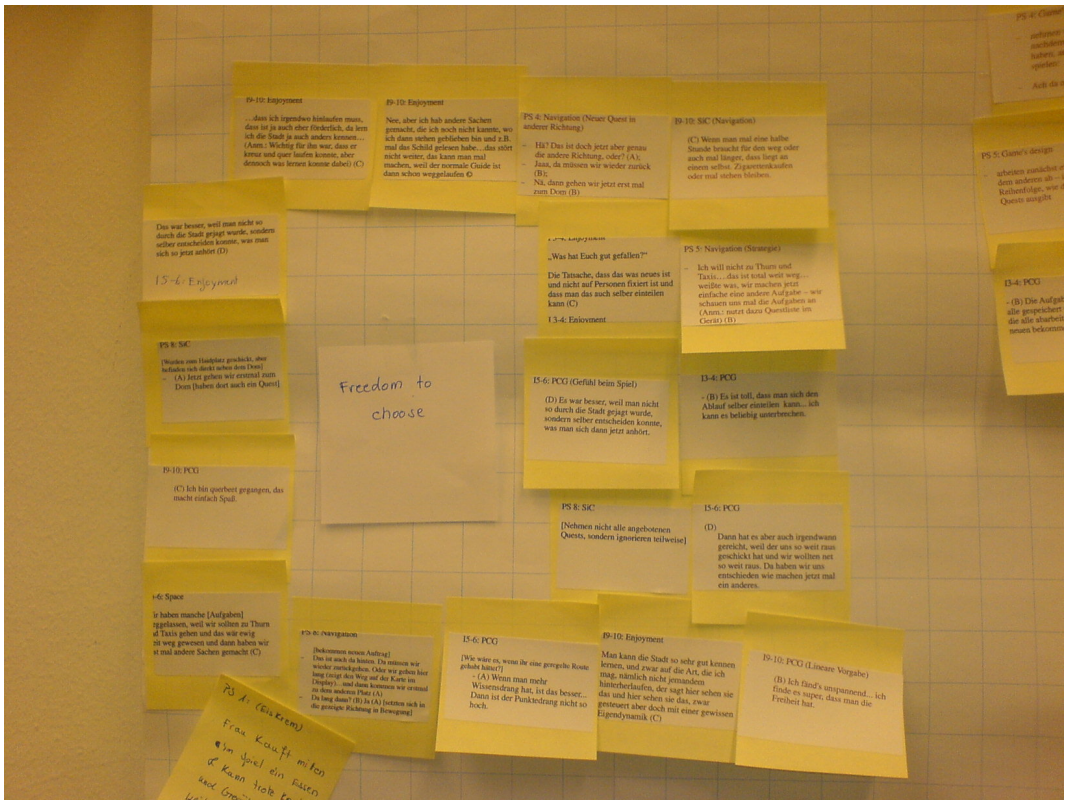

Figure 3.3: Statements about the "Freedom to choose"

Having the freedom to choose many players started to form strategies to cope with the upcoming characters. For instance Hannah stated: "When we finished the first quest, we show walk to the tower. On our way we met another character and do his quest first"[7]. Fae and Gabi even optimized their strategy to play very efficiently and always finished quests first which were spatially close to each other.

When we designed the game we had the feeling that people might get lost when they accept too many quests. To avoid that we limited the number of quests that can be accepted in parallel to three. Some of our players liked the idea. Irene: "But i like that... it would have been chaotic without"[8].

[7] "Als wir die erste [Aufgabe] erledigen wollten, sind wir am Turm vorbei gekommen. Wir haben dann entschieden erstmal unsere Aufgaben fertig zu machen, anstatt eine neue aufzunehmen"

[8] "Das fand ich aber auch gut... Sonst wäre es sehr chaotisch"

But the diagram also revealed that players were not satisfied with that constraint. When they entered a new hotzone our participants wanted to listen to the new characters. Some adapted to the situation by changing their strategy. Ebony: "Not all quests are saved and so we wanted to fulfill ours so we can receive new ones"[9].

Our goal was to combine freedom with not being confused and so we asked our players if a delete function would be a good solution. People liked that solution. For example Adriana said: "That would be a good idea. I would have deleted all of my quests"[10]. This feature was added with the next release.

One problem that came up often during the test was what we call the *Ping-Pong-Effect*. People were send from a character A to a character B and then B immediately send them back to character A with the next quest. Irene stated: "It was bad to be forced to walk back the same street we came from—we did not like that because we have been there before"[11].

Interestingly the effect was low when people were send to the same character later in the game and had some other quests between it. For example we asked Irene again how it would have been if she would have to solve some quests before coming back to the character: "That would have been okay"[12].

Our first strategy to solve this problem was to add a quest history which keeps track of the accepted quests and intelligently reduces the Ping-Ping-Effect. The history should be considered when a quest is chosen and it would not be picked randomly like it does right now. We noticed that this was obsolete after we added the delete functionality.

[9] "Die Aufgaben werden ja nicht alle gespeichert und daher wollten wir die alle abarbeiten, damit wir die Neuen bekommen"

[10] "Das wäre gut gewesen. Ich hätte komplett alle meine drei Aufgaben wieder gelöscht"

[11] "Also es war schon blöd von einer Strasse wieder zurück in die gleiche Strasse zu laufen—das hat uns also schon gestört... weil wir da direkt davor schon mal gewesen sind"

[12] "Das wäre in Ordnung gewesen"

Some players stated that the game time of an hour was too short. When they would play again they would not want to do all the quests they had done before. To attack that problem a saving functionality was added. When people play again later the current state of the game is rebuild.

3.2.3 Gestures

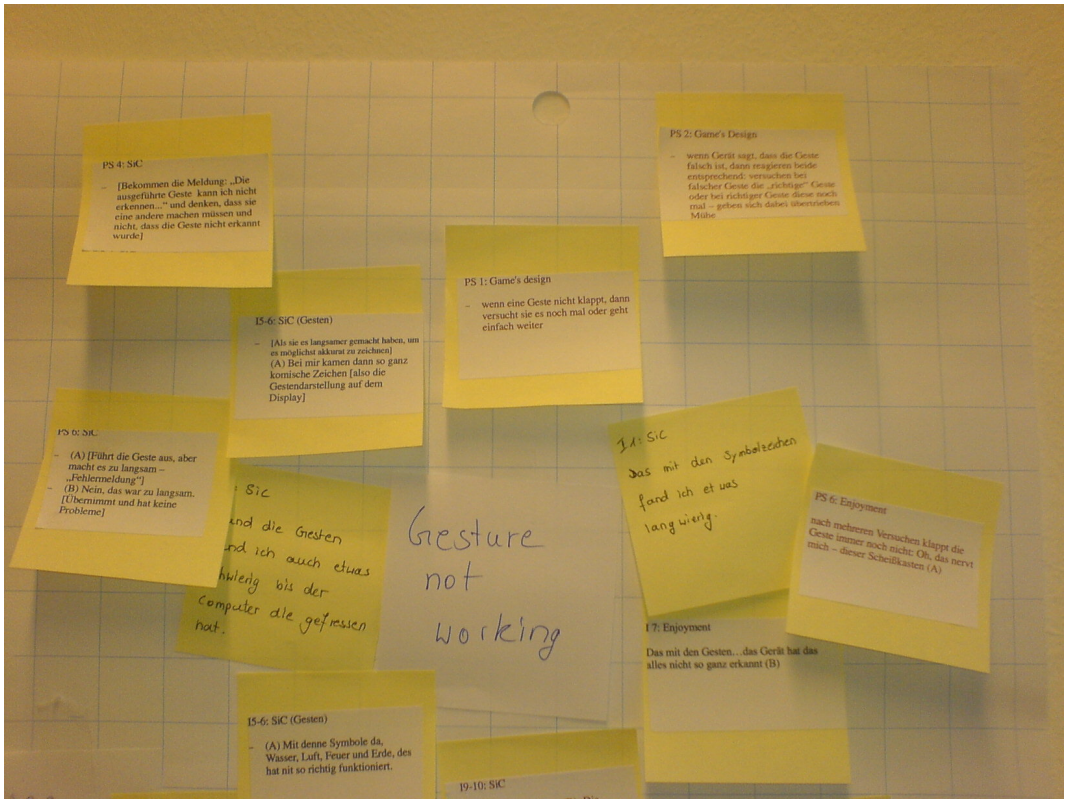

Figure 3.4: Statements about "Gestures not working"

Gestures

The Regensburg experiment was our first in field study. So the software was still in prototype state. One thing we noticed which did not work well for our participants were the gestures. The display on the device was very sensitive and if people did the gestures slowly the display showed moves in the wrong direction.

The problem is that when the gestures did not work the participants tried to make them as accurate as possible.

That resulted in an even worse feedback screen so people stopped doing the gesture and so the system could not recognize it. Lacy put it that way: "The gestures were not recognized"[13]. On the other hand, some players were surprised how tolerant the system was when they did their gesture. Aaron: "I like that the system accepted a round C—pretty tolerant i would say"[14].

From that we conclude that if the players had finished their gesture most of the time it had worked. To improve that we configured the gesture visualization so that it is not that sensitive anymore and people will hopefully finish their started gesture.

Interestingly some participants liked the challenge, that the non-working gestures offered, but others got frustrated if the system did not recognize their gestures even after many tries. Danny said: "I did a gesture five times and then thought: If it does not work now, you will just leave it alone"[15].

To cope with the situation some participants asked for buttons to press instead of gestures but others said that this would get rid of some important part of the game and shorten their experience. That got approved because after people finally figured out how to do the gestures properly they got happy about them. Irene says: "Yeah"[16] after doing the gesture several times before it got accepted.

First the system was redesigned to give every choice. First one has to try a gesture. If that did not work an rotating arrow appeared with which one can pick a symbol or go back to the gesture recognition.

After further testing we noticed that this version added a mode to the device. People pressed the button to do an-

[13] "Das mit den Gesten... das Gerät hat das alles nicht so ganz erkannt"

[14] "Was ich ganz gut finde, ist dass selbst wenn man ein rundes C macht, das Gerät selbst das noch erkennen würde—recht große Toleranz auf jeden Fall"

[15] "Ich habe so eine [Geste] fünf Mal gemacht und dann habe ich gedacht, wenn Du das jetzt nochmal machst und es nicht klappt, dann hast Du keinen Bock mehr, dann läßt Du es bleiben"

[16] "Na bravo... Yeah!"

other gesture and did not recognize the picking which confused our users. So instead the gesture recognition was improved to allow people to fulfill the gestures.

3.2.4 Focus of attention

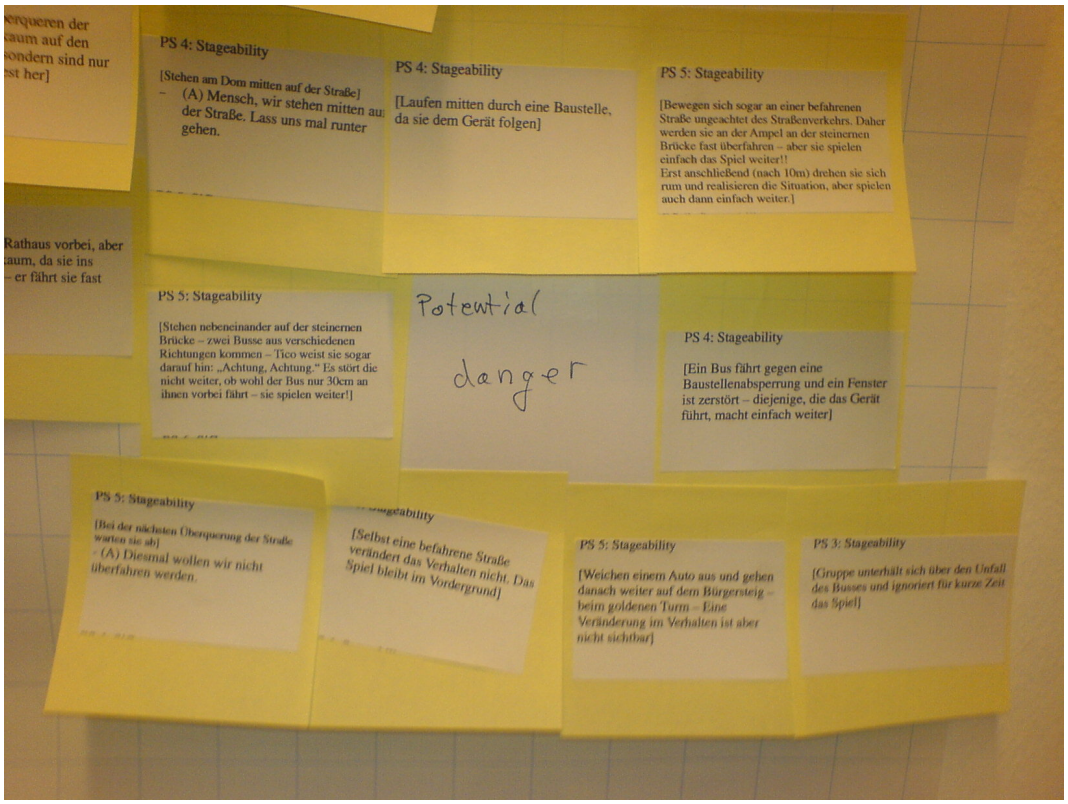

Figure 3.5: Statements about "Potential dangers"

Focus of attention

The REXplorer device was designed in a way the let people focus on the environment and not on the display. Most of the information was included as audio. The visual information the display provided was very low. It included low resolution pictures in black and white.

This worked out well for some players like Jackleen and Aaron who looked around in the environment while the characters were speaking. On the other hand, Ebony stated: "It was bad that one looked too much at the display and

listend to the little man talking"[17].

Even when the display contained few informations the players were very immersed in the game. For example Hannah and Irene got nearly hit by a car when they crossed a street while the traffic lights were red. They got scared for a few seconds but then kept playing.

To cope with the situation an explicit statement to be careful was added to the introduction movie and the instructions on the paper map.

During Fae's and Gabi's play session a bus hit a construction site a few meters away from them. They only looked up for a few seconds but then they kept playing like nothing had happened.

This shows that a game like REXplorer brings tension to the real world by attracting the attention not only of the players but also of bystanders.

In a city like Regensburg our players were surrounded by noise from cars, people and construction sites. Many dealt with these situations by using the *ear strategy*. They moved the device nearer to their ear. This worked for some but some others still were not able to hear the audio. Hannah stated: "Once a bus passed by, i did not hear it anymore, and i could not turn in louder"[18].

Daniela suggested: "It would be more comfortable to have ear phones"[19] Lacy also stated that the headphones were good because she would not disturb bystanders. This led to discussions in our group interviews. For example Jackleen said: "I would not like to wear ear phones. It would disturb me"[20] or Gabi stated: "I would feel silly walking around with ear phones. I think it would not be without danger,

[17] "Es war schade, dass man sehr viel auf das Display geschaut hat und was das Männchen da erzählt hat"

[18] "Bei uns ist zum Beispiel mal ein Bus vorbei gefahren und es ging dann nicht mehr lauter zu stellen"

[19] "Also mir wäre es angenehmer, wenn ich dann Kopfhörer hätte".

[20] "Also mich persönlich würde es stören ständig mit den Kopfhörern, das mag ich nicht so"

too."[21].

Actually the device has a button to repeat the character's text so that in extreme situations the text can be heard again when it is over. It seems like this function has to be made clearer in the introduction movie and the instructions.

3.2.5 Collaboration

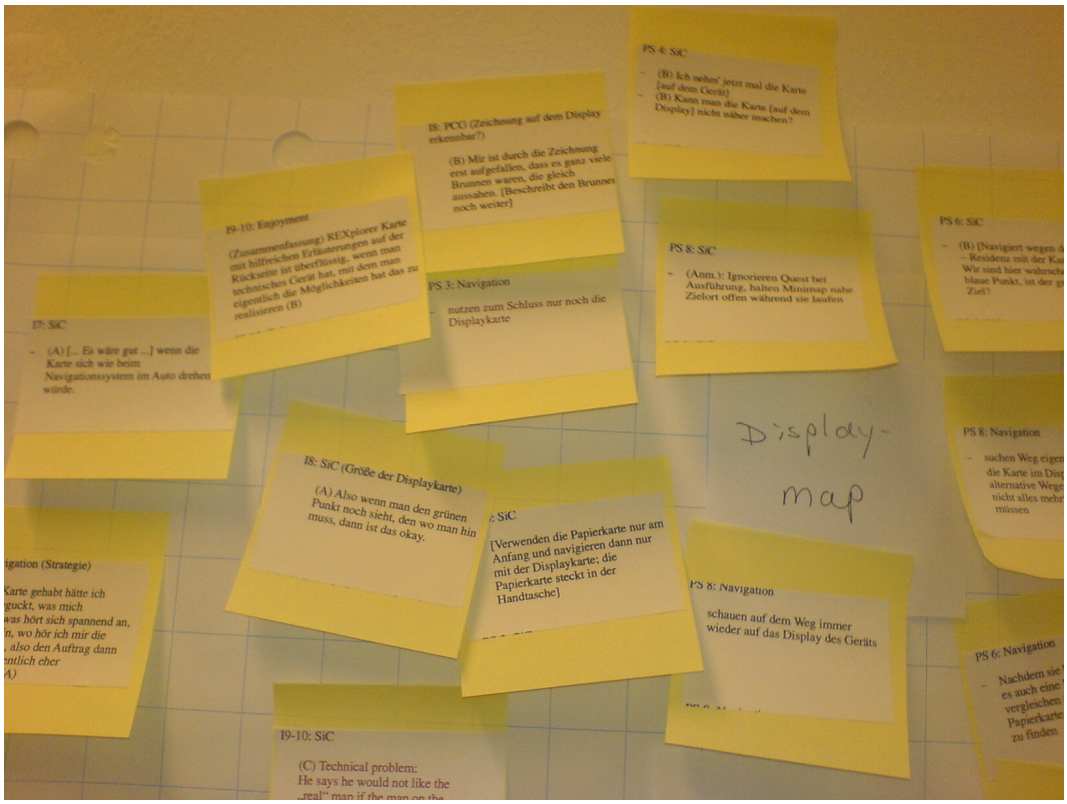

Figure 3.6: Statements about "Display map"

Collaboration

REXplorer was designed for more than one player. Many of our players discussed which gesture to do next. If the gesture recognition does not work they gave the device to the second player and let him try it out frequently. During the game they debated which route to take to fulfill the next quest.

[21] "Da würde ich mir genau so blöd vorkommen, wenn ich da mit Kopfhörern ständ. Außerdem wäre das sicherlich nicht ungefährlich!"

One element of the game design that supported the collaboration well was handing out a map beside the device. For instance Nancy said: "A map is good so that everyone has something to carry and feels integrated"[22]. Gabi added: "We have to go there"[23] and shows Fae where to go on the paper map. Besides discussing the route the map was used by Jackleen to show functions and gestures to the other player Aaron.

The device also contained a digital map. The layout of the maps was designed in a way that both maps could be accessed together. Aaron stated: "We used the map quite often to see where we have to go. Sometimes we even used the map on the device because of the points telling where to go"[24]. Fae and Gabi were not able to find their goal. To be sure which character to talk to they compared the display drawing and the drawing on the paper map.

Some other players decided to only use the device or the paper map. Interestingly some players only used the map on the device even if they said that they think it is too small. Gabi said: "Can't the map be zoomed?"[25]. Danny only used the device and even stated that the paper map is unnecessary because the device provides everything he needs.

Another element designed for collaboration was a low information screen described above. It is easier to connect to one another when people move their focus from the device to the environment.

Through the map the tasks were divided in the group. While one person used the map and searched for the way the other uses the device and collects the points. That seems to be important to develop a collaborative experience and integrates the second player without a device into the game.

[22] "Eine Karte ist gut, wenn man zu zweit ist, da jeder etwas in der Hand hat und sich jeder beteiligt fühlt und man sich austauschen kann"

[23] "Da müssen wir da hin"

[24] "Wir haben ziemlich oft den Stadtplan, also die Karte verwendet, teilweise auch die im Gerät drin, weil es da die Punkteanzeige gab, wo man hingehen muss"

[25] "Kann man die Karte [auf dem Display] nicht näher machen?"

3.3 Summary

Summary

Analyzing the user test gave some interesting insights into the field of pervasive games in general and into REXplorer in particular. Said that we have to state that it is questionable if the approach we took would be general enough to test different systems.

The border for video recording could be the degree of the spatial and temporal expansion. For example the play test of the game Feeding Yoshi, which we analyzed in chapter 1, took place for about a week. People went to work and only played the game when they had time for it. Such situations cannot be covered with video recording and a different approach has to be taken.

Another bottleneck is the amount of data. Creating the affinity diagram may not scale to a certain amount of data. To do an overall analysis of a playtest like the one conducted for Epidemic Menace, which took two days, may not be possible. Analyzing all that video would take too long to amortize. But then, in such situation the video collection could be limited like only recording a few hours of the test.

We should mention that recording play sessions at all can be very obtrusive and may change the user's behavior. People may react different if they have false assumptions about what they expected to do. To limit the effect interaction between the conductor and the participant during the test was almost permitted - only in situations like a device crash the conductor interfered and helped.

In our own test we actually noticed that not the clustering of the Post It - notes is the time consuming part, but creating them. Watching all videos and transcribing interesting comments in a group of two took about a week for 20 hours of video. It is recommend to increase the number of analyzers and split the questions. We believe that it would be optimal to explore one question per person. The analysis should be conducted in a quiet room without disturbance.

Altogether the affinity analysis took a lot of effort but the

results paid off. The findings gave a lot of insights which can be used for creating hypothesis and conduct further experiments. Nonetheless this approach still requires to be in place to conduct the test. A different approach, trying to bring the evaluation to the laboratory, will be introduced in the next chapter.

Chapter 4

Bringing evaluation to the labortory

"Fast, Cheap, and Good Usability Methods: Yes, You Can Have It All."

—Jakob Nielsen's Alertbox

The evaluation of ubiquitous computing applications and therefore pervasive games traditionally takes place in the field. In chapter 3 we saw that these may provide valuable information. The problem is that field methods may not always be appropriate for the fidelity of the prototype because of costs and effort.

As an alternative we try to use traditional methods to bring the experience of playing in the field to the laboratory. For comparison we conduct two user studies. We start with testing REXplorer in the field and collect errors with a questionnaire. Then we continue with the development of a game simulator, based on traditional methods for evaluation. Reccurring errors are discovered with the same questions.

But before analyzing the differences between the methods we start with an overview of the advantages of laboratory studies.

4.1 Advantages of a laboratory study

Advantages of a laboratory study

There are several reasons why conducting a laboratory study has advantages over a field study. A subset would be:

Equipment is available
When we did our user tests in the field we had to bring all the equipment we probably would use. If we forgot something it was hard to find a substitution. In the laboratory everything is available.

Costs for traveling and similar can be minimized
For our field study we had to travel to Regensburg. That included booking a hotel room and buying a train ticket plus paying for food. That caused additional costs.

Possible experiments can be conducted as extraneous variables can be controlled
In a field study it is never possible to control all extraneous variables as traffic can pass by, the weather can change and other things can happen, which cannot be prohibited. In the laboratory it is easier to fulfill that.

Possible lack of features for prototypes are easier to catch
In [Kjeldskov et al., 2004a] described that as *lack of control undermined the extendibility of the field condition* which we already discussed in chapter 2.

Field studies are very valuable because they test the system in the user's environment. Especially for location-based pervasive games like REXplorer it may be hard to simulate this environment in the laboratory.

In [Kjeldskov et al., 2004a] such a simulator is presented. Kjeldskov compared the results from a laboratory and a field study of *MOBILEWARD*, a context-aware mobile system and got positive results. In his conclusions he offers four key findings:

- Little added value of taking the evaluation into a field condition

- Lack of control undermined the extendibility of the field condition
- Both the lab and the field revealed context-aware related problems
- The clip-on camera facilitated high-quality data collection of mobile use

Altogether the conclusion could be that a laboratory study is enough to evaluate a ubiquitous computing system, but Kjeldskov already explains in his own paper that the system may suffer from its context and may not be generalizable.

[Carter et al., To appear.] concludes that such laboratory studies can be used for testing issues of aesthetics and standard graphical interface interaction, as well as for comparing possible solutions.

Even if it is problematic laboratory studies are valuable to support the iterative design process. While the costs of a field study may not be appropriate for middle fidelity prototypes a simulator can bring good and cheap results in such situations.

Our contribution will be the comparison of the evaluation of REXplorer in the laboratory and in the field to expand that discussion to pervasive games. The goal is also to narrow the gap between low and high - fidelity prototypes in this area.

We will start with the description of a play session in Regensburg as reference for comparison.

4.2 A second play test in Regensburg

A second play test in Regensburg

As a second source for beta testing and collecting data we conducted a field study with 30 pupils. The results were captured with a questionnaire.

Before the test the system still suffered from problems with GPS in the city so for that study the hotzone recognition was changed to location picking via a scroll down menu.

The focus was qualitative data as we did not only want to know how many but also which errors came up. Nonetheless the questionnaire was extended with quantitative data for better comparison.

4.2.1 Questionnaire design

Questionnaire design

The questionnaire was targeted at the magic circle. We took the spatial and social part into account. During the beta phase we were not able to test the saving mechanism. Therefore the temporal expansion was neglected.

Besides these points we wanted to learn about traditional areas which include the game design and the interaction with the device. As a result we added technical and conceptual sections.

Below we show the translated questions. The original german questions can be found in Appendix B.

Spatial aspect

Spatial aspect

1. Have you been disturbed during the game? If yes, how?
2. Did you ever get lost? If yes, how did you find your way back?
3. How did you decide which route to take to the next quest?
4. Have you ever been in danger? If yes, please describe.
5. How important was the map to you?
6. Do you think that the game fits Regensburg? If yes, why? If no, why not?
7. How did the game change the city for you?

8. Which role did the city take for you?

Technical aspect

Technical aspect

1. Did you have problems using the device?
2. Have you ever had problems to understand the character's voices? Please describe the situation.
3. Has the volume of the device been adequate? If not, in which situation did problems occur?
4. How did you carry the device? Has this been comfortable for you?
5. Did communicating with gestures fit the game? What could be improved?
6. Was the introduction movie adequate for introducing the game?

Social aspect

Social aspect

1. How did you like playing in public?
2. How did you share tasks in the group?
3. Have you ever felt excluded from the game? Please describe these situations.
4. What do you think was your role in the game?
5. Have you felt like playing the game?
6. I felt the game was (risky - safe)

Conceptual aspect

Conceptual aspect

1. How much would you pay to play REXplorer?
2. After you have seen the introduction movie, what did you expect the game to be like?

3. How would you describe REXplorer to a friend in one sentence?
4. What do you think about the length of the game?
5. Which character did you like most?
6. Has there been content which felt incomplete?
7. Did you ever cheat? What was the reason?
8. Have you develop interest in Regensburg besides the game? Where would you look for information?
9. Was the goal of the game clear to you? What was the goal and could you reach the goal? If not, why?
10. How did you manage your tasks?
11. Did you know when to do a gesture?
12. Did you ever get bored? Please describe the situation.
13. I think the game was (boring - entertaining).
14. I think the game was (new - common).
15. I think the game was (natural - unnatural).
16. I think the game was (lowbrow - informative).
17. I think the game was (exhausting - refreshing).
18. I think the game was (informative - non informative).
19. I think the game was (amusing - absurd).

These questions cover the aspects which we found interesting in the user study. The results were compared with the results from the laboratory test and can be found in a compilation below.

4.3 Designing REXperimentator

This section will cover the iterative design and implementation of REXperimentator, a lab simulator for the REXplorer game.

4.3.1 First iteration

REXperimentator as desktop application

The first requirements we set for REXplorer were that it must be able to simulate the game logic and the gesture input. For that it had to show the map plus an indicator where the player currently is and an area to display characters and buildings.

We decided to use a split view with the map on the left and the interaction elements on the right:

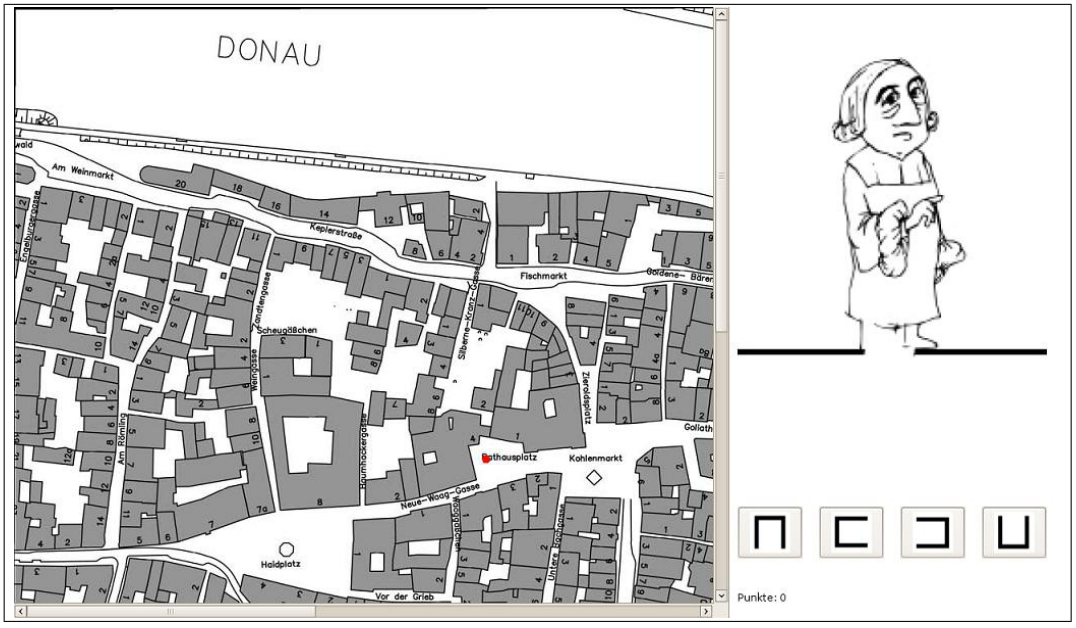

Figure 4.1: First version of Rexperimentator

This approach offered interesting ways to conduct experiments. Extraneous variables could be isolated to a high degree. Some of our initial ideas were to use points and audio as independent variables. It would have been easy to remove the score or exchange audio with displayed text.

However our own experience with the system and feedback from colleagues quickly revealed that the simulator does not capture the experience of the original system at all. So the results of possible experiments would not allow inference to REXplorer and decided to try a different approach which will be described in the next section.

4.3.2 Second iteration

Improving REXperimentator's design

As a traditional desktop computing application could not fulfill our needs according to the simulation of REXplorer we decided to reuse the existing system as much as possible.

Below is a list of REXplorer's expansion to the magic circle as described it in chapter 1.

Spatial expansion of REXplorer

Spatial expansion
REXplorer is played all over the city of Regensburg. The hotzones do not move but the players have to move from one hotzone to the next following the quests.

Temporal expansion of REXplorer

Temporal expansion
On the one hand, a game session takes about one and a half hour but on the other the current game status is saved so people can come back later and play again. As a souvenir a blog is created online which shows the route and photos taken. This can be watched at a later time.

Social expansion of REXplorer

Social expansion
Bystanders are not a directly part of the game. But REXplorer takes place in a public space and through the communication with gestures and the character's voices attracts attention. The game was designed to be playable in groups.

The expansions define the requirements for a successful laboratory simulation. As we did not test either the save and play again nor the blog creation function the temporal expansion was neglected.

Simulating the spatial expansion

Spatially REXplorer is played in the whole city of Regensburg. Players have to move through the streets to navigate from character to character and fulfill their quests.

Simulating movement

The problem with a laboratory study is that people cannot travel around like if they were in a city. To cope with the situation we offered a treadmill and a indication on a map to simulate movement. The result can be found in Figure 4.2.

Figure 4.2: Experimental setup without stepper - it was added right to the treadmill

The indicator spot was Wizard-of-Oz'ed to make development easy. Utilizing the iStuff - Toolkit we showed the according buildings when people moved into a hotzone. This should help people imagine the city.

4.3.3 Deleting extraneous variables

Deleting extraneous variables

After a pretest we noticed that some properties of our laboratory, which can be seen in Figure 4.2 above, would reduce immersion in the study.

For example behind our monitors is a window facade. The sight is comfortingly as it shows trees and a meadow. When the door is shut the room is quiet. If the windows are open wind rustling could be heard.

Therefore, before the test began, we close the windows and shut the window blinds. As a quiet room is not comparable with street noise the setup included playback of street noise. Free audio files under the creatives common license can be downloaded at the Freesound project[1] .

To not distract the user from the pictures shown on the center screen we positioned the map in a different angle and moved the third, not used screen a bit in the background. The distance between all three screens was the same. So the view should be focused on the building pictures. The treadmill and the stepper was centered in front of the monitors. The result can be seen in Figure 4.3.

Figure 4.3: Experimental setup with prepared room

First we introduced the people to the test and REXplorer. Than the intro movie was shown. The treadmill and the stepper had the affordances to lean on it. In the pretest we could see that people played while leaning so in the final user study we asked the participants to step away from the treadmill to do gestures.

If the device crashed during the game people had to walk to the starting point again to get it fixed. This happened in

[1] http://freesound.iua.upf.edu/

the Regensburg test also.

People played for an hour, the same time as in Regensburg. After the test they had to fill out the questionnaires. To make the users feel comfortable we laid some sweets out and offered water before and after the test. Right after the test the windows were opened again.

4.4 Comparing the results

Comparing the results

To compare the field study and the laboratory experiment we handed out the same questionnaires after each test. The structure of the questions were explained above but actually the players did not see our classification and answered in each question what felt right to them.

Total number of errors

For the sake of organization and analysis we decided to regroup the answers to fit the original structure. The following tables show errors that occurred during the user study. The row Q references the question in the questionnaire above in which the answer originally appeared. It should not be necessary to understand the answers. We start with a comparison of the total number of errors.

Expansion	Errors in R	Errors in AC	Common	Total
Spatial	7	3	3	7
Technical	21	23	21	23
Social	2	1	1	2
Conceptual	6	3	3	6
Sum	**36**	**30**	**28**	**38**

Table 4.1: Total number of errors in each setup

We can already see that results are often similar. The part that revealed the most errors is the technical one.

4.4.1 Spatial

Errors in the spatial section

We start the detailed comparison with the spatial expansion.

Q	Errors	R	A
Sp1	City Noise through construction sites, buses, ...	X	X
Sp2	Got lost but found back by using the map.	X	X
C12	Got bored when moving from place to place	X	X
Sp1	Traffic passing by disturbed the game	X	
Sp4	Players oversee traffic	X	
T3	City Noise through other tour guides	X	
So3	The players had problems to find the way	X	
Sum		**7**	**3**

Table 4.2: Comparison between errors that occurred in the spatial expansion (*Sp = spatial, T = technical, So = social, C = conceptual; # = Question number*)

Similarities
In both setups people complained about traffic noise. We think that this is an interesting finding. While we expected that result in Regensburg it shows that our simple setup with playback of random city sounds was successful. This is also a sign that we limited the calm nature around the building as extraneous variable.

Even more interesting is that both setups revealed that people can get lost in the city. The laboratory setup contained one big map of Regensburg displayed on the right screen in Figure 4.3 which indicated the actual position with a big red dot. The dot was used to simulate the movement for the participants and we expected that this would compromise the results according to the getting lost problem.

A conclusion from this could be that the paper map has to be evaluated again as it seems like it does not support orientation well.

Exclusively in the field
There are things that cannot be simulated with out experimental setup. That includes traffic passing by or getting in

danger because of the traffic. This is a shortcoming of our system that cannot be simulated easily.

While in both setups people got lost, only players in the field had problems to find the way. So this may be a result of the displayed map and the red dot, which make it is easy to keep an overview of the city.

If recorded audio of tour guides would be available it would be easy to play it back randomly when entering a hotzone and to extend the simulator to that case. The code for displaying a building when a hotzone is already there and so it would be easy to add a random component and audio playback.

Exclusively in the laboratory
The only things we could find exclusively in the laboratory were flaws with the setup like players who had problems to use the treadmill and a fire alarm that interrupting the game. This does not contribute to our comparison between of usability and playability errors in REXplorer and so these can be neglected.

Conclusion
The test brought up some interesting findings. The field study revealed problems with traffic, orientation and city guides. The last of these problems can be fixed. Recording a city guide and playing the voice back randomly would be relatively easy.

But then the traffic and orientation problem is unique to the city and cannot be tested in the laboratory without lots of effort. An approach to simulation would be using a system like Tatus, which we discussed in chapter 2, and script traffic. It is questionable if this extra effort would pay off. The consequences of these shortcomings will be summed up in the final conclusion at the end of the chapter.

4.4.2 Technical

Errors in the technical section

The second part of our detailed description deals with the technical part of our considerations.

Q	Errors	R	A
Sp1	The device crashed	X	X
T4	The device is too big	X	X
T4	The band is too long	X	X
T1	Interacting with the game including gestures and location picking was not clear	X	X
C11	Gesture picking after a wrong gesture was not understood	X	X
T1	Device's features like the questlist were overseen	X	X
T1	The instructions on the back of the map were overseen	X	X
T6	The introduction movie played back too fast	X	X
T6	The introduction movie should include sound	X	X
T6	The introduction movie text was fuzzy	X	X
T6	The movie missed essential information about how to use the device or start the game	X	X
Sp1	Device did not recognize gestures	X	X
T5	Trying to draw a gesture too often is boring. Possible retries should be limited.	X	X
T1	The feedback while drawing a gesture was wrong	X	X
T5	Gestures should be pickable everytime	X	X
T5	The gesture interaction should be more complex with more gestures	X	X
T1	Interaction with the device suffered from delays / unresponsiveness	X	X
T2	Some character's accent was not understandable	X	X
T2	Error messages and lengthy dialogues could not be interrupted	X	X
T2	Some times the sound was distorted	X	X
C12	Players expected vibration to indicate a hotzone	X	X
C9	One location could not be picked		X
C9	Suddenly the maximum number of quests was reached		X
Sum		**21**	**23**

Table 4.3: Comparison between errors that occurred in the technical expansion (*Sp = spatial, T = technical, So = social, C = conceptual; # = Question number*)

Similarities
The setups revealed very similar errors. All flaws that could be found in the field study were also discovered in the laboratory experiment. This shows that testing the device's stability and the interaction with the system are not bound to the city and can also be tested in laboratory setting. Below some of the findings and how we attacked them are explained in detail.

The most important finding was that the device crashed

from time to time. This goes together with an error reporting delays and unresponsiveness of the device. The players reacted on delays of a few seconds with repeatedly doing the gestures. That made the device crash. We reacted with putting more testing and debugging effort into development. The final product was stable.

Another flaw was that the participants did not know when to do a gesture and how to use the device's features to organize their game. The player also had problems to find their way in to the game, which is bad in particular as the play time is restricted to one hour.

The introduction movie was shown to make these things clear and the paper map we handed out included instructions on its back. But then people complained about the movie running too fast, including fuzzy text and they did not take a look on the back of the map. The movie has been improved to run slower and make gestures clearer. The instructions on the map have been extended and are shown to the people before playing.

Both setups revealed that the gesture recognition rejected gestures that seem to be right. They also explained that gestures should be pickable everytime. From our experience that only occurred because of the gesture recognition as people tend to like it if it works for them. In chapter 3 we already explained that the gesture feedback was changed to be not as sensitive. This was further improved after the test resulting in people ending the gestures and a better recognition.

Interestingly in both situations players asked for more complex interaction adding more gestures. On this hand this may be because they liked the challenge, as described in chapter 3, or the thought it would make things clearer.

The players were not able to interrupt dialogues with characters. This was mostly a problem when they did a gesture several times and they were requested to try again repeatedly. As it is important that the first dialogue presents the gesture only the option to interrupt the error message was added.

Exclusively in the laboratory
The laboratory experiment revealed two errors that were not found in the field study. We believe that the first one, a non pickable location, was rather found by accident then by the setup.

The second error revealed that the system was not initialized properly after a crash. This was not found in the field study because all pupils played at the same time while the laboratory experiment was conducted one group after the other. This does not show an advantage of the laboratory experiment because it probably would also occur when testing groups consecutively.

Conclusion
This section showed that evaluating the system in the laboratory revealed the same interaction flaws. As the technical part deals with the device and the interaction with it is not bound to the environment it is used in. While we can conclude that that this part can be tested equally in both environment, a more interesting conclusion would be if the simulator is required at all for this test.

The reference would be a study sitting at a table which we will discuss at the overall conclusion below.

4.4.3 Social

Errors in the social section

The social part of our test is judged below:

Q	Errors	R	A
T5	Doing a gesture repeatedly can be embarrassing	X	X
Sp1	Players got embarrassed by being watched	X	
Sum		**2**	**1**

Table 4.4: Comparison between errors that occurred in the social expansion (*Sp = spatial, T = technical, So = social, C = conceptual; # = Question number*)

Similarities
Both setups revealed that the players may be embarrassed

when doing the gestures repeatedly in public. Actually, while the participants of the field study really experienced the embarrassment, users in the laboratory only imagined that this may be a problem for them.

Exclusively in the field
In the laboratory there were no other people besides the participants and the conductor. So it was not able to show the problem by being watched.

Conclusion
Players can already imagine that they would be embarrassed when being watched. However it will not be possible to further test the social part in our laboratory setup as no other people are moving around in the room.

4.4.4 Conceptual

The last aspect we analyzed was the conceptual part. A detailed description can be found below:

Errors in the conceptual section

Q	Errors	R	A
C9	The playtime of one hour was too short	X	X
C2	The mistery around the Kindergrab was not unraveled	X	X
Sp1	The game was not understood	X	X
Sp6	Historical information was too superficial	X	
C7	Players cheated with asking for the way	X	
C7	Players cheated because they wanted to prevent the ping pong effect	X	
Sum		**6**	**3**

Table 4.5: Comparison between errors that occurred in the conceptual expansion (*Sp = spatial, T = technical, So = social, C = conceptual; # = Question number*)

Similarities
Both setups revealed that the play time is too short, a few players did not get the game and expected that the mystery about the Kindergrab was unraveled. This shows that it is not only possible to test interaction with the system in a laboratory study but also to evaluate if the content and the gameplay is understood.

Exclusively in the field
The field test revealed that the historical content was too superficial. As the field test was conducted with a class of pupils we could not control which expectations were aroused. In the laboratory study we only showed the video and did not introduce REXplrorer in detail. So this error may occur because of the expectations.

The way to cheat in REXplorer is to ask for the way or choose a location which is not the player's current location. The first one cannot be tested in the laboratory as there are no bystanders to ask. The second way occurred when players suffered from the ping pong effect we explained in chapter 3. Players in the laboratory suffered from the same problem but we could observe that they looked at the conductor at situation in which they wanted to cheat and then did not do it because they felt watched.

Conclusion
Summing it up the test of the conceptual part was possible. Further testing including a controlled experiment with a setup that makes sure that the same expectations are raised is needed to see if the superficial historical content cannot occur in the laboratory.

Cheating could be enabled by moving the conductor to another room. The indication point on the map is controlled with a keyboard and the iStuff Toolkit. So it should not be a problem to move from another room.

4.5 Summary

Summary and conclusions

The idea for the laboratory setup was to find a cheap way that reveals the same errors as a field test. This demand could not be reached. The only area which even showed even more errors than the field test was the technical part.

This is important for the game to function well but it is questionable if it pays off to put in the extra effort for the laboratory setup. A test while sitting at a table could reveal fewer but the most critical errors also. Examples are the

crashing device, the introduction movie playing too fast or uninterruptable error messages.

Examples for things that go beyond the device errors and could only be revealed with methods that go beyond low-fidelity testing are problems with traffic noise and the device volume or players getting lost in the city.

It is useful and could especially be used for middle fidelity prototypes not as exchange for the field test. This has to be done finally but the design process is supported and the gap is filled.

Ideas for further experiments and development are explained in the next chapter.

Chapter 5

Summary and future work

"The future is here. It's just not widely distributed yet."

—William Gibson

The evaluation of pervasive games is still an emerging field. The work at hand contributed by describing strategies for pervasive games. While we got useful results from the methods in use we think they still can be improved. This chapter summarizes our work so far and gives ideas about future development and research.

5.1 Summary and contributions

This work contributes to the field of evaluating pervasive games and supports an important phase of the iterative design approach—the evaluating. Therefore we began with describing the problems that occurred during evaluation due to the spatial, temporal and social expansions of the magic circle. Here we could already see that one traditional method alone would not be enough to test pervasive games.

After that we showed how a combination of traditional methods can be used to evaluate pervasive games. The chosen methods included product-interactive focus group interviews and video recordings. The practical use of the methods was attested by an exemplary field study.

The findings offered new insights into the pervasive game and a lot of valuable information. The affinity analysis approach should be general enough to be adapted by other pervasive games. The temporal expansion is a bottleneck as analyzing the video recordings can take a lot of time.

On the one hand, the field study showed detailed results, on the other hand, it should be conducted with high fidelity prototypes to justify the time and effort invested. So we identified a gap in the testing possibilities. Low-fidelity prototypes, like paper prototyping, can be used to identify flaws on a very low level while high-fidelity testing like field studies, offer ways to find very detailed results.

But there are not many methods available to test middle-fidelity prototypes like the results after only a few interations. To optimally support the iterative design process in that case, we developed a quick and cheap experimental setup for the laboratory, that simulates the experience of playing in the field.

We compared results from a field test and a laboratory experiment with a questionnaire we handed out in both conditions. The results showed that it is sensible to conduct such tests even if they do not find all errors which the field study offers. We suggest to use similar setups to test ideas during the iterative design process, but conduct at least one final field test to evaluate the system in the user's natural environment.

5.2 Future work

5.2.1 Improving field studies

The field study suffered from the high effort to analyze the video data. It would be interesting to see a merge of the QCCI tool, which can be found in chapter 2, and our approach. This would mean that our video would be completed with timestamps and analysis possibilities automatically. This would hopefully result in quicker analysis.

Another idea would be to integrate physiological measurements which can be collected automatically depending on which methods are protable enough to be used in a field study. This way we would be able to include not only the evaluation of interaction but also fun and enjoyment to test the playability of games.

5.2.2 Improving laboratory studies

The comparison between field and laboratory study suffered from the great amount of extraneous variables. One idea for further development would be to extract and adapt several variables, like different devices and sound volume, and measure their influence on the results.

An even more interesting approach would be to conduct experiments in the laboratory and see if the results can give conclusions about the field. Another point of interest is the comparison between our results and lower forms. For example it would be valuable to test if the experimental setup reveals more errors than sitting at a table and testing the device without moving at all.

Appendix A

Interview Guideline

Interview – Leitfaden

1. Vorstellen der Interviewer (namentlich vorstellen)

... es handelt sich um eine Evaluationsstudie und Ihre Meinung ist uns wichtig. Daher würden wir gerne ein Video – Interview durchführen. Die Ergebnisse daraus fließen in die finale Version unseres Projektes ein.

2. Ablauf des Gesprächs

Das Gespräch wird circa 30 Minuten dauern. Im Verlauf dessen möchten wir verschiedene Aspekte des Spiels aus Ihrer Sicht zu erfassen und Ihre Meinung dazu einfangen.

Doch bevor wir anfangen: Haben Sie noch Fragen?

3. Eingangsfrage

Sie haben gerade circa 1 Stunde mit unserem REXplorer – Spiel verbracht. Als erstes würden daher gerne wissen:

- Wie hat es Ihnen gefallen?

[Sollte es hier nicht weitergehen, so hilft eventuell die Frage:]

- Haben Sie Kritikpunkte am Spiel? Welche sind das? [Notieren !]

4. Aufspaltung

[Je nach Kritikpunkten kann auf den nächsten Bereich übergegangen werden. Es ist sinnvoll die Kritikpunkte zu notieren, falls verschiedene Bereiche auf einmal angesprochen werden.

Mögliche Übergänge zwischen den Bereichen wären:

Wir haben bisher den Bereich XY besprochen.

Wenn Sie an das Spiel denken, wir fanden Sie es im Bezug auf YZ?]

4.1 Räumlicher Aspekt

- Gab es für Sie irgendwelche Probleme bei der Orientierung im Spiel?

Goal: *How good worked the combination of the map and the controller? Is it enough to provide people with information they need for the game?*

- Konnten Sie sich in das Spiel vertiefen oder wurden sie durch

irgend etwas gestört? Durch was?
Goal: *Did the spatial component of the game add some disturbance like roads that had to be crossed, people are laughing at the participants when doing the gestures, ...? Were they in danger?*

- Hatten Sie eine Strategie um den passenden Weg zur Erfüllung der Aufgaben zu finden? Wie sah sie aus?
Goal: *Here i am trying to figure out how people actually play our game.*

4.2 Zeitlicher Aspekt

- Wie beurteilen Sie die Dauer des Spiels? Wie kommen Sie zu dieser Beurteilung?
Goal: *I would like to know if the hour of gaming was too short or if they even got bored.*

- Gab es Probleme in das Spiel hinein zu finden?
Goal: *Trying to find out if the introduction video is enough and i people need time to find their way around in the game.*

4.3 Sozialer Aspekt

- Wie war es für sie alleine / zu zweit zu spielen?
Goal: *Trying to find out if it is more interesting to play such games alone or in groups. Normally only one player can actually "play" with the controller. So is standing at their side interesting enough?*

- Was hatten Sie für ein Gefühl, wenn Passanten Ihnen beim Spielen zu gesehen haben?
Goal: *Trying to figure out the embarassing stuff again. Pervasive games are very new and people not playing need to get used to it. How does this work so far?*

4.4 Technischer Aspekt

- Welche Erfahrungen haben Sie im Umgang mit dem Zauberstab gemacht?
Goal: *How good does the controller work. Was it easy for people to play around with it?*

- Wie haben Sie den Transport des Gerätes empfunden?
Maybe: *Wie haben Sie das Gerät überhaupt transportiert?*
Goal: *I am trying to understand if people feel comfortable with carrying around the controller or if they even have better ideas how to transport it?*

4.5 Konzeptioneller Aspekt

- Was meinen Sie war Ziel des Spiels?
Goal: *Did they get the game? If i understood it right we want a game which does not need extra people to run around in the city. So the intro - / outro - movie must be enough to explain the game to people.*

- Welchen Teil des Spiels mochten Sie am meisten?

- Hatten Sie irgendwelche Probleme mit den Spielcharakteren zu kommunizieren? Welche waren das? Hatten Sie Probleme die richtigen Gesten zu finden?

- Im Spiel wurden Ihnen Aufgaben von den Charakteren erteilt. Wie sind sie damit umgegangen? Hatten Sie eine Strategie? Wie sah diese aus?
Goal: *Same thing as above. How do they play the game? Did we expect that?*

- Wie haben sie die Aufgaben verwaltet?
Goal: *Trying to answer the question: How good did they work with the controller? Were they avaible to figure out the log functions?*

5. Abschluss des Interviews

Tja, das war es im Prinzip auch schon. Wir möchten uns mit einer Kleinigkeit erkenntlich zeigen [Geschenk (Schlüsselanhänger??) überreichen], hoffen, dass es Ihnen Gefallen hat und Sie später vielleicht noch einmal Lust haben die finale Version des Projekts zu besuchen.

Appendix B

Questionnaire

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

rex
Regensburg Experience

RWTH AACHEN

Nachdem Du nun eine Stunde gespielt hast, würde uns Deine Meinung zu unserem Spiel interessieren.

Bitte nimm Dir etwas Zeit und beantworte die kommenden Fragen. Deine Anregung / Kritik fließt in unser Spiel ein und hilft uns, es noch besser zu machen.

Zu Deiner Person

Alter:

Geschlecht:

- Magst Du Computerspiele? Was magst Du daran oder eben nicht?

- Hat Dich dieses Spiel motiviert? Wenn ja, wie?

- Hast Du schon mal an einer Stadtführung teilgenommen? Falls ja, wo war das und wie hat es Dir gefallen?

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

rex
Regensburg Experience

RWTH AACHEN

Räumlicher Aspekt

- Bist Du während des Spiels gestört worden? Wenn ja, wodurch?

- Hast Du Dich jemals verlaufen? Falls ja, wie hast Du reagiert?

- Wie hast Du entschieden, welche Route Du zur nächsten Aufgabe nimmst?

- Gab es eine Situation, in der Du in Gefahr warst? Falls ja, beschreibe sie bitte.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

rex
Regensburg Experience

RWTH AACHEN

- Wie wichtig war es für Dich eine Karte zu haben?

unwichtig						sehr wichtig

- Findest Du, dass das Spiel nach Regensburg passt? Wenn ja, warum? Wenn nein, warum nicht?

- Wie hat sich die Stadt durch das Spiel für Dich verändert?

- Welche Rolle hat die Stadt für das Spiel gespielt?

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

rex
Regensburg Experience

RWTH AACHEN

Technischer Aspekt

- Wie bist Du mit der Bedienung des Gerätes zurecht gekommen?

- Hattest Du jemals Probleme, die Stimmen der Charaktere zu verstehen? Bitte beschreibe die Situation.

- War für Dich die Lautstärke des Gerätes ausreichend? Falls nein, in welchen Situationen war dies nicht der Fall?

- Wie hast Du das Gerät getragen? War das angenehm für Dich?

- Fandest Du die Kommunikation per Gesten passend? Was könnte verbessert werden?

- Hat der Einführungsfilm Dir auf ausreichende Weise erklärt, wie das Spiel funktioniert?

Sozialer Aspekt

- Wie fandest Du es, in der Öffentlichkeit zu spielen?

- Wie habt ihr die Aufgaben in der Gruppe verteilt?

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

rex
Regensburg Experience

RWTH AACHEN

- Gab es Situationen, in denen Du Dich vom Spiel ausgeschlossen gefühlt hast? Beschreibe diese bitte.

- Was glaubt Du welche Rolle Du in diesem Spiel angenommen hast?

- Hattest Du Lust dieses Spiel zu spielen?

gar nicht						sehr

- Ich empfand das Spiel als

riskant						sicher

Konzeptioneller Aspekt

- Wie viel würdest Du bezahlen um REXplorer spielen zu können?

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

rex
Regensburg Experience

RWTH AACHEN

- Nachdem Du unsere Einführung gesehen hast Welche Vorerwartung hattest Du an das Spiel?

- Wie würdest Du einem Freund REXplorer in einem Satz beschreiben?

- Wie hast Du die Spielzeit empfunden?

zu kurz						zu lang

- Welches war Dein Lieblingscharakter?

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

rex
Regensburg Experience

RWTH AACHEN

- Gab es Inhalte, die Dir unvollständig vorkamen? Beschreibe diese bitte.

- Hast Du im Spiel einmal gemogelt? In welcher Situation? Gab es einen Grund dafür?

- Hast Du über das Spiel hinaus Interesse an Regensburg gewonnen? Wo würdest Du Dir zusätzliche Informationen besorgen?

- War Dir das Ziel des Spiels klar? Was war das Ziel und konntest Du es erreichen? Falls nein, warum nicht?

- Wie hast Du Deine Aufgaben verwaltet?

- War Dir klar, wann eine Geste auszuführen ist und wann nicht?

- Hast Du Dich während des Spiels gelangweilt? Beschreibe bitte die Situationen in denen das der Fall war.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

rex
Regensburg Experience

RWTH AACHEN

– Ich habe das Spiel empfunden als

langweilig						unterhaltend
neuartig						gewöhnlich
natürlich						unnatürlich
anspruchslos						lehrreich
ermüdend						erfrischend
informativ						nicht informativ
amüsant						albern

Vielen Dank für Deine Teilnahme. Wir hoffen, dass es Dir gefallen hat.

Das REXplorer - Team

Bibliography

Macintosh Human Interface Guidelines. Addison-Wesley Pub. Co, 1992.

R. Anastasi, N. Tandavanitj, M. Flintham, A. Crabtree, M. Adams, J. Row-Farr, J. Iddon, S. Benford, T. Hemmings, S. Izadi, et al. Can You See Me Now? A Citywide Mixed-Reality Gaming Experience.

Rafael Ballagas, Steffen P. Walz, Sven Kratz, Claudia Fuhr, Eugen Yu, Martin Tann, Jan Borchers, and Ludger Hovestadt. Rexplorer: A mobile, pervasive spell-casting game for tourists. In *To appear in CHI '07 extended abstracts on Human factors in computing systems*, San Jose, CA, USA, 2007. ACM Press.

L. et al Barkhuus. Picking pockets on the lawn: The development of tactics and strategies in a mobile game. *Proc. Ubicomp*, pages 358–374, 2005.

J. Barton and V. Vijayaraghavan. UBIWISE, A Simulator for Ubiquitous Computing Systems Design. Technical report, HPL-2003-93. 2003, Palo Alto, CA: Hewlett-Packard Labs. 17., 2003.

M. Bell. Interweaving mobile games with everyday life, 2006.

Steve Benford and Mauricio Capra. Design and evaluation literature review. (IPerG-Deliverable-D5.2), 2006.

Steve Benford, Andy Crabtree, Martin Flintham, Adam Drozd, Rob Anastasi, Mark Paxton, Nick Tandavanitj, Matt Adams, and Ju Row-Farr. Can you see me now? *ACM Trans. Comput.-Hum. Interact.*, 13(1):100–133, 2006.

Hugh Beyer. *Contextual Design: Defining Customer-Centered Systems*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1997.

Jan Borchers. *A Pattern Approach to Interaction Design*. John Wiley and Sons, 2001.

S. Carter, J. Mankoff, S. Klemmer, and T. Matthews. Exiting the cleanroom: On ecological validity and ubiquitous computing. *HCI Journal*, To appear.

Scott Carter, Jennifer Mankoff, and Jeffrey Heer. Momento: support for situated ubicomp experimentation. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 125–134, New York, NY, USA, 2007. ACM Press.

M. Chalmers, L. Barkhuus, M. Bell, B. Brown, M. Hall, S. Sherwood, and P. Tennent. Gaming on the edge: Using seams in pervasive games. *Proc. Pervasive, PerGames workshop*, 2005.

Alan DIX. *Human-Computer Interaction*. Prentice Hall, Englewood Cliffs, 2003. ISBN 0130461091.

C. Fabricatore, M. Nussbaum, and R. Rosas. Playability in Action Videogames: A Qualitative Design Model. *Human-Computer Interaction*, 17(4):311–368, 2002.

S. Hinske, M. Lampe, C. Magerkurth, and C. Röcker. Classifying pervasive games: On pervasive computing and mixed reality. 2007.

Johan Huizinga. *Homo Ludens*. Beacon Press, June 1971. ISBN 0807046817.

Jesper Kjeldskov and Jan Stage. New techniques for usability evaluation of mobile systems. *Int. J. Hum.-Comput. Stud.*, 60(5-6):599–620, 2004b.

Jesper Kjeldskov, Mikael B. Skov, Benedikte S. Als, and Rune T. H¯egh. Is it worth the hassle? exploring the added value of evaluating the usability of context-aware mobile systems in the field. In *Mobile HCI*, volume 3160 of *Lecture Notes in Computer Science*, pages 61–73. Springer, 2004a.

Smith-Jackson-T. L. Nussbaum M. A. Tomioka K. Lee, Y. S. and Y Bhatkhande. Use of product-interactive focus groups for requirement capture and usability assessment. *Proceedings of the 48th Annual Human Factors and Ergonomics Conference*, pages 2461–2465, 2004.

Irma Lindt, Jan Ohlenburg, Uta Pankoke-Babatz, and Sabiha Ghellal. A report on the crossmedia game epidemic menace. *Comput. Entertain.*, 5(1):8, 2007. ISSN 1544-3574.

Ohlenburg J.-Pankoke-Babatz U. Ghellal S. Oppermann L. Lindt, I. and M. Adams. Designing cross media games. *In Proc. PerGames Workshop at Pervasive 2005*, 2005.

L. Liu and P. Khooshabeh. Paper or interactive?: a study of prototyping techniques for ubiquitous computing environments. *Conference on Human Factors in Computing Systems*, pages 1030–1031, 2003.

Carsten Magerkurth, Adrian David Cheok, Regan L. Mandryk, and Trond Nilsen. Pervasive games: bringing computer entertainment back to the real world. *Comput. Entertain.*, 3(3):4–4, 2005.

R.L. Mandryk and K.M. Inkpen. Physiological indicators for the evaluation of co-located collaborative play. *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 102–111, 2004.

David R. Millen. Rapid ethnography: time deepening strategies for hci field research. In *DIS '00: Proceedings of the conference on Designing interactive systems*, pages 280–286, New York, NY, USA, 2000. ACM Press.

M. Montola. Exploring the edge of the magic circle. defining pervasive games. *In Proc. of DAC 2005*, 2005.

M. Montola. Domain of pervasive gaming. (IPerG-Deliverable-D5.3B), 2006.

R. Morla and N. Davies. Evaluating a Location-Based Application: A Hybrid Test and Simulation Environment. *IEEE Pervasive Computing*, 3(3):48–56, 2004.

J. Nielsen. Iterative User-Interface Design. *Computer*, 26 (11), 1993.

Eva Nieuwdorp. The pervasive interface: Tracing the magic circle. *Proceedings of DiGRA 2005 Conference*, 2005.

D. Norman. *The Design of Everyday Things*. Basic Books, New York, 2002.

Eleanor O'Neill, Martin Klepal, David Lewis, Tony O'Donnell, Declan O'Sullivan, and Dirk Pesch. A testbed for evaluating human interaction with ubiquitous computing environments. In *TRIDENTCOM '05: Proceedings of the First International Conference on Testbeds and Research Infrastructures for the DEvelopment of NeTworks and COMmunities (TRIDENTCOM'05)*, pages 60–69, Washington, DC, USA, 2005. IEEE Computer Society.

Jennifer Preece. *Interaction Design*. J. Wiley and Sons, New York, 2002.

Katie Salen and Eric Zimmerman. *Rules of Play : Game Design Fundamentals*. The MIT Press, October 2003.

I. Satoh. Flying Emulator: Rapid Building and Testing of Networked Applications for Mobile Computers. *Proceedings of Conference on Mobile Agents (MAí2001), LNCS*, 2240:103–118.

Ben Shneiderman. *Designing the User Interface*. Addison Wesley Longman, 1998.

A.K. Sinha and J.A. Landay. Capturing user tests in a multimodal, multidevice informal prototyping tool. *Proceedings of the 5th international conference on Multimodal interfaces*, pages 117–124, 2003.

S. Walz and R. Ballagas. Digra paper. 2007.

Mark Weiser. The computer for the 21st century. *SIGMOBILE Mob. Comput. Commun. Rev.*, 3(3):3–11, 1999.

Index

Typeset July 30, 2007